Tennessee Law Review

Volume 75
Issue 1 *2007-2008*

Article 1

2008

Volume 75

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 75 | Fall 2007 | Number 1

CONTENTS

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 75 Winter 2008 Number 2

SYMPOSIUM: LOOKING FORWARD: THE NEXT SIXTY YEARS OF CLINICAL LEGAL EDUCATION

CONTENTS

DEDICATION: A TRIBUTE TO DOUGLAS A. BLAZE, DIRECTOR, THE UNIVERSITY OF TENNESSEE LEGAL CLINIC 1993-2006

Doug Blaze began directing The University of Tennessee (UT) Legal Clinic in the fall of 1993. Under Doug's leadership, the UT Legal Clinic undertook a timely examination of its clinical offerings, expanded the clinical programs to involve more students in the clinical experience, and increased clinical offerings to meet the needs of students who did not see themselves in a litigation practice. Doug also sought to augment the opportunities for students to engage in pro bono work, to provide service to those unable to afford counsel, and to begin the process of involving the students in the obligation of the profession to give back to society. His work on the local, state and national level brought recognition to the UT Legal Clinic and the law school.

When Doug arrived, the UT Legal Clinic had two primary branches: a criminal clinic and a civil clinic. The civil clinic consisted of four separate clinics: a housing clinic, a social security benefits clinic, an unemployment benefits clinic, and a general litigation clinic. Under Doug's leadership, the Clinic faculty examined the lawyering experiences that the respective clinics were providing to our students. While all of the clinics provided students with opportunities to interview and counsel clients, few students were exposed to the full range of lawyering tasks that lawyers typically employ in practice. The faculty, under Doug's direction, merged the civil clinics and the criminal clinic into one single advocacy clinic—a clinic that sought to combine cases in a way that would provide greater opportunities for students to perform a more complete array of lawyering roles.

This revision of our clinical offerings better serves our students and our state. Many of our students, upon graduation from law school, work in practice settings in which they have primary responsibility for client representation. In their respective practice settings, they often have little or no access to a mentoring program. Providing a clinical experience that more nearly replicates practice arms our students with the skills they use in the day-to-day practice of law. Clients of these former students are more likely to have a lawyer who has not only read the law, but has also performed the skills requested to competently address the client's legal needs.

The Advocacy Clinic reaffirmed our commitment to the principles that underlay the creation of the UT Legal Clinic in 1947. As Doug Blaze noted in his article *Déjà Vu All Over Again*,[1] "clinical education does not have an inherent, pre-determined set of goals. The challenge then, as now, is determining those objectives that can most effectively be achieved for any

1. 64 TENN. L. REV. 939 (1997).

particular program. Moreover, selection of educational objectives is fluid, changing with time and audience."[2]

As director of the Clinic, Doug was cognizant of the need for our students to understand some basic lawyering skills before entering practice. He saw the need to broaden and deepen our offerings in order to provide a more meaningful clinical experience. He masterminded the change in our clinical offerings so as to better achieve the five goals of clinical teaching articulated by John Bradway and effectuated by Charlie Miller when he established the UT Legal Clinic in 1947.[3] First, the student is to receive practical experience in order to bridge the gap between theory and practice.[4] Second, the clinical experience should provide a forum for the student to synthesize substantive and procedural law through application.[5] Third, the clinic should be a laboratory for the students to study the client as a whole, not just as someone with a legal problem.[6] Fourth, clinical education should bring context to professionalism.[7] And finally, the clinic should allow a student to take a case from start to finish and to learn to plan strategically.[8]

Doug also saw that some of our law students avoided participating in the Clinic because they did not see themselves as litigators after graduation from law school. Others sought specialization. The former sought a transactional law practice or perhaps a concentration in mediation. Under Doug's direction, a business law clinic and a mediation clinic were established. For students seeking to specialize in criminal law, a public defender externship and a prosecutorial externship were created. A judicial externship now presents students with a view of judicial decision-making. A domestic violence clinic exposes students to family law. And a death penalty clinic provides students with an appellate experience in a complex area of law.

Doug believes in the concept of equal access to justice for all. As Director of the UT Legal Clinic, he guided us in the selection of cases that met the unserved needs of those unable to afford counsel in the Knoxville area. When the Legal Services Corporation forbade representation of public housing tenants facing eviction for alleged drug or criminal charges,[9] the UT Legal Clinic undertook representation of those tenants whenever we could. But Doug's work on behalf of the indigent did not stop there. He was instrumental in establishing a vibrant pro bono program at the UT College of Law.[10] He served

2. *Id.* at 947.
3. *See id.* at 947–50.
4. *Id.* at 947.
5. *Id.* at 948.
6. *Id.*
7. *Id.*
8. *Id.*
9. 45 C.F.R. § 1637 (2002).
10. Doug found that 33% of UT law students participate in pro bono programs in collaboration with 60 alumni lawyers, providing nearly 1400 hours of pro bono service annually. Douglas A. Blaze, *Toward Equal Access to Justice: Rethinking the Role of Law*

as faculty advisor of that program for many years, and through his service, he introduced countless students to the opportunity and obligation to provide service to those unable to afford counsel—to the homeless, to children, to families, to immigrants, and to battered spouses, among others. The program has been so successful that a UT law student normally receives the Tennessee Bar Association's annual award for outstanding service by a law student in this state.

Several years ago I was asked by the Access to Justice Coordinator of the Tennessee Bar Association why the UT College of Law had such a robust pro bono program. I told her that, in my opinion, it was because of Doug Blaze. He was a charismatic leader who led by example. He was someone students respected and wanted to emulate. And what he taught them about pro bono was correct: it was the right thing to do.

Doug has advocated for an even larger role for law schools in the provision of pro bono services. He believes that those unable to afford counsel would be better served if law schools partnered with the private bar in the provision of legal services to the poor. Not only do law schools have resources, students, connections to the bar, and alumni, but relationships with other departments in the university allow for a more holistic, interdisciplinary approach to the social and legal problems presented by the clients.[11] Moreover, a partnership between the bar and the law student body would not be limited by the restrictions placed upon legal services programs.

Doug's commitment to public service in his personal capacity is outstanding as well. He gives tirelessly of his time to countless community and bar projects. He sets a pace for the rest of us that is nearly impossible to meet. Not long after his arrival at the University of Tennessee, he was asked to serve on a Tennessee Bar Association committee charged with the task of devising a program to assist recent law graduates as they enter law practice—a bridge-the-gap program. This program would have provided a mentoring experience to transition the graduate from the classroom to law practice. It would have underscored any clinical experience the graduate had participated in while in law school, introduced the student to the business side of law practice, and stressed the ethical aspects of the lawyer-client relationship. Unfortunately, the Tennessee Supreme Court did not implement the program.

A second major undertaking by Doug was his leadership role in the Tennessee Alliance of Legal Services (TALS). Doug served as chair of TALS from 2001 to 2004, during the time that the Legal Services Corporation was advocating for consolidation of legal services programs within states. This was a difficult time because local programs and local boards often resisted consolidation. Doug skillfully mediated the disputes between local programs and successfully guided the consolidation in Tennessee.[12] As a result of his

---

Schools, 2 TENN. J. LAW & POL'Y 66, 71 (2006).

11. *Id.* at 72.

12. In 2003 there were eight legal services programs in Tennessee: Memphis Area Legal Services, West Tennessee Legal Services, Legal Services of South Central Tennessee, Legal

outstanding work, TALS presented him with the B. Riney Green award in 2003.

Doug has continued to work with TALS and legal services lawyers since his time as president of that organization. At the annual TALS meetings, Doug has designed and implemented training programs for legal services attorneys. The programs have been universally well-received, with attorneys constantly asking him for more and more training.

Finally, let me attempt to describe Doug Blaze, the clinical teacher. Doug has always been an enthusiastic, energetic, engaged teacher. He is neither shy nor retiring. When my office was just a few doors down from Doug's, I often heard him discussing cases with his students. He believes in careful, thoughtful preparation of all clients' cases. When he was actively teaching in the Clinic, he always strove to allow the student to take the lead and to defer, whenever appropriate, to the student's judgment about how to handle an aspect of the case. But he also actively participated in the planning process with his students, which is necessary to ensure quality representation, to evaluate the student's performance, and to engage the student in self-reflection.

As teachers, we do not always accomplish our pedagogical goals. I was sitting in my office once when I heard Doug's booming voice exclaim, "You did what?!" It seems that two of his students had decided to skip the planning and strategy session before they negotiated their client's case. And unbeknown to Doug, the students had set up a negotiation with the assistant district attorney. I recall that the students obtained a good result for the client, but it would be fair to say that skipping the planning session and neglecting to tell Doug about the negotiation was akin to waving the proverbial red flag in front of the bull. I do not think the students ever forgot the message. Doug is a demanding teacher who is interested not only in getting a good outcome for the client but also in the process by which the outcome is derived. Sometimes we have to reign in the bull, but never the message.

As a teacher, Doug always respected and trumpeted client autonomy. In a civil case, I witnessed Doug and his students counsel a client to accept an offer in a particular case. It appeared that the client should have accepted the offer, but he elected to go forward with a trial; the client ultimately lost the trial, a

---

Services of Middle Tennessee, Rural Legal Services, Knoxville Legal Aid Society, Legal Services of Southeast Tennessee, and Legal Services of Upper East Tennessee. Each program had a board of directors consisting of local attorneys and local clients, as well as its own set of priorities for the types of cases it would handle. Consolidation had the potential of diluting the influence that local attorneys and clients might bring to bear on their respective programs—influence upon the actions of the legal services staff and upon the types of cases the offices would accept. There was also the potential that some administrators would lose their positions of power, for if there were fewer programs, there would be fewer directors, assistant directors, etc. Instead of eight programs, Tennessee now has four: Legal Aid of East Tennessee, Memphis Area Legal Services, Inc., Legal Aid Society of Middle Tennessee and the Cumberlands, and West Tennessee Legal Services, Inc. *See* Legal Services Corporation, Tennessee Program Information, http://www.lsc.gov/map/state_T32_R50.php (last visited Feb. 19, 2008).

result that netted less than the offer he had rejected. At the end of the hearing, the client walked out of the courtroom, came over to the students and to Doug, and thanked them for their work. In particular, the client stated that even though he had lost the trial, he was grateful that he had the opportunity to tell the judge his side of the case. The client was able to tell his story, and that was what was important to him. Doug is the kind of teacher who can bring such a result to fruition and discuss with the students—both beforehand and post mortem—the lessons to be learned from such a case.

On a personal note, Doug is a great colleague. Perhaps because he is an outstanding lawyer and teacher, he is someone I can go to for sound legal advice and strategy. I know of no other lawyer who I believe has a better analytical mind. He can dissect legal problems, rendering the parts understandable and manageable. He makes the problems comprehensible and their solutions, if not attainable, at least intelligible. He is always accessible. He even appears to enjoy the enterprise. We will miss him as our director, but he has left us with a strong foundation and we look forward to many years ahead with him as our clinical colleague.

JERRY P. BLACK, JR.*

*Associate Professor of Law and Former Director of Clinical Programs, University of Tennessee College of Law.

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 75 | Spring 2008 | Number 3

CONTENTS

TENNESSEE LAW REVIEW

Member of the National Conference of Law Reviews

Volume 75 | Summer 2008 | Number 4

CONTENTS

MEDICAL FUTILITY STATUTES: NO SAFE HARBOR TO UNILATERALLY REFUSE LIFE-SUSTAINING TREATMENT

THADDEUS MASON POPE*

ABSTRACT

Over the past fifteen years, a majority of states have enacted medical futility statutes that permit a health care provider to refuse a patient's request for life-sustaining medical treatment (LSMT). These statutes typically permit the provider to unilaterally stop LSMT where it would not provide "significant benefit" or would be contrary to "generally accepted health care standards." These safe harbors are vague and imprecise, however. Consequently, providers have been reluctant to utilize these medical futility statutes.

The uncertainty concerning these statutes most likely cannot be reduced. States have been unable to reach a consensus on substantive measures of medical inappropriateness. Only a purely process-based approach like that outlined in the Texas Advance Directives Act (TADA) has proven effective in inducing the conduct that medical futility statutes intended. Therefore, while the specific contours of TADA must be refined, policymakers in other states should look to the TADA as a model.

Table of Contents

---

* Visiting Assistant Professor of Law, Widener University School of Law (2007–2008); Assistant Professor of Law, University of Memphis Cecil C. Humphreys School of Law; J.D., Georgetown University Law Center, 1997; Ph.D., Georgetown University, 2003; B.A., University of Pittsburgh, 1992. The author received valuable criticism from presenting earlier drafts at the American Society of Law, Medicine & Ethics Health Law Teachers Conference (Baltimore, Md., June 2006), the Southeastern Association of Law Schools Annual Meeting (Palm Beach, Fla., July 2006), law faculty presentations at DePaul University College of Law and William Mitchell College of Law, and, most usefully, the Saint Louis University Law School Health Law Scholars Workshop (Sept. 2006). Thanks to the participants at these conferences for their comments, particularly to Michael Allen, Kathy Cerminara, Kelly Dineen, Jesse Goldner, Sandra Johnson, and Rob Schwartz. Thanks to Dr. Kenneth Leeds and the LACBA Bioethics Committee for inspiring this project. Finally, thanks to the excellent research assistance of Timothy Stehli. This Article was supported by a generous summer research grant from the University of Memphis.

INTRODUCTION

Esther Hutchison is a 97-year-old woman with metastasized cancer in her liver, kidneys, and lungs.[1] She will never again be conscious. Her medical treatment includes mechanical ventilation support and artificial nutrition and hydration. Pursuant to an advance directive, Mrs. Hutchison's daughter is her mother's agent for health care decisions. She wants the health care team to "do everything" to save her mother's life.

But, given her situation, Mrs. Hutchison's health care providers are uncomfortable with continuing to provide her with life-sustaining medical treatment (LSMT).[2] They want to switch her to comfort care.[3] Several meetings with the treatment team, ethics committee, social workers, and clergy have failed to change the daughter's treatment request. Now, the treatment team wants to withdraw treatment *without* the daughter's consent. The relevant health care law seems to authorize this unilateral action,[4] but the team and the hospital are unwilling to proceed. They are reluctant to do what they think is right and what the law allows.

During the 1990s, a significant number of professional medical associations and individual health care providers and institutions formally concluded that, under some circumstances, in cases of intractable conflict such as Mrs. Hutchison's, it would be appropriate for health care providers to

---

1. This is a fictional case based on the facts of many cases discussed in this Article.

2. LSMT refers to medical interventions that sustain the patient's life, but are not effective in helping the patient recover from a terminal condition or persistent vegetative state. These interventions may include assisted ventilation, artificial nutrition and hydration, renal dialysis, surgical procedures, blood transfusions, and the administration of drugs. Following the statutory convention, this Article refers to LSMT as a category. *See, e.g.*, ALA. CODE § 22-8A-3(8) (LexisNexis 2006); 755 ILL. COMP. STAT. ANN. 40/10 (West 2007). Yet, as Edmund Pellegrino notes, "[e]ach treatment must be evaluated in terms of its end . . . ." Edmund D. Pellegrino, *Decisions at the End of Life—The Abuse of the Concept of Futility*, PRACTICAL BIOETHICS, Summer 2005, at 3, 5.

3. *See infra* notes 79–80 and accompanying text.

4. This Article uses the term "unilateral action" to describe the situation in which the health care provider overrides a patient's or surrogate's request for LSMT. Where the provider acts unilaterally, she acts contrary to the instructions of the legally authorized decision maker. This usage is consistent with most of the literature. *See, e.g.*, Kathryn L. Moseley et al., *Futility in Evolution*, 21 CLINICS GERIATRIC MED. 211, 216 (2005). While the term "unilateral action" is also sometimes used to refer to a situation where the provider stops LSMT when the patient is incompetent and no surrogate is reasonably available, this Article does not cover such cases. In such a situation, there is no overriding authority because the provider typically becomes the authorized decision maker.

unilaterally withhold or withdraw LSMT.[5] But most health care providers were unwilling to act on these policies and guidelines without sufficient legal protection.[6] Many state legislatures responded by enacting statutes that purport to provide this protection and to authorize health care providers to unilaterally withhold or withdraw LSMT.[7]

But, as exemplified in Mrs. Hutchison's case, these unilateral decision statutes have failed to achieve their intended purpose. Today, even with explicit statutory authorization and grants of immunity, health care providers are *still* reluctant to unilaterally withhold or withdraw medically inappropriate LSMT.[8]

Futility disputes are becoming increasingly common.[9] Because providers want adequate legal authority to make unilateral decisions, it is important to diagnose the effects, or lack thereof, of the unilateral decision statutes. This Article reviews the history and effects of the unilateral decision statutes. Certainly, there are ongoing academic and legislative debates concerning whether unilateral decision making is even good public policy. Rather than directly engaging that debate, this Article assesses these statutes on their own terms.

Part One of this Article provides a brief overview of medical futility[10]

---

5. *See infra* notes 53, 280-82 and accompanying text.

6. *See infra* notes 276-90 and accompanying text.

7. *See infra* notes 291, 297-310 and accompanying text.

8. *See infra* notes 382-83, 387-403 and accompanying text.

9. *See infra* notes 84–88 and accompanying text. *See generally* Thaddeus Mason Pope & Ellen A. Waldman, *Mediation at the End-of-Life: Getting Beyond the Limits of the Talking Cure*, 23 OHIO ST. J. ON DISP. RESOL. 143 (2007) (explaining why mediation has failed as a mechanism for resolving the growing number of futility disputes).

10. This Article uses the term "medical futility" to describe only a type of dispute. *Cf.* Anne L. Flamm, *The Texas "Futility" Procedure: No Such Thing as a Fairy Tale Ending*, 11 LAHEY CLINIC MED. ETHICS J. 11, 11 n.1 (2004) (using "futility" for "sake of brevity" to describe situations where patients or surrogates demand LSMT that the health care provider believes to be un-useful or harmful); John Fletcher, *The Baby K Case: Ethical and Legal Considerations of Disputes about Futility*, 2 BIOLAW: A LEGAL AND ETHICAL REPORTER ON MEDICINE, HEALTH CARE AND BIOENGINEERING S:219, S:231-:233 (1994) (using "futility" to describe a "type of moral dispute"); Edmund D. Pellegrino, *Futility in Medical Decisions: The Word and the Concept*, 17 HEC FORUM 308, 309 (2005) (using "futility" as a clinical concept to describe the point where medical treatment "can no longer serve any recognizable good for the patient"). The term's pervasiveness in the literature justifies this much. Because the term is so troubled; however, this Article abandons "medically futile" in favor of "medically inappropriate" when referring to a type of treatment or intervention. *Cf.* Michael Ardagh, *Futility Has No Utility in Resuscitation Medicine*, 26 J. MED. ETHICS 396, 399 (2000) ("The words futile and futility should be abandoned . . . ."); Jeffrey T. Berger, Letter to the Editor, *Advance Directives, Due Process, and Medical Futility*, 140 ANNALS INTERNAL MED. 402, 403 (2004) (noting concept of medically appropriate, rather than futility, "integrates the society-based authority under which physicians operate with the physicians' fiduciary obligations to patients"); Raanan Gillon, *Futility—Too Ambiguous and Pejorative a Term?*, 23 J. MED. ETHICS 339, 339 (1997) (describing "futility" term as unclear and too complex); Eric M. Levine, *A New*

disputes, including both how they arise and how they are resolved. Part Two summarizes the leading definitions of "medical inappropriateness." These include brain death and physiological futility, where there is literally nothing that medicine can offer the patient. Other definitions of "medical inappropriateness" include concepts that are less scientifically measurable and more value-laden, including quantitative futility, qualitative futility, and generally accepted health care standards.

But definitions are not enough. Taking unilateral action has been and still is fraught with legal risks. Part Three outlines legal constraints on the unilateral withholding and withdrawing of LSMT. In particular, this Part reviews potential civil, criminal, and disciplinary sanctions that could result. Then, Part Four canvasses state legislation that purports to relieve providers from these constraints by authorizing the unilateral limitation of LSMT.

Part Five examines the effects of these unilateral decision statutes. While some evidence suggests that unilateral decision statutes facilitate the informal resolution of disputes, they do not provide a workable solution against intractable disputes. The unilateral decision statutes were meant to permit providers to decline to comply with requests for medically inappropriate treatment. But providers continue to comply with such requests. Not only have most health care institutions never adopted a futility policy, but most of those that have a futility policy have never implemented it. Yet, there is a notable exception in Texas where providers do unilaterally stop LSMT.

Part Six analyzes why the unilateral decision statutes have failed to achieve their intended objective. In particular, this Part contends that despite statutory authorization and grants of immunity, providers are "chilled" from unilaterally stopping treatment because of legal uncertainty. There are three potential sources of this uncertainty: (1) the vagueness of the state statutes, (2) their potential federal preemption, and (3) their potential unconstitutionality. Since Texas providers are subject to the same federal and constitutional restrictions, this Article posits that the relevant "chilling" uncertainty must come from the vagueness of the state statutes.

Finally, Part Seven offers some suggestions on how to eliminate this statutory vagueness. There are two primary options: (1) legislate concrete, measurable, and predictable clinical criteria; or (2) legislate a concrete, measurable, predictable process. No consensus exists on the precise, legislatable measures of medical inappropriateness. Apparently then, only a purely process-based approach, like the one adopted in Texas, can effectively protect the conduct that medical futility statutes were designed to protect. While the current formulation of that process may not be sufficiently fair and

---

Predicament for Physicians: The Concept of Medical Futility, the Physician's Obligation to Render Inappropriate Treatment, and the Interplay of the Medical Standard of Care, 9 J. L. & HEALTH 69, 84 n.104 (1994-1995) (avoiding use of terms "futile" and "useless" by replacing them with "medically inappropriate"); Thaddeus Mason Pope, *Is Public Health Paternalism Really Never Justified? A Response to Joel Feinberg*, 30 OKLA. CITY U. L. REV. 121, 202-06 (2005) (discussing the dangers of employing "thick" terms).

rigorous, Texas's pure process approach wizens (if not eliminates) uncertainty and should serve as a model for other states.

I. OVERVIEW OF MEDICAL FUTILITY

A. Dying in America

Modern advances in science and medicine have made possible the prolongation of the lives of many seriously ill individuals, without always offering realistic prospects for improvement or cure.[11] "Halfway" technologies such as mechanical ventilation and artificial nutrition and hydration can sustain biological life for practically indefinite periods of time but may not themselves lead to improvement or cure.[12]

As a consequence of the availability of these life-sustaining technologies, most deaths in America occur in an institutional setting such as a hospital.[13] Most of these institutional deaths are the result of an intentional, deliberate decision to stop LSMT and allow death.[14] Nancy Dubler explains that "[d]eath is a negotiated event; it happens by design. . . . 70% of the 1.3 million Americans who die in health care institutions do so after a decision has been made and implemented to forego some or all forms of medical treatment."[15]

---

11. *See* Alan Meisel & Bruce Jennings, *Ethics, End-of-Life Care, and the Law: Overview, in* LIVING WITH GRIEF: ETHICAL DILEMMAS AT THE END OF LIFE 63, 63 (2005) ("Most of the cases and dilemmas that have shaped the law on end-of-life care have involved patients whose lives could be prolonged by new medical treatments and technologies, but whose health, functioning, quality of life, and even conscious awareness itself could not be restored."). *See generally,* WILLIAM H. COLBY, UNPLUGGED: RECLAIMING OUR RIGHT TO DIE IN AMERICA 57-71 (2006) (discussing the ascent of medical technology in futility cases); JOHN D. LANTOS & WILLIAM MEADOW, NEONATAL BIOETHICS: THE MORAL CHALLENGES OF MEDICAL INNOVATION 18–52 (2006) (describing moral controversies arising from advances in neonatal care).

12. John Lantos, *When Parents Request Seemingly Futile Treatment for their Children,* 73 MOUNT SINAI J. MED. 587, 588 (2006); Gay Moldow et al., *Why Address Medical Futility Now?,* MINN. MED., June 2004, at 38, 38.

13. *See* Thomas Wm. Mayo, *Living and Dying in a Post-Schiavo World,* 38 J. HEALTH L. 587, 587–88 n.3 (2006) (citing S. 570, 109th Cong. § 2(a)(1) (2005)) (stating that eighty percent of deaths in America occur in hospitals).

14. *See* Arthur E. Kopelman, *Understanding, Avoiding, and Resolving End-of-Life Conflicts in the NICU,* 73 MOUNT SINAI J. MED. 580, 580 (2006) ("Eighty percent of the deaths that occur in the neonatal intensive care unit (NICU) are preceded by decisions to limit, withhold, or withdraw life support . . . ."); Pellegrino, *supra* note 2, at 3 ("[T]he majority of patients in modern hospitals today die as a result of a deliberate decision to withhold or withdraw treatment."); Thomas J. Prendergast & John M. Luce, *Increasing Incidence of Withholding and Withdrawal of Life Support from the Critically Ill,* 155 AM. J. RESPIRATORY & CRITICAL CARE MED. 15, 15 (1997) ("[W]ithholding or withdrawal of life support precedes 40 to 65% of deaths in intensive care facilities.").

15. Nancy Dubler, *Limiting Technology in the Process of Negotiating Death,* 1 YALE J.

B. The Right to Die

For some individuals the possibility of extended life is meaningful and beneficial. For others, the artificial prolongation of life may provide nothing beneficial and serve only to extend suffering and prolong the dying process. To accommodate these varying attitudes, the rise of modern life-sustaining medical technologies was accompanied by the rise of patient autonomy.[16]

During the 1970s and 1980s, appellate courts across the country decided numerous cases in which patients and patients' families wanted to withdraw or withhold LSMT but health care providers were reluctant to cede to such requests.[17] These cases firmly established the right of patients to refuse LSMT.[18] These cases also established the right of surrogates to exercise this right for patients who were incompetent and unable to exercise it for themselves.[19]

Today, all states have laws enabling patients and surrogates to refuse medical care.[20] Patients and surrogates decide whether LSMT is beneficial

---

HEALTH POL'Y L. & ETHICS 297, 297 (2001) (reviewing MANAGING DEATH IN THE INTENSIVE CARE UNIT: THE TRANSITION FROM CURE TO COMFORT (J. Randall Curtis & Gordon D. Rubenfield eds., 2001) [hereinafter MANAGING DEATH]); *see* Thomas J. Prendergast et al., *A National Survey of End-of-Life Care for Critically Ill Patients*, 158 AM. J. RESPIRATORY & CRITICAL CARE MED. 1163, 1163 (1998). *See generally* COLBY, *supra* note 11, at 95–107 (discussing the correlation between the increasing life expectancy and the rising use of LSMT).

16. *See* Matthew S. Ferguson, *Ethical Postures of Futility and California's Uniform Health Care Decisions Act*, 75 S. CAL. L. REV. 1217, 1230 (2002) ("As we moved into the 1990s, however, patients became consumers of medical technology, often forcing the hands of their doctors by seeking to determine when treatment should be applied.").

17. *See generally* ALAN MEISEL & KATHY CERMINARA, THE RIGHT TO DIE § 2 (3d ed. 2005 & Supp. 2007) [hereinafter THE RIGHT TO DIE] (discussing legal development of end-of-life decision making); CLAIRE C. OBADE, PATIENT CARE DECISION-MAKING: A LEGAL GUIDE FOR PROVIDERS chs. 7-8 (1991 & Supp. 2006) (providing case law guidance for balancing patient rights with medical responsibilities).

18. *See generally* THE RIGHT TO DIE, *supra* note 17, at § 2 (tracing the right to die from its common law roots to Supreme Court jurisprudence); OBADE, *supra* note 17, at chs. 7-8 (discussing exceptions to the general rule requiring treatment and the legal bases for such exceptions).

19. *See generally* THE RIGHT TO DIE, *supra* note 17, at § 4; OBADE, *supra* note 17, at chs. 9, 11. This Article employs the term "surrogate" to refer to all those who are authorized to make health care decisions on behalf of the patient, whether appointed by the patient herself (e.g, agents, surrogates), by a court (e.g., guardians, conservators), or by default legal rules (e.g., surrogates). Most patients are unable to communicate with providers at the time decisions are made about stopping LSMT. *See* MANAGING DEATH, *supra* note 15, at 364; Seth Rivera et al., *Motivating Factors in Futile Clinical Interventions*, 119 CHEST J. 1944, 1945 (2001) ("None of the patients were able to participate in the decision-making process of their own care since they were universally too impaired."). Therefore, these decisions are usually made by surrogates.

20. *See generally* THE RIGHT TO DIE, *supra* note 17, at § 7; OBADE, *supra* note 17, at app. A.

given their own values and particular circumstances.[21] Health care providers must generally comply with decisions to refuse LSMT.[22]

C. Nature of Medical Futility Disputes[23]

A medical futility dispute arises when a health care provider seeks to stop LSMT that the patient or surrogate wants continued.[24] A medical futility dispute is sometimes referred to as a "reverse right to die,"[25] a "right to life,"[26] a "duty to die,"[27] or even an "involuntary euthanasia"[28] situation. In a classic right to die situation, the patient or the surrogate wants to limit LSMT but the

---

21. *See infra* notes 37-50 and accompanying text.

22. *See, e.g.*, Rodriguez v. Pino, 634 So. 2d 681, 683 (Fla. Dist. Ct. App. 1994) (finding doctor who complied with patient's initial refusal of LSMT could not be held subsequently liable for patient's death); Osgood v. Genesys Reg'l Med. Ctr., No. 94-26731-NH (Genesee County Cir. Ct. Feb. 16, 1996), *noted in* THE RIGHT TO DIE, *supra* note 17, §11.01[A] (Supp. 2005) (awarding $16.6 million verdict where the ICU provided LSMT to a patient in contravention of her agent's demands); Estate of Leach v. Shapiro, 469 N.E.2d 1047, 1051 (Ohio Ct. App. 1984) (authorizing battery action for maintaining a PVS patient on a respirator against her previously expressed wishes); *see also* Elena N. Cohen, *Refusing and Forgoing Treatment: Liability Issues*, *in* 3 TREATISE ON HEALTH CARE LAW §§ 18.07[1] & 18.07[2] nn.45-56 (Alexander M. Capron & Irwin M. Birnbaum eds., 2005) (detailing the success of claims for providing unwanted treatment on various legal grounds); *Barriers to End of Life Care—Not in My ER, Not in My Nursing Home*, 11 L. & HEALTH CARE NEWSL., Spring 2004, at 16, 20 [hereinafter *Barriers*] (reporting Maryland state agency fined nursing home for failing to heed resident's advance directive); Amy Lynn Sorrel, *Lawsuit Showcases DNR Liability Twist for Doctors*, AM. MED. NEWS, Feb. 5, 2007, *available at* http://www.ama-assn.org/amednews/2007/02/05/prl20205.htm (noting that courts in Florida increasingly hold providers liable for providing unwanted LSMT).

23. There is an enormous literature on the definition of "medical futility" and the ethical justifiability of unilateral decisions. This Article provides neither a conceptual analysis nor a normative defense of "medical futility." While these issues provide essential context, this Article focuses on the effects of the unilateral decision statutes and on the effectiveness of their safe harbors.

24. *See* Flamm, *supra* note 10, at 11 n.1 (medical futility describes situations where patients or surrogates demand LSMT which the health care provider believes to be un-useful or harmful)

25. *See, e.g.*, Mayo, *supra* note 13, at 602 n.68; *see also* THE RIGHT TO DIE, *supra* note 17, § 13.01[B] at 13-4 (referencing the "reverse end-of-life").

26. *See, e.g.*, Nancy Neveloff Dubler, *Conflict and Consensus at the End of Life*, 35 HASTINGS CTR. REP. (SPECIAL REPORT), Nov.-Dec. 2005, at S19; Leigh B. Middleditch Jr. & Joel H. Trotter, *The Right to Live*, 5 ELDER L.J. 395, 397 (1997); Wesley J. Smith, *Suing for the Right to Live*, THE DAILY STANDARD, Mar. 11, 2004, *available at* http://www.weeklystandard.com/Content/Public/Articles/000/000/003/836zeecs.asp.

27. *See, e.g.*, Smith, *supra* note 26.

28. *See, e.g.*, Mary Ann Roser, *Debate Heats Up on 10-Day Medical Law*, AUSTIN AM.-STATESMAN, Aug. 10, 2006, at B1.

health care provider resists.[29] This is represented as situation (3) in the diagram below. In contrast, in a futility situation, the roles are reversed such that the health care provider wants to limit LSMT and the patient or the surrogate resists.[30] This is represented as situation (2) in the diagram below.

	Provider: "LSMT yes"	Provider: "LSMT no"
Patient/Surrogate: "LSMT yes"	(1) Consensus – no dispute	(2) *Medical futility dispute*
Patient/Surrogate: "LSMT no "	(3) Classic right to die dispute	(4) Consensus – no dispute

In a futility dispute, it is the health care provider, rather than the patient or surrogate, who judges LSMT as unbeneficial.[31] In other words, it is the health care provider who wants to stop the train when the patient or surrogate says, "keep going."[32]

Often the surrogate and the health care provider's disagreement over whether LSMT provides a benefit is caused by a failure in communication; the surrogate and provider perceive the situation differently.[33] In other cases, the disagreement is normative.[34] Whether for factual or normative reasons,

---

29. *See, e.g.*, Mayo, *supra* note 13, at 587.

30. For the sake of economy, this Article assumes that there are only two relevant players: the patient and the health care provider. Of course, things are actually often far more complicated. When, as is often the case, the patient is incompetent, it may not always be clear who is the appropriate decision maker or there may be intra-family disagreement as to the proper action. *See, e.g.*, *In re* Doe, 418 S.E.2d 3, 7 (Ga. 1992) (finding hospital could not enter DNR order where mother agreed to DNR order for daughter but father did not); Lebreton v. Rabito, 650 So. 2d 1245, 1246-47 (La. Ct. App. 1995) (allowing daughter's lawsuit against physicians for withdrawing LSMT from father because withdrawal was authorized by wife/mother but strongly suggesting that her claim had no merit); NANCY NEVELOFF DUBLER & CAROL B. LIEBMAN, BIOETHICS MEDIATION: A GUIDE TO SHAPING SHARED SOLUTIONS 10 (2004); Troyen A. Brennan, *Ethics Committees and Decisions to Limit Care*, 260 JAMA 803, 806 (1988) (remarking that health care provider's recommendation of DNR order for an incompetent patient is controversial where there was no family present to make decision or family was divided over choice). Similarly, on the provider side there may be disagreement among residents, nurses, or attending physicians. *See* Warthen v. Toms River Cmty. Mem'l Hosp., 488 A.2d 229, 230 (N.J. Super. Ct. App. Div. 1985) (reviewing termination of nurse's employmet for refusing to administer dialysis to terminally ill patient); Arthur U. Rivin, *Futile Care Policy: Lessons Learned from Three Years' Experience in a Community Hospital*, 166 W. J. MED. 389, 390 (1997).

31. *See* K. Francis Lee, *Postoperative Futile Care: Stopping the Train When the Family Says "Keep Going"*, 15 THORACIC SURGERY CLINICS 481, 481 (2005).

32. *Id.*

33. *See* JOSEPH J. FINS, A PALLIATIVE ETHIC OF CARE: CLINICAL WISDOM AT LIFE'S END 82–86 (2006); *see also infra* notes 70-72, 89-92 and accompanying text.

34. *Cf.* FINS, *supra* note 33, at 82-86 (describing how most, but not all, disagreements between patients and surrogates or providers are caused by miscommunication).

however, the provider and surrogate disagree because they have different goals.[35] The patient's goals might include cure, amelioration of disability, palliation of symptoms, reversal of disease processes, or prolongation of life. The provider, on the other hand, might judge these goals to be impossible, virtually impossible, or otherwise inappropriate under the circumstances.[36]

1. Patient and Surrogate Reasons for Insisting on Treatment

Surrogates are often inclined to request that "everything [be] done."[37] There are many reasons that surrogates insist on continuing treatment that their health care provider considers medically inappropriate. Surrogates might think that the health care provider's prognosis is wrong, perhaps distrusting that the patient is receiving proper care either because of their race or socioeconomic status[38] or because of their provider's financial incentives.[39] A significant volume of scientific literature demonstrates that patients from racial and ethnic minorities more frequently and more adamantly demand LSMT.[40]

---

35. *Cf.* Thomas Wm. Mayo, *Health Care Law*, 53 SMU L. REV. 1101, 1109-10 n.78 (2000) ("[T]he disagreement is over what constitutes a 'benefit' to the patient . . . ."). The Supreme Court observed that when questioning the benefit of LSMT, the relevant question to ask is "effective at doing what?" LSMT for Nancy Cruzan, after all, was "100 percent effective at sustaining life." Transcript of Oral Argument at *28, Cruzan v. Dir., Mo. Dept. Health, 497 U.S. 261 (1989) (No. 88-1503).

36. *See infra* Part III (providing definitions of medical inappropriateness).

37. *See, e.g.*, LAWRENCE J. SCHNEIDERMAN & NANCY S. JECKER, WRONG MEDICINE: DOCTORS, PATIENTS, AND FUTILE TREATMENT 22-34 (1995) [hereinafter WRONG MEDICINE]; John Ellement, *Woman Suing MGH Tells Court of Distress*, BOSTON GLOBE, Apr. 8, 1995, at B18; Donalee Moulton, *Death, Denial and the Law*, 40 MED. POST (Toronto), May 4, 2004, at 29 ("[T]his is the recommendation of a doctor or health-care team not to do anything further, to stop treatment of not proceed with a treatment. It is a recommendation patients and families often refuse to accept").

38. *See, e.g.*, FINS, *supra* note 33, at 78 ("An especially difficult dynamic can arise when the family believes that the patient's dire condition was precipitated by a medical error or if they are suspicious that substandard care is being provided because the patient is from a traditionally marginalized population."); Lee, *supra* note 31, at 483; Moseley, *supra* note 4, at 212-13; Mary Ellen Wojtasiewicz, *Damage Compounded: Disparities, Distrust, and Disparate Impact in End-of-Life Conflict Resolution Policies*, 6 AM. J. BIOETHICS, Sept.-Oct. 2006, at 8-12; Pam Belluck, *Even as Doctors Say Enough, Families Fight to Prolong Life*, N.Y. TIMES, Mar. 27, 2005, at A1 (reporting that some "patients and families . . . are skeptical of doctors' interpretations or intentions").

39. *See* Pope & Waldman, *supra* note 9, at 164-65.

40. *See, e.g.*, William Bayer et al., *Attitudes Toward Life-Sustaining Interventions Among Ambulatory Black and White Patients*, 16 ETHNICITY & DISEASE 914 (2006); Ursula K. Braun et al., *Decreasing Use of Percutaneous Endoscopic Gastronomy Tube Feeding for Veterans with Dementia—Racial Differences Remain*, 53 J. AM. GERIATRIC SOC'Y 242 (2005); Marion Danis, *Improving End-of-Life Care in the ICU: What's to be Learned from the Outcomes Research*, 6 NEW HORIZONS 110 (1998); Michael N. Diringer et al., *Factors Associated with Withdrawal of Mechanical Ventilation in a Neurology/Neurosurgery Intensive Care Unit*, 29 CRITICAL CARE

Even if not distrustful of health care providers, surrogates might be in denial or under a "therapeutic illusion" that the patient can recover or that a new therapy will come along.[41] Access to online medical information makes surrogates more confident in opposing providers' recommendations.[42] Even in the face of clear and dire medical facts, family members often hold out hope that the patient will "beat the odds."[43]

Even when surrogates appreciate that the odds are exceedingly slim, they may believe that those odds are still worth pursuing. They might believe that God will perform a miracle.[44] They might otherwise be compelled by religious or cultural traditions.[45]

MED. 1792 (2001); Kevin Fiscella, *Socioeconomic Status in Healthcare Outcomes: Selection Bias or Biased Treatment,* 42 MED. CARE 939 (2004); Joanne Mills Garrett et al., *Life-Sustaining Treatments During Terminal Illness: Who Wants What?*, 8 J. GEN. INTERNAL MED. 361 (1993); Faith P. Hopp & Sonia A. Duffy, *Racial Variations in End-of-Life Care*, 48 J. AM. GERIATRIC SOC'Y 658 (2000); Hilary Waldman, *End-of-Life Care, Viewed in Stark Black and White*, L.A. TIMES, Feb. 6, 2006, at F5. *But see* Amber E. Barnato et al., *Racial Variation in End-of-Life Intensive Care Use: A Race or Hospital Effect?*, 41 HEALTH SERVICES RES. 2219, 2219 (2006) (arguing that differences were attributable to the use of hospitals with higher ICU use rather than to racial differences).

41. *See* Middleditch & Trotter, *supra* note 26, at 402-03 (discussing modern "culture's persistent denial of death's reality"); Stacey A. Tovino & William J. Winslade, *A Primer on the Law and Ethics of Treatment, Research, and Public Policy in the Context of Severe Traumatic Brain Injury*, 14 ANNALS HEALTH L. 1, 2 n.5, 26 n.153 (2005) (discussing "therapeutic illusions" where patients have "false hopes despite the lack of future benefit").

42. Julie Sneider, *Medical Ethics Experts See Shift in Care Disputes*, MILWAUKEE BUS. J., Apr. 22, 2005, *available at* http://www.bizjournals.com/milwaukee/stories/2005/04/25/focus2.html.

43. *See* Clare Dyer, *Doctors Need not Ventilate Baby to Prolong His Life*, 329 BMJ 995, 995 (2004) (reporting that two mothers of terminally ill infants rejected medical advice because their babies were "'fighters'. . . [and] had lived longer than doctors had predicted . . . ."); Todd Ackerman, *Hospital Rules to Unplug Baby Girl: Leukemia Patient's Parents Scramble to Find New Care Facility*, HOUSTON CHRON., Apr. 30, 2005, at B1 (reporting that the mother of Knya Dismuke-Howard, a six-month old girl with leukemia in her brain, multiple organ failure, and a life-threatening antibiotic-resistant infection stated, "I think she can beat the odds . . . . She's a fighter."); Belluck, *supra* note 38, at A1 ("Extraordinary medical advances have stoked the hopes of families."); Bill Murphy, *Life and Death Matter Goes to Court: Comatose Man's Relatives Fighting State Law, Hospital to Keep Him Alive*, HOUSTON CHRON., Mar. 18, 2001, at A37 (reporting that relatives opposed to removing life support did not "share the conclusion that [patient's] condition [was] hopeless"). *Cf. In re* Guardianship of Schiavo, 851 So. 2d 182, 186 (Fla. Dist. Ct. App. 2003) ("[W]e understand why a parent . . . would hold out hope . . . . If Mrs. Schiavo were our own daughter, we could not but hold to such a faith.").

44. *See, e.g., In re* Baby K, 832 F. Supp. 1022, 1026 (E.D. Va. 1993) ("The mother opposes the discontinuation of ventilator treatment . . . because she believes that all human life has value . . . . [and] that God will work a miracle if it is his will."); Lee, *supra* note 31, at 483; Robert Sibbald et al., *Perception of "Futile Care" Among Caregivers in Intensive Care Units*, 177 CANADIAN MED. ASS'N J. 1201, 1204 (2007); *Parents Fear Home Delay May Keep 'Miracle' Baby Charlotte in Hospital*, BIRMINGHAM POST (UK), Jan. 7, 2006, at 3 (reporting that

The surrogates may feel a sense of responsibility or guilt with respect to their relationship to the patient.[46] They might be too grief stricken to stop treatment.[47] Or they might—consistent with the technological imperative in

---

parents of Charlotte Wyatt were "committed Christians" who believed that "miracles do happen")Ed Yeates, *Parents Fight to Keep Son on Life Support* (KSL TV5 broadcast Oct. 13, 2004) (transcript on file with Tennessee Law Review) (parents sought an injunction to stop physicians from disconnecting their son from life support even though he was declared dead because "we performed a miracle and I don't see why we can't do that again").

45. *See, e.g.*, Rideout v. Hershey Med. Ctr., 30 Pa. D. & C.4th 57, 62 (Dauphin County Ct. C.P. Dec. 29, 1995) (No. 872S1995), 1995 WL 924561 (parents opposed to removing ventilator from daughter because of "religious belief that all human life has value and should be protected"); James Bopp, Jr. & Richard E. Coleson, *Child Abuse by Whom?—Parental Rights and Judicial Competency Determinations: The* Baby K *and* Baby Terry *Cases*, 20 OHIO N.U. L. REV. 821, 841 (1994) ("I cannot make that decision to terminate life. God did not give me that power." (quoting Brief of Appellant, *In re* Achtabowski, No. 93-1247-AV, at 39 (Mich. Cir. Ct. Aug. 12, 1993) (No. 170251))); Lee, *supra* note 31, at 483; John Carvel, *Muslim Family Lose Right-to-Life Appeal*, GUARDIAN, Sept. 2, 2005, *available at* http://www.guardian.co.uk/print/0,,5276201-103690,00.html (noting the "family's religious conviction"); Bill Murphy, *Comatose Man Dies After Battle Over Life Support: Family Cited Spiritual Beliefs*, HOUSTON CHRON., Mar. 23, 2001, at A29 (reporting that for spiritual and cultural reasons, the family of Joseph Ndiyob sought an injunction preventing Memorial Hermann Hospital from removing Ndiyob's life support); Emily Ramshaw, *Children Fight to Save Mom: Carrollton Hospital Seeks to End Care of Woman with Brain Injury*, DALLAS MORNING NEWS, Aug. 18, 2006, at B1 (injured woman's children believed that their mother, "Ruthie Webster, [was] deeply religious and believe[d] only God should give and take life"); Kevin Rollason, *Jewish Kin Say Pulling Plug Would Be a Sin,* WINNIPEG FREE PRESS, Dec. 11, 2007, at A4 (family of Samuel Golubchuk sought injunction against removing LSMT since Orthodox Jews believe life must be extended as long as possible); Benjamin Weiser, *The Case of Baby Rena: Who Decides When Care is Futile?—A Question of Letting Go: A Child's Trauma Drives Doctors to Reexamine Ethical Role*, WASH. POST, July 14, 1991, at A1 (discussing the religious views of Baby Rena's foster parents).

46. Lee, *supra* note 31, at 483 ("Many [surrogates] believe it is morally wrong to end a patient's life intentionally or to allow a patient's life to end without available interventions."); John J. Paris et al., *Has the Emphasis on Autonomy Gone Too Far? Insights from Dostoevsky on Parental Decisionmaking in the NICU*, 15 CAMBRIDGE Q. HEALTHCARE ETHICS 147, 147 (2006); Jan Hoffman, *The Last Word on the Last Breath*, N.Y. TIMES, Oct. 10, 2006, at F1 ("Families often believe that consenting to a D.N.R. order implies they are giving up on their loved one, signing a death warrant . . . ."); Ann Wlazelek, *Pendulum Swings in Life-Saving Efforts; Hospitals' Policies on Doing All They Can to Keep Patients Alive Have Changed*, THE ALLENTOWN MORNING CALL, June 13, 2004, at A1 ("'It's dangerous to give the family the last word since guilt and a desire to do everything for pop makes it emotionally impossible to stop treatment.'" (quoting Arthur Caplan)).

47. *See, e.g.*, Alexander Morgan Capron, *Abandoning a Waning Life*, 25 HASTINGS CTR. REP., July–Aug. 1995, at 24 (reporting that Massachusetts General Hospital wrote a unilateral DNR because "the family's unpreparedness for their mother's death did 'not justify mistreating the patient.'"); Ezekiel J. Emanuel & Linda L. Emanuel, *Proxy Decision Making for Incompetent Patients: An Ethical and Empirical Analysis*, 267 JAMA 2067, 2067–68 (1992) (discussing that many family members find that they cannot let the patient go).

American medicine[48]—simply believe that the patient is entitled to everything.[49] Whatever the reason, more and more surrogates want their health care providers to "do everything to save [the patient's] life."[50]

2. Provider Reasons for Resisting Treatment

In some circumstances, health care providers resist surrogate requests that "everything be done." Such resistance stems from a significant consensus that some requests for treatment are inappropriate and that health care providers should not comply with them.[51] While no consensus exists on the specific

---

48. This is the mindset that because doctors *can* use a given technology, they *should* use that technology. *See* Kathy Cerminara, *Dealing with Dying: How Insurers Can Help Patients Seeking Last-Chance Therapies (Even When the Answer Is "No")*, 15 HEALTH MATRIX: J. L.-MED. 285, 296 (2005) (commenting that this "technological imperative" has subordinated the general availability of health care services to the pursuit of medical research); Robert L. Fine, *The History of Institutional Ethics at Baylor University Medical Center*, 17 BAYLOR U. MED. CTR. PROC. 73, 81-82 (2004) (explaining how medical innovation causes "moral tension" in regards to "distributive justice and fairness"). *See generally* VICTOR R. FUCHS, WHO SHALL LIVE? HEALTH, ECONOMICS, AND SOCIAL CHOICE (1974) (describing the limitations that economics places on how health care resources are allocated in terms of both equity and efficiency).

49. *See, e.g.*, Kopelman, *supra* note 14, at 582–85; Alan Meisel, *The Role of Litigation in End of Life Care: A Reappraisal*, 35 HASTINGS CTR. REP. (SPECIAL REPORT), Nov.–Dec. 2005, at S49 ("A vocal proportion of the population . . . believes that life per se is a pearl beyond price and must be preserved at all costs . . . . This set of beliefs [is] known as 'vitalism' . . . ."); Rivin, *supra* note 30, at 392; James W. Walter, *Medical Futility—An Ethical Issue for Clinicians and Patients*, PRACTICAL BIOETHICS, Summer 2005, at 1, 1, 6. Particularly where LSMT is covered by insurance, it is financially easy for surrogates to insist on continued treatment. All the economic and social costs are external. The insurer pays through other policyholders. Health care providers, particularly nurses, bear the emotional burden of treating the patient. *See* Robert M. Taylor & John D. Lantos, *The Politics of Medical Futility*, 11 ISSUES L. & MED. 3, 9 (1995) (comparing the benefit to family and friends for prolonging the patient's life and the burden subsequently carried by the medical professionals and insurance companies); *see also* Todd Ackerman, *St. Luke's Postpones Removal of Life Support: Man's Family Has Until 3 p.m. to Explore Any Possible Appeals,* HOUSTON CHRON., Mar. 12, 2005, at B1 ("[T]he family understands there is no hope . . . [but] 'the decision when life support is removed should be [the family's], not a corporation's.'").

50. *See* News Release, Pew Res. Ctr. for the People and the Press, *More Americans Discussing— and Planning—End-of-Life Treatment: Strong Public Support for Right to Die* 24 (Jan. 5, 2006), *available at* http://people_press.org/reports/pdf/266.pdf (reporting that between 1990 and 2005, the percentage of Americans who want a doctor to "do everything to save life" increased from 15% to 22%); *see also* Sneider, *supra* note 42 ("[M]ore families are challenging doctors who believe additional medical treatment of a critically ill patient is unwarranted.").

51. For this reason, this Article starts with the controversial presumption that the law should facilitate health care providers' ability to unilaterally terminate LSMT. However, some physicians do not resist patient requests for inappropriate LSMT for several reasons. First, some treating physicians judge that the conflict is not worth the trouble, especially when they

criteria and conditions under which providers may decline to comply with requests for LSMT, the appropriateness of unilateral refusals has long been accepted.[52] In fact, a plethora of professional medical associations have issued policy statements supporting the unilateral withholding and withdrawal of inappropriate LSMT.[53]

---

will soon shift off rounds for that patient. *See, e.g.*, Capron, *supra* note 47, at 24 (reporting Catherine Gilgunn's original attending physician eventually deferred to the surrogate's request to continue LSMT, but a month later, the new attending physician did not); Susan Carhart, *Process Approach to End-of-Life Care Fails to Eliminate Ethical, Political Issues*, 11 BNA HEALTH L. REP. 1755, 1756 (2002) ("'[I]t's not worth the hassle . . . .'" (quoting Stephen Streat)). Second, some physicians accede to requests for inappropriate LSMT because they do not want to admit defeat. *See, e.g.*, WRONG MEDICINE, *supra* note 37, at 25–28; ROBERT ZUSSMAN, INTENSIVE CARE: MEDICAL ETHICS AND THE MEDICAL PROFESSION 109 (1992) (noting that doctors are "inclined towards activism"); MANAGING DEATH, *supra* note 15, at 377; Rivin, *supra* note 30, at 392 (discussing the physicians' "attitude that death is the enemy" which leads to a "compulsion to be thorough and to leave no possibility untried"); Tovino & Winslade, *supra* note 41, at 27 (discussing vitalism and the "heroic urge to rescue"). Third, some providers accede because of their own religious or cultural convictions. Rivin, *supra* note 30, at 392 tbl.2. Fourth, some agree with the requests out of a "desire to please the patient's family." *Id.* Fifth, some providers accede because of reimbursement incentives. *See* Tovino & Winslade, *supra* note 41, at 27.

52. *See, e.g.*, 2 HIPPOCRATES, *The Art*, *in* HIPPOCRATES 193 (W.H.S. Jones trans. 1923) (purpose of medicine includes "to refuse to treat those who are overmastered by their diseases, realizing that in such cases medicine is powerless"); PLATO, THE REPUBLIC 100 (408b) (Richard W. Sterling & William C. Scott trans., 1985) ("But they thought a man constitutionally sickly and intemperate was of no use to himself or anyone else. They believed that the art of medicine ought not to be squandered on his ilk and that he should not receive treatment even if he were richer than Midas."); Lee, *supra* note 31, at 484 ("According to Hippocrates, 'to attempt futile treatment is to display an ignorance that is allied with madness.'") (citing L. EDELSTEIN, ANCIENT MEDICINE: SELECTED PAPERS OF LUDWIG EDELSTEIN 97-98 (O. Temkin & C.L. Temkin eds., 1967)).

53. *See, e.g.*, AMA COUNCIL ON ETHICAL AND JUDICIAL AFFAIRS, CODE OF MEDICAL ETHICS §§ 2.035, 2.037 (2006–2007), *available at* http://www.ama-assn.org; AMA Council on Ethical and Judicial Affairs, *Medical Futility in End-of-Life Care: Report of the Council on Ethical and Judicial Affairs*, 281 JAMA 937, 938 (1999) [hereinafter AMA Council]; Am. College of Obstetricians and Gynecologists, *Committee Opinion: Opinion No. 362: Medical Futility*, 109 OBSTETRICS & GYNECOLOGY 791 (Mar. 2007); Am. Thoracic Soc'y, *Withholding and Withdrawing Life-Sustaining Therapy*, 144 AM. REV. RESPIRATORY DISEASE 726, 728 (1991); College of Physicians and Surgeons of Manitoba, *Statement No. 1602: Withholding and Withdrawing Life Sustaining Treatment* (Jan. 30, 2008), *available at* http://www.cpsm.mb.ca/statements/1602.pdf; Soc'y of Critical Care Med. Ethics Comm., *Consensus Statement Regarding Futile and Other Possibly Inadvisable Treatments*, 25 CRITICAL CARE MED. 887, 888 (1997); *2004 House of Delegates Action on Resolutions and Board Reports*, 103 WIS. MED. J. 91, 91 (2004) [hereinafter *House of Delegates Action*] (referencing the Wisconsin Medical Society Resolution 1-2004, which "establishes a legally sanctioned extra-judicial process for resolving disputes regarding futile care"). *See generally* BRIT. MED. ASSOC., WITHHOLDING AND WITHDRAWING LIFE-PROLONGING MEDICAL TREATMENT (2001) (providing guidance for such action).

The policy statements are primarily motivated by four concerns, the most significant of which is professional integrity. Physicians do not want to be indentured servants,[54] "reflexive automatons,"[55] "vending machines,"[56] "prostitutes,"[57] or "grocers"[58] beholden to provide whatever treatment patients or surrogates want. After all, medicine is not a "consumer commodity like breakfast cereal and toothpaste."[59]

The medical profession is a self-governing one with its own standards of professional practice.[60] The "integrity of the medical profession" is an important societal interest that must be balanced against patient autonomy.[61]

---

54. E. Haavi Morreim, *Profoundly Diminished Life: The Casualties of Coercion*, 24 HASTINGS CTR. REP., Jan.-Feb. 1994, at 18 ("The physician-patient relationship is not an irrevocable indentured servitude . . . .").

55. WRONG MEDICINE, *supra* note 37, at 58, 103-04 (stating that physicians are not obligated to do everything a patient wants).

56. WRONG MEDICINE, *supra* note 37, at 9; Lawrence J. Nelson & R.M. Nelson, *Ethics and the Provision of Futile, Harmful, or Burdensome Treatment to Children*, 20 CRITICAL CARE MED. 427, 431 (1992).

57. WRONG MEDICINE, *supra* note 37, at 126. Dr. Schneiderman has more recently further developed this analogy, noting that "there were some things [prostitutes] would not do no matter how much they were paid." LAWRENCE J. SCHNEIDERMAN, EMBRACING OUR MORTALITY: HARD CHOICES IN AN AGE OF MEDICAL MIRACLES 123 (2008).

58. Ellen Goodman, *The Shift from Dr. Partner to Dr. Provider*, BOSTON GLOBE, Oct. 24, 1993, at 85.

59. George J. Annas, *Asking the Courts to Set the Standard of Emergency Care—The Case of Baby K*, 330 NEW ENG. J. MED. 1542, 1545 (1994) (arguing for avoidance of the scenario where "physicians will do whatever patients want (as long as they can pay for it), because medicine will be seen as a consumer commodity like breakfast cereal and toothpaste"); *see also* Tom Tomlinson & Diane Czlonka, *Futility and Hospital Policy*, 25 HASTINGS CTR. REP., May-June 1995, at 29 ("[T]he value assumptions made in cases of futility will have to receive their warrant from . . . values for the profession."). *But see* Eric Gampel, *Does Professional Autonomy Protect Medical Futility Judgments?*, 20 BIOETHICS 92, 97 (2006) (arguing that while limits on physician autonomy are set by the norms of the medical community rather than by individual providers, those limits do not extend to the futility context).

60. Gampel, *supra* note 59, at 97 (referencing the "right of the medical profession to be a self-governing body, one which defines its own standards of professional practice").

61. *See generally* Washington v. Glucksberg, 521 U.S. 702, 731 (1997) ("The State also has an interest in protecting the integrity and ethics of the medical profession."); Superintendent of Belchertown State Sch. v. Saikewicz, 370 N.E.2d 417, 425 (Mass. 1977) ("The interest of the State in prolonging a life must be reconciled with the interest of an individual to reject the traumatic cost of that prolongation."); *In re* Quinlan, 355 A.2d 647, 663 (N.J. 1976) ("[T]he unwritten constitutional right of privacy . . . is broad enough to encompass a patient's decision to decline medical treatment under certain circumstances . . . ."); Ferguson, *supra* note 16, at 1239-43 (noting that the UHCDA attempts to protect the ethical integrity of the medical profession). The legal profession is similar to the medical profession in this respect. While generally the client is in charge, a lawyer can withdraw from representation if "the client insists upon taking action that the lawyer considers repugnant . . . ." ABA MODEL OF RULES PROF'L. CONDUCT R. 1.16(b)(4) (2006). Lawyers also have obligations under Rule 11 of the Federal

Indeed, patient autonomy "has never been construed as requiring a health professional to provide a particular type of treatment."[62] Since the medical profession determines the goals and values of medicine, it can judge certain requests as inconsistent with those goals and values.[63]

In particular, many health care providers do not consider the practice of medicine to include measures aimed solely at maintaining corporeal existence and biologic functioning.[64] Under these circumstances, providers feel that continued LSMT is just "bad medicine . . . medicine being used for the wrong ends."[65] Moreover, health care providers find it gruesome, distressing, and demoralizing to provide treatment that harms patients.[66]

Rules of Civil Procedure—lawyers cannot file frivolous lawsuits even if the client demands it. *See* 2 JAMES WM. MOORE, MOORE'S FEDERAL PRACTICE THIRD EDITION § 11.11[1] (3d ed. 2007).

62. Loane Skene, *Disputes about the Withdrawal of Treatment: The Role of Courts*, 32 J. L. MED. & ETHICS 701, 701 (2004) (citing Schwartz, *infra* note 105, at 32). Nevertheless, other legal principles (e.g., nondiscrimination) have been construed to require providers to provide treatment that they deemed inappropriate. *See infra* Part III.

63. *See* Gampel, *supra* note 59, at 97 (stating that a health care provider "may refuse treatments which the medical profession gauges to be inappropriate, i.e. as being inconsistent with the basic goals and values of medicine").

64. *See, e.g.*, College of Physicians and Surgeons of Manitoba, *supra* note 53, at 15-S4 ("A patient is not just a physical being, but a person with a body, mind and spirit expressed in a human personality of unique worth.").

65. *See* Weiser, *supra* note 45, at A1 (quoting Dr. Murray Pollack).

66. *See* ZUSSMAN, *supra* note 51, at 123–38; Robert A. Burt, *The Medical Futility Debate: Patient Choice, Physician Obligation, and End-of-Life Care*, 5 J. PALLIATIVE MED. 249, 253 (2002); Betty R. Ferrell, *Understanding the Moral Distress of Nurses Witnessing Medically Futile Care*, 33 ONCOLOGY NURSING F. 922 (2006); Terese Hudson, *Are Futile-Care Policies the Answer? Providers Struggle with Decisions for Patients Near the End of Life*, 68 HOSPITALS & HEALTH NETWORKS, Feb. 20, 1994, at 26, 28; Stacey Burling, *Penn Hospital to Limit Its Care in Futile Cases: Severely Brain-Damaged Patients Won't Get Certain Treatments, as a Rule*, PHILA. INQUIRER, Nov. 4, 2002, at A1; Hoffman, *supra* note 46, at F1 ("'[D]oing CPR [to end-stage patients] felt not only pointless, but like I was administering final blows to someone who had already had a hard enough life.'" (quoting Dr. Daniel Sulmasy)); Liz Kowalczyk, *Hospital, Family Spar Over End-of-Life Care*, BOSTON GLOBE, Mar. 11, 2005, at A1 [hereinafter Kowalczyk, *Hospital, Family Spar*] ("Howe's longtime doctors and nurses believe[d] . . . that keeping her alive [was] tantamount to torture."); Liz Kowalczyk, *Mortal Differences Divide Hospital and Patient's Family*, BOSTON GLOBE, Sept. 28, 2003, at A1 [hereinafter Kowalczyk, *Mortal Differences*] (reporting physician and nurse refused to participate in continued aggressive treatment of Barbara Howe); Elisabeth Rosenthal, *Rules on Reviving the Dying Bring Undue Suffering, Doctors Contend*, N.Y TIMES, Oct. 4, 1990, at A1 ("Doctors and nurses . . . describe anger and anguish at being forced by a patient or family to inflict pain on the dying, knowing that it is to no avail."); Gregory Scott Loeben, Medical Futility and the Goals of Medicine 98 (1999) (unpublished Ph.D. dissertation, University of Arizona) (on file with Tennessee Law Review) ("If such judgments are meant to benefit anyone, it makes more sense to say that it is the physician . . . uncomfortable with the role [he is] being asked to play . . . ."). *Cf.* TOM L. BEAUCHAMP & JAMES F. CHILDRESS, PRINCIPLES OF BIOMEDICAL ETHICS 38 (5th ed.

Second, in addition to professional integrity, providers resist inappropriate treatment requests out of concern for the patient. Continued interventions can be inhumane, invasive, pointless, intrusive, cruel, burdensome, abusive, degrading, obscene, violent, or grotesque.[67] For example, CPR can be painful, causing rib or sternal fractures in a majority of cases.[68] Health care providers want to shorten and ease patient suffering; they do not want to cause or prolong it.[69]

A third reason that providers resist requests for inappropriate treatment is that they do not want to offer false hope. If they acted as though a medically inappropriate option were "available," this would create a psychological burden on surrogates to elect that option regardless of their prior wishes.[70] Naturally, families want to at least take all reasonable measures. Yet, it is unfair and

---

2001) (defending the physician's right to autonomy and "conscientious objection" where the patient's request for something is "morally objectionable").

67. *See, e.g.*, *In re* Doe, 418 S.E.2d 3, 4 (Ga. 1992) (failing to reach hospital's allegation that continued treatment of a patient with degenerative neurological disease would constitute "medical abuse"); Wendland v. Sparks, 574 N.W.2d 327, 328-29 (Iowa 1998) (ignoring testimony that doctor's unilateral decision not to attempt CPR was "an act of mercy" because the patient's prospects for quality of life were "not good"); *In re* Dinnerstein, 380 N.E.2d 134, 137 (Mass. App. Ct. 1978) (characterizing LSMT as "pointless, even cruel, prolongation of the act of dying"); Brief of Appellants at 3, *In re* Baby "K," 16 F.3d 590 (4th Cir. 1994) (No. 93-1899), 1993 WL 13123742 ("This tragic case involves a parent's attempt to require physicians to provide to a dying infant treatment that is medically unreasonable, invasive, burdensome, inhumane, and inappropriate."); John Altomare & Mark Bolde, Note, *Nguyen v. Sacred Heart Medical Center*, 11 ISSUES L. & MED. 199, 200 (1995) (observing hospital alleged continued treatment was "cruel and inhumane"); Martha Kessler, *Court Orders Hospital to Comply with Decisions Made Under Health Proxy*, 13 BNA HEALTH L. REP. 527, 527 (2004) (reporting Massachusetts General Hospital successfully argued to a Boston court that CPR for Barbara Howe would be "severe, invasive and harmful"); Kowalczyk, *Hospital, Family Spar*, *supra* note 66, at A1 ("'[T]his inhumane travesty has gone far enough . . . . This is the Massachusetts General Hospital, not Auschwitz.'" (quoting Dr. Edwin Cassem)).

68. *See generally* WRONG MEDICINE, *supra* note 37, at 94 ("[A]ttempted cardiopulmonary resuscitation could involve forceful, even violent, efforts at compressing the chest cage to the point of fracturing ribs . . . ."); Paul C. Sorum, *Limiting Cardiopulmonary Resuscitation*, 57 ALB. L. REV. 617, 617 (1994) ("The patient will usually receive the following interventions: manual compressions of the chest . . . ; one or more jolts of electricity to the chest . . .; and intravenous medications and fluids.").

69. *See* WRONG MEDICINE, *supra* note 37, at 100–01 ("[P]hysicians . . . should be *encouraged* or *required* to refrain from using futile treatments."); Capron, *supra* note 47, at 24 (unilateral termination can sometimes avoid "mistreating the patient").

70. *See, e.g.*, Annas, *supra* note 59, at 1543 (calling the provision of mechanical ventilation to Baby K after birth a "medical misjudgment" that gave the mother a false impression); Allan S. Brett, *Futility Revisited: Reflections on the Perspectives of Families, Physicians, and Institutions*, 17 HEC FORUM 276, 281-82 (2005) (discussing the "psychologically difficult conundrum for families" in futility cases). *But cf.* Fletcher, *supra* note 10, at S:224 (suggesting that the court documents in *Baby K* showed the physicians had certain reasons to support intubation).

deceptive to offer an option where none actually exists.[71] If health care providers offered ineffective treatment, they would risk losing public confidence.[72]

Lastly, providers resist inappropriate treatment requests in an effort to maximize the utility of scarce resources.[73] Providers want to be good "steward[s]"[74] of both "hard" resources like ICU beds and "soft" resources like health care dollars.[75] While costs have seldom been a consideration in defining when treatment is inappropriate,[76] there is little doubt that costs have been a major impetus for increasing attention on medical futility.[77] Thus, the issue of

---

71. *See* Howard Brody, *The Physician's Role in Determining Futility*, 42 J. AM. GERIATRICS SOC'Y 875, 876-77 (1994) (unethical to mislead patients by falsely raising hopes); Hudson, *supra* note 66, at 28 (quoting Dr. John Popovich's argument that "physicians who offer futile, meaningless care are charlatans"); Paris, *supra* note 46, at 150 (discussing how offering futile options gives false hope and unrealistic expectations to family members ultimately leading to "demands for more and more interventions and the risk of further complications"); Tomlinson & Czlonka, *supra* note 59, at 28, 30 (offering futile care is "a bogus choice, and the offer of it is a deception"; rather, providers should seek "acceptance" of a plan for a futility judgment rather than "consent").

72. *See* Brody, *supra* note 71, at 876-77 (discussing the importance of maintaining the medical profession's integrity).

73. *See* Rosenthal, *supra* note 66, at A1 ("Doctors and nurses . . . question whether futile resuscitations, which can costs thousands of dollars and tie up precious intensive care beds, make sense in an era of rising health costs."). *Cf.* WRONG MEDICINE, *supra* note 37, at 42 (treating 14,000 to 25,000 patients in a permanent vegetative state has estimated cost between $1 billion and $7 billion per year); Leonard M. Fleck, *Just Health Care Rationing: A Democratic Decisionmaking Approach*, 140 U. PA. L. REV. 1597, 1611 (1992) (estimating that Missouri spent nearly $1 million to keep Nancy Cruzan in a persistent vegetative state for eight years).

74. S.H. Miles, *Informed Demand for "Non-Beneficial" Medical Treatment*, 325 NEW ENG. J. MED. 512, 514 (1991).

75. *See infra* notes 188-97 and accompanying text.

76. *See infra* notes 188-97 and accompanying text. *But see* Murphy, *supra* note 43, at A37 (while Joseph Ndiyob's lack of health insurance and costs approaching $500,000 did not influence his attending physician's recommendation to stop treatment, the hospital's "medical futility review committee" did consider "whether the hospital should expend resources on a terminal patient rather than one who may recover").

77. *See* Rideout v. Hershey Med. Ctr., 30 Pa. D. & C.4th 57, 62 (Dauphin County Ct. C.P. Dec. 29, 1995) (No. 872S1995), 1995 WL 924561 (noting that a day after learning that patient's private health insurance was almost exhausted, hospital entered DNR order for patient); Nat'l Conference of Comm'rs on Uniform State Laws, Proceedings in Comm. of the Whole, Uniform Health-Care Decisions Act, Aug. 2, 1993, at 269-70 (statement of Comm'r King Hill) (noting that "medically ineffective" refers to costs). Mr. Hill explained:

> "This says to the physician that you don't have to institute some new radical $200,000 procedure if it's only going to keep the patient alive for two or three months, even though there may be many articles in the journals that say that's an accepted health-care standard for a [twenty-two] year old."

Nat'l Conference of Comm'rs on Uniform State Laws, Proceedings in Comm. of the Whole,

medically futile treatment is likely to increase in the future as concerns about costs for such treatment grows.[78]

Uniform Health-Care Decisions Act, Aug. 2, 1993, at 269-70 (statement of Comm'r King Hill); *see also* J.K. MASON & G.T. LAURIE, *Medical Futility*, *in* MASON AND MCCALL SMITH'S LAW AND MEDICAL ETHICS 539, 571-74 (7th ed. 2006) ("[T]he law clearly accepts that resource allocation forms a proper part of medical decision making."); Hudson, *supra* note 66, at 26 (noting that "economic losses for the hospital" motivated the futility of care policy at Santa Monica Hospital); Lantos, *supra* note 12, at 588-89 (discussing the "fundamental economic element" involved in futility determinations); Middleditch & Trotter, *supra* note 26, at 404 ("[T]he right to live may have less to do with societal conceptions of death or the legal doctrine of patient autonomy and more to do with money."); Rivin, *supra* note 30, at 389 (describing how the futile care policy developed directly from a review of the medical center's "financial losers"); Taylor & Lantos, *supra* note 49, at 7 ("We believe that the futility debate was more immediately motivated by changes in the way doctors and hospitals are paid."); Benjamin Weiser, *The Case of Baby Rena: Who Decides When Care is Futile?—Who Should Decide When Treatment is Futile? In Many Cases, Physicians Are Asking Whether Patient Autonomy Has Gone Too Far*, WASH. POST, July 14, 1991, at A19 ("It is not a coincidence that futility emerged as an issue in the mid-1980s only after the government limited hospital reimbursement for many patients."). Costs were similarly a motivation for moving from cardiopulmonary to neurological criteria for death. *See, e.g.*, Henry K. Beecher et al., *A Definition of Irreversible Coma, Report of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death*, 205 JAMA 337, 337 (1968) ("Our primary purpose is to define irreversible coma as a new criterion for death").

78. *See* MARK A. HALL, MARY ANNE BOBINSKI & DAVID ORENTLICHER, HEALTH CARE LAW AND ETHICS 3 (6th ed. 2003) ("[T]he Baby K situation may become more typical as a result of greater pressure on physicians to limit medical costs."); JOAN M. KRAUSKOPF ET AL., ELDERLAW: ADVOCACY FOR THE AGING § 13:26, at 500-01 (2d ed. 1993); THE RIGHT TO DIE, *supra* note 17, § 13.01[C] at 13-5, § 13.09 at 13-43; Donald J. Murphy, *The Economics of Futile Interventions*, *in* MEDICAL FUTILITY AND THE EVALUATION OF LIFE-SUSTAINING INTERVENTIONS 123, 133 (Marjorie B. Zucker & Howard D. Zucker eds., 1997) (arguing that the "economics of futile interventions deserves more study"); Ronald Bailey, *Pulling the Plug on Unwilling Patients: Should the High Cost of Living Affect Your Chances of Dying?*, REASON, Feb. 10, 2006, *available at* http://www.reason.com/news/printer/35016.html ("[I]t is clear that in the real world of limited medical resources that the 'authorities,' whether private or governmental, will unavoidably be making similar life and death decisions in the future."); Miran Epstein, *Legitimizing the Shameful: End-of-Life Ethics and the Political Economy of Death*, 21 BIOETHICS 23 (2007); Gampel, *supra* note 59, at 98 (predicting "managerial pressures on [health care providers] to use and extend the category of futility . . . ."); Kowalczyk, *Mortal Differences*, *supra* note 66, at A1 ("[H]ospitals will go to court more often to remove patients from life support, 'as health care becomes more of a scarce commodity . . . .'" (quoting law professor Charles Baron)); Wlazelek, *supra* note 46 ("'[B]ecause of the rising cost of health care, someone like the government or insurers will dictate that if you have X, Y, or Z you will not get the care.'" (quoting Joseph Vincent)); *cf.* CONGRESSIONAL BUDGET OFFICE, TECHNOLOGICAL CHANGE AND THE GROWTH OF HEALTH CARE SPENDING (Jan. 2008) (urging less and more cost-effective use of medical technology).

3. Limits on Resisting Treatment

Whatever might be their motivations for stopping LSMT, health care providers generally recognize two important limits on the extent to which they will resist a surrogate's request for LSMT: (1) comfort care and (2) accommodation. First, even when LSMT is stopped, providers will continue to administer comfort care.[79] They will continue to ensure the patient's comfort by providing services that include oral and body hygiene, reasonable efforts to offer food and fluids orally, medication, positioning, warmth, appropriate lighting, and other measures aimed at relieving pain and suffering or respecting the patient's dignity and humanity.[80] In short, stopping treatment does not mean stopping care.

Second, even when they consider continued LSMT to be inappropriate, providers will generally make a short-term accommodation of the surrogate's wishes.[81] Providers will respect patient treatment goals such as providing time to resolve personal matters, grieving, and allowing time to say goodbye.[82] Brain dead patients are oftentimes maintained on life support for several hours or days as a matter of sensitivity to religious, cultural, or moral values.[83]

---

79. *See, e.g.*, CONN. GEN. STAT. § 19a-573(a) (1993) ("Notwithstanding the [unilateral decision] provisions . . . , comfort care and pain alleviation shall be provided in all cases."); MINN. STAT. § 145B.13(1) (1991) (upon withdrawal of LSMT, there should be a "a continuation of appropriate care to maintain the patient's comfort, hygiene, and human dignity and to alleviate pain"); N.J. STAT. ANN. § 26:2H-67(b) (1991) (decision to forego LSMT does not impair care and comfort obligations); OR. REV. STAT. § 127.642 (2005) (care to provide comfort and cleanliness should be administered after withdrawal of LSMT); Pellegrino, *supra* note 10, at 309 ("Care [and] comfort . . . are never futile.").

80. *See, e.g.*, MINN. STAT. § 145B.13(1); OR. REV. STAT. § 127.642.

81. *See, e.g.*, Erich H. Loewy & Richard A. Carlson, *Futility and Its Wider Implications: A Concept in Need of Further Examination*, 153 ARCHIVES INTERNAL MED. 429, 429-30 (1993) (defending extending treatment for a reasonable time to allow family to come to terms with the situation, because while medically inappropriate, treatment may have social value).

82. *See* THE RIGHT TO DIE, *supra* note 17, § 13.08[A] at 13-40 ("[T]reatment might be rendered despite its certain or probable lack of medical benefit occurs when the patient or family has personal, 'non-medical' reasons for wanting the treatment . . . ."); WRONG MEDICINE, *supra* note 37, at 166; Carhart, *supra* note 51, at 1755 ("'[W]hy not just leave the machines on for two weeks?'" (quoting health law attorney Shirley Paine)); Fletcher, *supra* note 10, at S:236 (arguing that physicians should be permitted to discontinue treatment "after a grace period of adjustment"); Pellegrino, *supra* note 10, at 315-16 (urging a "permissive" rather than an "overly rigorous" application of futility because the family needs "time to adjust" and a patient might like to see "a grandchild born, or have a last meeting with family or friends"); Skene, *supra* note 62, at 701 (arguing for the "broader aspect of patients' 'best interests'"); Tomlinson & Czlonka, *supra* note 59, at 29 (providers must consider "nonbiomedical goals"); David M. Zientek, *The Texas Advance Directives Act of 1999: An Exercise in Futility?*, 17 HEC FORUM 245, 253 (2005) (urging certain goals to be respected, such as "support[ing] life until a child overseas in the military can return home for a last visit" or "continu[ing] life support to allow for spiritual preparation for death").

83. *See* Dority v. Superior Court of San Bernardino County, 193 Cal. Rptr. 288, 289 (Cal.

D. The Resolution of Futility Disputes

The disagreement between surrogates and providers regarding continued LSMT produces a significant number of futility disputes each year.[84] Fortunately, the vast majority of these disputes are resolved internally and informally through good communication and mediation practices.[85] The standard dispute resolution process consists of six roughly chronological stages.[86] Most futility disputes are resolved within the first five stages.[87]

Ct. App. 1983) (describing hospital policy of keeping brain dead children on life support "until the parents were emotionally able to realize what the medical opinion was"); Lorry R. Frankel & Chester J. Randle Jr., *Complexities in the Management of a Brain-Dead Child, in* ETHICAL DILEMMAS IN PEDIATRICS: CASES AND COMMENTARIES 135, 137 (Lorry R. Frankel et al. eds., 2005) ("On rare occasions, life support will be continued for a few more hours, pending arrival of other family members."); Myra J. Edens et al., *Neonatal Ethics: Development of a Consultative Group*, 86 PEDIATRICS 944, 947 (1990) ("[T]reatment is continued for a period of time to allow the parents to come to terms with the hopelessness of [the] . . . condition."); Rasa Gustaitis, *Right to Refuse Life-Sustaining Treatment*, 81 PEDIATRICS 317, 319 (1988) ("[C]hildren have not infrequently been kept alive on life-support equipment for the sake of others . . . ."); George J. Annas, *When Death Is Not the End*, N.Y. TIMES, Mar. 2, 1996, at 19 ("Maintaining a corpse in an intensive care unit for a few days may be reasonable as a matter of sensitivity to religious or moral beliefs . . . ."); Yeates, *supra* note 44 (reporting that parents tried to maintain the life support of their six-year old boy after he was declared dead by the doctors). In some jurisdictions this is required by statute or regulation. *See, e.g.*, N.J. STAT. ANN. § 26:6A-5 (1991) (exemption to accommodate patient or family's religious beliefs); N.Y. COMP. CODES R. & REGS. tit. 10, § 400.16(e)(3) (1987) (allows for accommodations for an "individual's religious or moral objection" to determinations of death).

84. One study found 974 futility disputes in sixteen hospitals over an average four-year period. *See* Emily Ramshaw, *Bills Challenge Care Limits for Terminal Patients: Some Say 10 Days to Transfer Isn't Enough Before Treatment Ends*, DALLAS MORNING NEWS, Feb. 15, 2007. According to the American Hospital Association, there are 5,700 hospitals in the United States. American Hospital Association, Fast Facts on US Hospitals, at 1 (2007), http://www.aha.org/aha/content/2007/pdf/fastfacts2007.pdf. If the study's sample is representative, then that rate of fifteen futility disputes per hospital per year means that there are tens of thousands of futility disputes nationwide. However, there is reason to think that this sample is not representative. One reason is that the sample is from Texas, a state where physicians became more willing to resist inappropriate treatment requests after enactment of an effective statutory safe harbor. *See* Robert L. Fine & Thomas Wm. Mayo, *Resolution of Futility by Due Process: Early Experience with the Texas Advance Directives Act*, 138 ANNALS INTERNAL MED. 743, 745 (2003) [hereinafter Fine & Mayo] (upon passage of the statutory safe harbors, futility consultations increased 67%); *see also infra* Parts V.C, VII.B (discussing the Texas Advance Directives Act).

85. *See infra* notes 89-95 and accompanying text.

86. These stages track the process recommended by the AMA and endorsed by most regional and facility policies. *See* AMA Council, *supra* note 53, at 939 (discussing the steps of fair process in futility cases).

87. *See infra* note 110 and accompanying text (discussing that few cases ever reach the final stage of the process and thus, are presumably resolved in one of the previous stages).

Nevertheless, a small but significant number of cases do proceed to the sixth and final stage, where the provider must unilaterally decide whether to stop treatment.[88]

Stage One: Ensure Good Communication by the Health Care Team. It is best to avoid a futility dispute in the first place through careful communication—clarifying the goals of treatment, its possible outcomes, and the patient's values and wishes.[89] Many commentators argue that much more can and should be done in this respect.[90] Nevertheless, through education and persuasion, the surrogate and the provider usually reach agreement.[91] Most disputes are avoided or resolved at this stage.[92]

Stage Two: Bring in a Consultant. If the health care team is unable to convince the surrogate to end LSMT, then the team typically employs an individual consultant or mediator to negotiate an agreement between the physician and patient.[93] Professor Nancy Dubler explains that a bioethics mediator "facilitates a discussion between and among the parties to the

---

88. *See infra* notes 111-14 and accompanying text.

89. *See* Chad Bowman, *Disputes Over End-of-Life Care Treated Increasingly with Mediation*, 9 BNA HEALTH L. REP. 1527, 1527 (2000) ("'If you communicate well enough, often enough, and clearly enough, you will not have futility issues.'" (quoting attorney Shirley J. Paine)); Ursula Braun et al., *Defining Limits in Care of Terminally Ill Patients*, 334 BMJ 239, 239 (2007) ("Doctors should make clear that good medical care does not always mean doing everything that is technically possible . . . ."); Fine & Mayo, *supra* note 84, at 745 ("Most end-of-life consultations ease the transition from curative to a palliative model of care and occur in the absence of any particular conflict between parties."); Stanley A. Nasraway, *Unilateral Withdrawal of Life-Sustaining Therapy: Is It Time? Are We Ready?*, 29 CRITICAL CARE MED. 215, 217 (2001) (recommending "preemptive actions" to prevent conflicts from taking place).

90. For example, some commentators recommend that health care providers should not offer non-indicated options because the family will feel guilty if they do not do everything. *See supra* notes 46-50, 70-71 and accompanying text. Alternatively, providers should offer inappropriate options only as a time-limited trial to be stopped if unsuccessful. *See, e.g.*, Tovino & Winslade, *supra* note 41, at 52-53.

91. Lantos, *supra* note 12, at 589 ("Generally, in such situations, doctors explain [the situation] to the patients or their surrogates, the latter understand and accept the situation, and treatment is withheld or withdrawn."). Of course, some disputes may be resolved not only though persuasion but also through manipulation and coercion. *Cf.* THE RIGHT TO DIE, *supra* note 17, § 13.09 at 13-41 ("Some (perhaps most) futility cases can be resolved at the bedside, without the necessity of litigation, by acquiescence of one of the parties to the view of the other . . . .").

92. *See* Robert L. Fine, *The Texas Advance Directives Act of 1999: Politics and Reality*, 13 HEC FORUM 59, 71-72 (2001) (reporting that "within a day or two of learning of the [dispute resolution] process," families often agree to substitute comfort care in place of LSMT); Giles R. Scofield, *Medical Futility: Can We Talk?*, 18 GENERATIONS 66, 67 (1994) (reporting evidence that 94% of patients agree with their physician's recommendation to not attempt LSMT); Tomlinson & Czlonka, *supra* note 59, at 34 ("[A]lmost all cases are resolved at this [first] stage.").

93. *See generally* Pope & Waldman, *supra* note 9, at 155-58 (reviewing the relevant literature on mediation in futility disputes).

conflict[,]" helping the parties "to identify their goals and priorities and to generate, explore, and exchange information and options."[94] For many futility disputes, "mediation can provide a process to assist in the formation of a care plan that meets the needs of the patient and family and respects professional commitments."[95]

Stage Three: Go to the Hospital Ethics Committee. If the provider and surrogate still disagree about the appropriate treatment for the patient, the provider will typically ask the institutional ethics committee to intervene.[96] The committee usually, though not always, agrees with the treating physician's recommendation to stop LSMT.[97]

Upon receiving the committee's decision, the surrogate may agree to terminate care.[98] This acquiescence might stem from the passage of additional time and the opportunity for more careful deliberation, making the surrogate feel more secure about such a decision.[99] Moreover, if the ethics committee indicates that it will authorize the unilateral withdrawal of treatment, the surrogate may likely feel relieved from the burden of that decision.[100]

Stage Four: Change the Decision Maker. In some cases, the health care provider may doubt that the surrogate's decision reflects the patient's actual

---

94. Dubler, *supra* note 26, at S24-S25.

95. *Id.* at S25.

96. *See* Hudson, *supra* note 66, at 26 ("The bioethics committee gets involved in about 2 percent of cases . . . because by the time an ethics committee conference is scheduled, the issue has often been resolved . . . or the patient dies before the conference is held.").

97. *See, e.g.*, Fine & Mayo, *supra* note 84, at 745 tbl.3 (reporting that one hospital's ethics committee agreed with the attending physician 90% of the time).

98. Zientek, *supra* note 82, at 250. Doctor Zientek reported that "[o]f the 43 cases deemed futile, in 37 cases the family agreed to withdrawal of treatment, while in six cases they refused to accept withdrawal. Of these six cases, the families of three agreed to shifting to comforting measures a 'few days' after receiving the committee's formal report." *Id.*; *see* Fine & Mayo, *supra* note 84, at 745 (reporting that family decision makers accepted the committee's judgment 86% of the time); Belluck, *supra* note 38, at A1 ("Ethics committees resolve most cases, often through repeated family discussions over weeks or months.").

99. *See* Robert D. Truog & Christine Mitchell, *Futility—From Hospital Policies to State Laws*, 6 AM. J. BIOETHICS 19, 20 (2006).

100. *See* Robert L. Fine et al., *Medical Futility in the Neonatal Intensive Care Unit: Hope for a Resolution*, 116 PEDIATRICS 1219, 1221 (2005) ("[T]he family was relieved because they had 'put up the good fight' . . . but now the decision was out of their hands."); Fine & Mayo, *supra* note 84, at 745 ("'If you are asking us to agree with the recommendation to remove life support from our loved one, we cannot. However, . . . if the law says it is OK to stop life support, then that is what should happen.'"); Lantos, *supra* note 12, at 589 ("The concept of futility . . . has a moral role in helping absolve patients or surrogates of the moral obligation to continue treatment."); Hoffman, *supra* note 46, at F1 ("Families often believe that consenting to a D.N.R. order implies they are giving up on their loved one, signing a death warrant, turning their backs on hope."); Wlazelek, *supra* note 46 ("'It's dangerous to give the family the last word since guilt and a desire to do everything for pop makes it emotionally impossible to stop treatment.'" (quoting Arthur Caplan)).

preferences or best interests.[101] Under these circumstances, providers may try to switch the legally authorized decision maker to one that will agree with their recommendation to cease LSMT.[102] One strategy providers sometimes employ to make the switch is to argue that LSMT constitutes abuse or neglect where it primarily imposes burdens such as pain.[103] That is a difficult task because the provider is usually not questioning whether the surrogate's decisions truly reflect the patient's preferences or whether the surrogate is acting in the patient's best interests.[104] Rather, the provider is just disagreeing with the decision maker's determination.[105]

---

101. *See infra* notes 102-05 and accompanying text.

102. *See In re* Baby K, 832 F. Supp. 1022, 1031 (E.D. Va. 1993) (remarking that mother's treatment decision need not be respected if it "would constitute abuse or neglect"); Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1076 n.3 (La. Ct. App. 1998). The *Causey* court noted that if a surrogate insists on inappropriate treatment, "the usual procedure . . . is to transfer the patient or go to court to replace the surrogate or override his decision. The argument would be that the guardian or surrogate is guilty of abuse by insisting on care which is inhumane [or that the surrogate is not fulfilling their statutorily provided role]." *Causey*, 719 So. 2d at 1076 n.3.

103. *See, e.g.*, *Baby K*, 832 F. Supp. at 1031 (discussing whether continuing LSMT constituted abuse); *In re* Doe, 418 S.E.2d 3, 6-7 (Ga. 1992) (discussing but declining to decide whether LSMT constituted "medical abuse"); *see also* Gustaitis, *supra* note 83, at 318-19 (suggesting use of child abuse laws to override parental requests for inappropriate treatment).

104. *See, e.g.*, *In re* Howe, No. 03 P 1255, 2004 WL 1446057, at *3, *21 (Mass. Prob. & Fam. Ct. Dept. Mar. 22, 2004) (refusing Massachusetts General Hospital's request to replace Barbara Howe's daughter as her proxy); *State of Minnesota District Court –Probate Court Division County of Hennepin Fourth Judicial District*, 7 Issues L. & Med. 369, 372 (1991) (discussing *In re Wanglie,* where the court denied the hospital's request for conservator because the patient's husband was the appropriate person to articulate her wishes); Weiser, *supra* note 45, at A18 (reporting how the district court rejected the hospital's attempt to replace a mother who was demanding LSMT with a court-appointed guardian); *cf. In re* Guardianship of Schiavo, No. 90-2908GD-003, 2000 WL 34546715, at *7 (Fla. Cir. Ct. Feb. 11, 2000) (denying Theresa Schiavo's parents' motions to transfer guardianship from her husband). *But see In re* Guardianship of Mason, 669 N.E.2d 1081, 1085-87 (Mass. App. 1996) (affirming probate court's entry of DNR order and overriding patient's son's decision because he was in "denial about the deterioration [of] his mother"); Bopp & Coleson, *supra* note 45, at 825-26 ("Baby Terry's parents had 'specific incompetence' to choose Baby Terry's medical treatment.") (citing Brief of Appellant at 5, *In re* Achtabowski, No. 93-1247-AV (Mich. Cir. Ct. Aug. 12, 1993) (No. 170251)). Judicial hostility to surrogate shopping in these cases does seem to be waning. *See* Thaddeus Mason Pope, *Reassessing the Judicial Treatment of Futility Cases*, 9 MARQ. ELDER'S ADVISOR (forthcoming 2008), *available at* http://ssrn.com/abstract=1078983.

105. *See* Robert L. Schwartz, *Autonomy, Futility, and the Limits of Medicine*, 1 CAMBRIDGE Q. HEALTHCARE ETHICS 159, 161 (1992) (arguing that whether Mr. Wanglie was his wife's best substitute decision maker was the wrong question). Professor Schwartz posits that "[t]he real question . . . [should have been] whether the continuation of ventilator support and gastrostomy feeding were among the reasonable medical alternatives that should have been available to Mrs. Wanglie or her surrogate decision maker, whoever that might be." *Id.* at 161-62.

Frequently, surrogate decision makers are often replaced in child abuse cases where the parent is the alleged abuser.[106] In such cases, it is naturally assumed that the parent would *not* be acting in the best interest of the child by insisting on continued LSMT. This assumption arises particularly where the child's death could result in murder charges against the parent. There is no such clarity in the typical futility case.

Stage Five: Attempt Transfer. If the surrogate cannot be replaced and the provider and surrogate still do not agree, then the health care provider should do one of the following: (1) find a new provider or (2) attempt to transfer the patient to another institution willing to comply with the surrogate's treatment requests.[107] While this is rarely successful, it does sometimes resolve a few additional disputes.[108]

Stage Six: Implement the Unilateral Decision to Stop Treatment. Only after diligently making all of the foregoing attempts to resolve the conflict should a provider take unilateral action to stop LSMT against the wishes of the patient or surrogate.[109]

---

106. *See, e.g.*, Tabatha R. v. Ronda R., 564 N.W.2d 598, 602, 605 (Neb. 1997). In that case, the Department of Social Services took temporary custody of an infant in a persistent vegetative state and requested withdrawal of LSMT over parents' objections; consequently, the court ruled that parental rights must first be terminated since this would result in the death of the child. *Id.*; Pam Belluck, *Custody and Abuse Cases Swirl Around a Troubled Girl on Life Support*, N.Y. TIMES, Dec. 6, 2005, at A18 (reporting Massachusetts juvenile court granted DSS request to remove life support from child in their custody against the wishes of child's adoptive parents); *Clackamas County Judge to Rule on Brain-Damaged Baby*, COLUMBIAN, Apr. 24, 2004, at C8 (reporting state advocate for brain damaged baby took custody of child and requested juvenile court to grant a DNR order); *see also* Child & Family Servs. of Cent. Manitoba v. R.L., 123 Man. R. (2d) 135, 154 D.L.R. (4th) 409 (Man. App. 1997) (allowing the providers to enter a DNR at the direction of Child & Family Services over the parents' objections).

107. Most institutional and professional association model futility policies provide for transfer. *See* sources cited *supra* note 53. This is consistent with the law of tortuous abandonment, which requires that physicians assist their patients in finding a new provider before terminating a treatment relationship. *See, e.g.*, Payton v. Weaver, 182 Cal. Rptr. 225, 227 (Cal. Ct. App. 1982) (dealing with the problem of a disruptive dialysis patient and the lack of accepting institutions); Stella L. Smetanka, *Who Will Protect the 'Disruptive' Dialysis Patient?*, 32 AM. J.L. & MED. 53, 71-79 (2006) (discussion of cases and "no duty to treat"). Transfer is also required by most state health care decision making statutes. *See* sources cited *infra* note 369.

108. *See infra* notes 339-52, 369 and accompanying text.

109. *See* MICHAEL D. CANTOR ET AL., NATIONAL CENTER FOR ETHICS IN HEALTH CARE, DO-NOT-RESUSCITATE ORDERS AND MEDICAL FUTILITY: A REPORT BY THE NATIONAL ETHICS COMMITTEE OF THE VETERANS HEALTH ADMINISTRATION 1, 8 (2000) [hereinafter VHA-NEC REPORT] (arguing that unilateral decisions "should be reserved for exceptionally rare and extreme circumstances after thorough attempts" to resolve disagreements have failed); THE RIGHT TO DIE, *supra* note 17, § 13.04[B] at 13-22 ("[S]ometimes only litigation can break the impasse between demanding families and resistant health care professionals."); Timothy Bowen & Andrew Saxton, *New Developments in the Law—Withholding and Withdrawal of Medical*

While most cases will never reach this stage,[110] a significant percentage will.[111] One recent five-year study of sixteen hospitals found that in approximately sixty-five cases, the hospitals decided to unilaterally stop LSMT.[112] Another study of nine hospitals found that they decided to unilaterally stop LSMT in 2% of 2,842 cases.[113] Furthermore, there are strong reasons to suspect that the rate of intractability and unilateral hospital action will rise.[114]

II. LEADING DEFINITIONS OF "MEDICAL INAPPROPRIATENESS"

"Medical inappropriateness" is a term with a contentious history because commentators argue it has different meanings in different contexts. While there is a consensus that LSMT is inappropriate where the patient is brain dead or where the requested treatment simply will not work (i.e., physiological futility), these definitions cover only a tiny fraction of the relevant cases.[115] In most disputes, providers employ a notion of quantitative or qualitative futility, considering either the likelihood that the treatment will succeed or the quality of life that it can provide the patient.[116] These definitions of medical inappropriateness, however, are value-laden determinations, lacking consensus support from the medical community, the bioethical community, and the public.[117]

Treatment, 14 AUSTL. HEALTH L. BULL. 57, 60 (2006).

110. *See* Brennan, *supra* note 30, at 807 ("In all cases [where unilateral DNR orders were entered], the families either ultimately accepted this reasoning or ceased insisting that invasive procedures be used.").

111. *See* Pope & Waldman, *supra* note 9, at 158-61; *see also* Fine, *supra* note 48, at 79 (noting that five of twenty-nine cases went through the whole process, although two died and three agreed to withdraw before treatment was unilaterally stopped); Daniel Garros et al., *Circumstances Surrounding End of Life in a Pediatric Intensive Care Unit*, 112 PEDIATRICS 1171, 1173 (2003) (in 1 out of 68 cases, no complete agreement could be reached between the surrogates and providers).

112. Ramshaw, *supra* note 84. About half of the patients in the study died or were transferred to other facilities before treatment was actually stopped. *Id.*

113. Tex. H.R. Comm. on Pub. Health, 80th Leg., *Interim Report*, at 36 (2006) [hereinafter Interim Report] (citing written testimony of Greg Hooser).

114. The reasons for surrogate insistence are becoming more prevalent. *See supra* notes 37-50 and accompanying text. At the same, provider resistance may increase with changes in reimbursement and an increased focus on palliative care.

115. *See infra* notes 118-56 and accompanying text.

116. *See infra* notes 157-66, 175-77 and accompanying text.

117. *See infra* notes 167-73, 178-211 and accompanying text.

A. Brain Death

Perhaps the clearest case of medically inappropriate care is LSMT requested for a brain dead patient.[118] Since the 1950s, health care providers have been able to artificially maintain respiration and circulation even for a patient whose brain had completely and irreversibly ceased to function.[119] In light of this possibility to maintain breathing and a heart beat with technology, the previously accepted standard for determining death—the cessation of cardiopulmonary function—was too limited.[120] Consequently, every state soon adopted the cessation of all brain function as an alternative method for determining death.[121]

There is a consensus that it is ethically, legally, and medically appropriate to stop LSMT for a brain dead patient.[122] The adoption of the Uniform Determination of Death Act has "alleviate[d] concern among medical practitioners that legal liability might be imposed" for stopping LSMT for a brain dead patient.[123] Indeed, defining a patient as dead provides such legal clarity that many have argued for broadening the statutory standards for the determination of death.[124]

---

118. *See* David C. Blake, *Bioethics and the Law: The Case of Helga Wanglie: A Clash at the Bedside—Medically Futile Treatment v. Patient Autonomy*, 14 WHITTIER L. REV. 117, 126 (1993).

119. *See* James L. Bernat, *The Whole-Brain Concept of Death Remains Optimum Public Policy*, 34 J.L. MED. & ETHICS 35, 35 (2006).

120. *See id.*

121. *Id.* at 36; Kirsten Rabe Smolensky, *Defining Life from the Perspective of Death: An Introduction to the Forced Symmetry Approach*, 2006 U. CHI. LEGAL F. 39, 43-48 (2006).

122. *See, e.g.*, Gallups v. Cotter, 534 So. 2d 585, 589 (Ala. 1988) (affirming summary judgment for defendants on outrage claim); Dority v. Superior Court of San Bernardino County, 193 Cal. Rptr. 288, 290-91 (Cal. Ct. App. 1983); Cavagnaro v. Hanover Ins. Co., 565 A.2d 728, 731 (N.J. Super. Ct. Law Div. 1989) (finding insurer need not pay medical and hospital expenses after brain death because not incurred for treatment); *In re* Long Island Jewish Med. Ctr., 641 N.Y.S.2d 989, 992 (N.Y. Sup. Ct. 1996) (holding that hospital can withdraw LSMT from brain dead child over parent's objections); Alvarado v. N.Y. City Health & Hosps. Corp., 547 N.Y.S.2d 190 (N.Y. Sup. Ct. 1989), *vacated by* 550 N.Y.S.2d 353, 354 (N.Y. App. Div. 1990) (finding condition of infant did not constitute brain death as defined by statute); Marshall B. Kapp, *Legal Liability Anxieties in the ICU*, *in* MANAGING DEATH, *supra* note 15, at 234.

123. *In re* Bowman, 617 P.2d 731, 738 (Wash. 1980) (citing Uniform Brain Death Act § 1, 12 U.L.A. (Supp 1980)).

124. Some have proposed extending the definition to include patients in a permanent vegetative state and anencephalic infants. *See, e.g.*, E. Haavi Morreim, *Futilitarianism, Exoticare, and Coerced Altruism*, 25 SETON HALL L. REV. 883, 886 n.11, 888 n.22 (1995). *But see* Alexander Morgan Capron, *Anencephalic Donors: Separate the Dead from the Dying*, 17 HASTINGS CTR. REP., Feb. 1987, at 5 ("It would be unwise to amend the Uniform Determination of Death Act to classify anancephalics as 'dead.'"); David T. McDowell, Note, *Death of an Idea: The Anencephalic as an Organ Donor*, 72 TEX. L. REV. 893, 930 (1993) (arguing that society would be "worse off" if the legal definition of death were extended to include the

B. Physiological Futility

Apart from brain death, the narrowest and perhaps most clearly defined definition of medically inappropriate care is referred to as "physiological futility."[125] Physiologically futile interventions are inappropriate because they do not produce a measurable effect on the patient.[126] In essence, the requested treatment has a zero percent chance of being effective.[127]

Physiological futility is true to the etymological origins of the term "futility."[128] The Latin word *futilis* refers to "actions or instruments which were inherently leaky and therefore ill-suited for achieving [their] desired ends."[129] The classic illustration of *futilis* comes from Greek mythology; the daughters of King Danaus were condemned to Hades and forced to draw water in leaky containers.[130] Because of the leaks, the daughters could not achieve the goals of their actions.[131]

Commentators have offered a multitude of colorful examples of physiological futility, including the following:[132] (1) prescribing laetrile or

---

anencephalic). As Roger Dworkin notes, "Definition is dangerous because it allows us to avoid analysis and do bad things to persons without concern by defining them out of existence." ROGER B. DWORKIN, LIMITS: THE ROLE OF THE LAW IN BIOETHICAL DECISION MAKING 112 (1996).

125. *See, e.g.*, Fletcher, *supra* note 10, at S:232 (discussing the narrow meaning of treatment that is "physiologically ineffective"); Dale L. Moore, *Challenging Parental Decisions to Overtreat Children*, 5 HEALTH MATRIX: J. L.-MED. 311, 315-16 (1995) (briefly explaining the concept of physiologically futile treatment).

126. Moore, *supra* note 125, at 315-16. Sometimes this can be known *ex ante* as a matter of science. Other times, physiological futility cannot be determined until after one or more failed attempts with a specific patient.

127. *Id.* at 316 ("[A] clear example of treatment that is 'futile' in the 'physiologically futile' sense: it simply did not (and was not destined to) work.")

128. *See* SUSAN RUBIN, WHEN DOCTORS SAY NO: THE BATTLEGROUND OF MEDICAL FUTILITY 42 (1998).

129. *Id.*

130. *See* PLATO, *supra* note 52, at 59 ("These they bury in the mud of Hades; some are also compelled to fetch water in a sieve.")

131. *See id.* This assumes that the leaks were so substantial that all the water drained out between the river Styx and the destination. If the leaks were slower such that not all of the water was drained, then the daughters could have achieved their goal, at least to some degree. This situation would be analogous to qualitative futility. *See infra* Part II.D.

132. *See, e.g.*, Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1074 (La. Ct. App. 1998). The court argued that "[t]he problem is not with care that the physician believes is harmful or literally has no effect. For example, radiation treatment for Mrs. Causey's condition would not have been appropriate. This is arguably based on medical science." *Id.*; FINS, *supra* note 33, at 79-80 (offering examples such as "infus[ing] septic patients with fluids and pressors to hold a blood pressure[,] . . . . intubation in a patient with an obstructing tracheal mass[, or] . . . . 'call[ing] the code' . . . [i]f one can not get a rhythm or bring the pH up to normal range"); Moore, *supra* note 125, at 315-16 (CPR on patient with renal failure who had not had dialysis); Morreim, *supra* note 124, at 894, 896 (offering examples where disability would render the

pasque-flower tea for cancer,[133] (2) prescribing antibiotics for a viral illness,[134] (3) performing a heart transplant for a patient dying of liver failure,[135] (4) performing CPR in the presence of cardiac rupture or severe outflow obstruction,[136] (5) offering chemotherapy for an ulcer,[137] (6) giving a penicillin shot for a head cold,[138] (7) performing an appendectomy to calm a patient's fears that they may have appendicitis,[139] and (8) treating the dead with mechanical ventilators and pressors.[140]

With physiological futility, the provider does not make any assessment that the effect is unlikely, too small, or not worthwhile.[141] The provider does not characterize whether the effect is a "benefit" or not.[142] Instead, health care providers can readily ascertain physiological futility based solely upon their clinical knowledge.[143] Thus, there is no room for normative disagreement.[144]

---

treatment "utterly pointless").

133. Fletcher, *supra* note 10, at S:232; Schwartz, *supra* note 105, at 160.

134. Levine, *supra* note 10, at 74; *see also* Robert M. Veatch & Carol M. Spicer, *Medically Futile Care: The Role of the Physician in Setting Limits*, 18 AM. J. L. & MED. 15, 18 (1992) (prescribing antibiotics for the common cold).

135. Taylor & Lantos, *supra* note 49, at 4.

136. Fletcher, *supra* note 10, at S:232. A similar example entails a blood transfusion where the recipient is hemorrhaging at a rate that exceeds the maximum rate of transfusion. Levine, *supra* note 10, at 74; *see also* American Heart Association, *2005 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care—Part 2: Ethical Issue*, 112 CIRCULATION IV-6, IV-7 (2005), *available at* http://circ.ahajournals.org/cgi/content/full/112/24_supp/IV-6 ("[A]ll patients in cardiac arrest should receive resuscitation unless . . . "[n]o physiological benefit can be expected because vital functions have deteriorated despite maximal therapy (e.g., progressive septic or cardiogenic shock)."); Veatch & Spicer, *supra* note 134, at 18 (CPR is physiologically futile where performed on a patient who last breathed three hours prior to administering the care).

137. Wesley J. Smith, *Death by Ethics Committee: Refusing to Treat Lives Deemed Unworthy of Living*, NAT'L REV., Apr. 27, 2006, *available at* http://www.nationalreview.com/smithw/smith200604271406.asp.

138. Marcia Angell, *The Case of Helga Wanglie: A New Kind of "Right to Die" Case*, 325 NEW ENG. J. MED. 511, 512 (1991).

139. THE RIGHT TO DIE, *supra* note 17, § 13.07[B] at 13-38; *see also* WRONG MEDICINE, *supra* note 37, at 157 ("Nor is a surgeon obligated to perform a prophylactic appendectomy to assuage a patient's fears that her recurrent abdominal pains are due to appendicitis.").

140. FINS, *supra* note 33, at 79-80 (offering examples such as "infus[ing] septic patients with fluids and pressors to hold a blood pressure").

141. *See* Gampel, *supra* note 59, at 96 (contrasting refusals to provide physiologically futile treatment with refusals because the treatment is "inappropriate . . . [and] the risks outweigh the potential benefits, or because the patient's request is irrational or ill-considered given the low odds or limited benefit involved").

142. *See id.*

143. *See* Levine, *supra* note 10, at 79 ("Characterizing a treatment as 'useless' based on the extremely low chance that a physiological effect will occur requires an opinion that this low probability is not worth pursuing, *not a scientific determination* that the physiological effect sought is scientifically impossible.") (emphasis added).

The basis for refusing treatment is an empirical one: the treatment simply will not work.[145] Even the biggest opponents of unilateral decision making concede that "[r]efusals of requests for such 'physiologically futile care' would be proper and professional."[146]

However, this objectivity comes at a steep price because physiological futility has a very limited applicability.[147] First, the vast majority of cases are not as clear-cut as those described in the previous four paragraphs. Decisions on withholding and withdrawing treatment are usually based on mere probabilities as opposed to certainties.[148] Most providers find it difficult to be certain that there is a 100% probability that any given intervention will have zero effect.[149]

Second, physiological futility has limited applicability because it is too demanding, requiring the absence of an "effect" on any part of the patient's anatomy, physiology, or chemistry.[150] Because technology permits many "effects," such as keeping a heart beating, obtaining true physiological futility rarely occurs.[151] One must be careful to distinguish between physiological

144. Because physiological futility is so much more easily justified, hospitals often attempt to characterize (or mask) the care that they seek to unilaterally withdraw as physiologically futile. *Cf.* Gampel, *supra* note 59, at 96. For example, Baylor Hospital argued that mechanical ventilation for Tirhas Habtegiris was "medically inappropriate, on scientific grounds alone." Yet, Baylor conceded that it would "keep [the] suffering patient alive." Baylor Health Care System, Tirhas Habtegiris Case: Baylor Response, http://www.baylorhealth.com/articles/habtegiris/response.htm (last visited Oct. 15, 2007) [hereinafter Baylor Response]. More convincingly, Massachusetts General Hospital argued this theory of medical inappropriateness to the jury in the *Gilgunn* case. Capron, *supra* note 47, at 26 (noting the defendants argued that "CPR 'could not produce the desired physiological change' . . . [and] would not only be ineffective but would be harmful").

145. *See* Moore, *supra* note 125, at 315-16.

146. *See, e.g.*, Smith, *supra* note 137.

147. *See* FINS, *supra* note 33, at 80-81 ("[T]he narrowness of the physiologic definition is also its greatest weakness . . . ."); Bowman, *supra* note 89, at 1527 ("'With the exception of a small number of cases, it's not possible to say with certainty that care will provide no benefits.'" (quoting Dr. Gregg Bloche)); Brett, *supra* note 70, at 293 ("[T]he vast majority of contentious cases do not involve physiologic futility."); Gampel, *supra* note 59, at 94 ("[I]f clinical certainty of a zero chance of success were required, there would be little if any room for the use of the concept of futility in medical practice."); Levine, *supra* note 10, at 82 n.92 ("Treatment is strictly physiologically futile only when it is *certain* that the physiological effect sought from the treatment cannot be achieved").

148. *See* WRONG MEDICINE, *supra* note 37, at 14, 97, 136 (discussing the inaccuracies of assessing "quantitative probabilities" in health care); Bowen & Saxton, *supra* note 109, at 59 (noting that published guidelines for the withdrawal of LSMT contemplate probabilities).

149. *See* WRONG MEDICINE, *supra* note 37, at 14 ("[O]ne can never be absolutely certain of the outcome.").

150. *See* FINS, *supra* note 33, at 80 ("[T]he physiologic definition is the narrowest definition of medical futility. It is a clinical determination based on narrow physiologic parameters.").

151. *See id.* ("A physiologic definition simply asks whether the infection could be resolved

futility, where there is no effect, and the more typical situation of qualitative futility, where there is some effect, albeit one judged to offer no meaningful "benefit."[152] Therefore, physiological futility is a narrow category covering few cases.[153]

While many states explicitly permit the unilateral termination of physiologically futile interventions, no state with a unilateral decision statute relies *solely* on a physiological futility standard of medical appropriateness.[154] New York's standard most closely resembles this idea, articulating that "'[m]edically futile' means that cardiopulmonary resuscitation will be unsuccessful in restoring cardiac and respiratory function or that the patient will experience repeated arrest in a short time period before death occurs."[155] Yet, even the language of this statute recognizes that the CPR might work, just not for a sufficient time to be considered worthwhile.[156]

---

with antibiotics. If so, the treatment is not physiologically futile, even though the "restoration" of health will be a pre-morbid state of severe cognitive impairment.").

152. *Cf. id.* at 80-81 ("[P]atients are more than their physiology.").

153. *See, e.g.*, Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1074 (La. Ct. App. 1998) ("Strictly speaking, if a physician can keep the patient alive, such care is not medically or physiologically 'futile;' however, it may be 'futile' on philosophical, religious, or practical grounds.").

154. For example, Maryland allows providers to refuse "medically ineffective treatment." MD. CODE ANN., HEALTH–GEN. § 5-611(b)(1) (West 2007). But even Maryland makes clear that this is not limited to physiological futility but also includes medical procedures that, to a reasonable degree of medical certainty, will not do the following: "(1) Prevent or reduce the deterioration of the health of an individual; or (2) Prevent the impending death of an individual." MD. CODE ANN., HEALTH–GEN. § 5-601(n) (West 2007); *see also* GA. CODE ANN. § 31-39-2(4) (2006) ("'Candidate for nonresuscitation' means a patient who . . . (C) Is a person for whom cardiopulmonary resuscitation would be medically futile in that such resuscitation will likely be unsuccessful in restoring cardiac and respiratory function or will only restore cardiac and respiratory function for a brief period of time . . . ."); OKLA. STAT. ANN. tit. 63, § 3131.4(C)(2) (West 1999) (providing that CPR is not required where it would not prevent imminent death); OR. REV. STAT. ANN. § 127.580(b) (West 2005) ("Administration of such nutrition and hydration is not medically feasible or would itself cause severe, intractable or long-lasting pain."); OR. REV. STAT. ANN. § 127.635(c) (West 2005) (LSMT is not required where it would not benefit the patient or only cause them pain); S.D. CODIFIED LAWS § 59-7-2.7 (2004) (artificial nutrition or hydration may be withheld or withdrawn if "the attending physician reasonably believes that the principal's death will occur within approximately one week" or that the nutrition or hydration "cannot be physically assimilated by the principal").

155. N.Y. PUB. HEALTH LAW § 2961(12) (McKinney 1993). This statute permits unilateral decisions only in the absence of a contrary decision. *See infra* notes 376-77 and accompanying text.

156. Therefore, this statute employs a standard of "imminent demise futility," not physiological futility because the "patient will die shortly regardless of the intervention." Amir Halevy et al., *The Low Frequency of Futility in an Adult Intensive Care Unit Setting*, 156 ARCHIVES INTERNAL MED. 100, 101 (1996); *see* Amir Halevy & Baruch A. Brody, *A Multi-Institution Collaborative Policy on Medical Futility*, 276 JAMA 571, 571 (1996).

C. Quantitative Futility

While a physiological standard of medical inappropriateness is objective, a quantitative standard is subjective.[157] Though it seemingly possesses the precision of mathematics, a quantitative standard cannot be determined by reference to science alone; a quantitative standard can only be set through "reasonable consensus."[158] It is "not so much a realistic, factual or scientific concept as it is a pragmatic or useful one."[159]

Some evidence suggests that a quantitative standard of medical inappropriateness is practically implementable.[160] Proponents note that clinical studies and scoring systems can provide enough information about their likelihood to provide an empirical basis for establishing some percentage thresholds.[161] Indeed, percentages have been developed for certain patient populations.[162]

Furthermore, proponents of a quantitative standard of medical inappropriateness contend that the standard is not only workable but also ethically justified.[163] By employing such a standard, the provider is only determining whether the requested treatment can achieve the patient's goals. This determination would not necessarily challenge those goals.[164]

In fact, this is a well-established role for health care providers because they already interpret conditions specified in patients' advance directives.[165] If the advance directive states, "Treat me as long as *x*," then health care providers must determine when or whether *x* is obtainable.[166] For example, if the goal

---

157. *See* WRONG MEDICINE, *supra* note 37, at 162 ("This proposal is not an 'objective' . . . definition.").

158. *See id.*

159. Lisa Anderson-Shaw et al., *The Fiction of Futility: What to Do with Policy?*, 17 HEC FORUM 294, 295 (2005).

160. *See* WRONG MEDICINE, *supra* note 37, at 148 (discussing the use of clinical studies and scoring systems to determine overall probabilities of a treatment's effectiveness).

161. *Id.*

162. *See, e.g.*, L. Esserman et al., *Potentially Ineffective Care—A New Outcome to Assess the Limits of Critical Care*, 274 JAMA 1544, 1544-51 (1995); A. Rauss et al., *Prognosis for Recovery from Multiple Organ System Failure: The Accuracy of Objective Estimates of Chances for Survival*, 10 MED. DECISION MAKING 155-62 (1990).

163. *See, e.g.*, Nancy S. Jecker, *Medical Futility: A Paradigm Analysis*, 19 HEC FORUM 13, 25-29 (2007).

164. *See, e.g.*, AMA Council, *supra* note 53, at 937; Anderson-Shaw, *supra* note 159, at 301; Rivin, *supra* note 30, at 389 (defining "futile care" where "further treatment . . . cannot, within a reasonable possibility, cure, ameliorate, improve, or restore a quality of life that would be satisfactory *to the patient*") (emphasis added); Tomlinson & Czlonka, *supra* note 59, at 33 (criticizing the precise formulation of Rivin's policy).

165. *See infra* note 166.

166. The New Jersey advance directive statute, for example, permits patients to indicate that they want LSMT withheld or withdrawn where it "is *likely* to be ineffective or futile in prolonging life, or is *likely* to merely prolong an imminent dying process." N.J. STAT. ANN. §

for a patient in a persistent vegetative state were full recovery, a provider could determine that continued treatment would be quantitatively futile. In contrast, if the goal were family contact before death, continued treatment might not be quantitatively futile.

However, a quantitative standard of medical inappropriateness suffers from two serious problems. First, where should legislatures set the threshold percentage for quantitative futility? One percent? One-tenth of a percent? Any level is likely to be controversial and arbitrary. Second, even if lawmakers are able to settle upon a threshold percentage, then how exactly do doctors ascertain whether that threshold standard is satisfied with respect to a particular patient? Any quantitative threshold would be impossible to apply with precision across a wide variety of patients and cases.

Where, if at all, should the threshold percentage be set? The most prominent proponent of quantitative futility, Lawrence Schneiderman, argues that "a treatment should be regarded as medically futile if it has not worked in the last 100 cases . . . ."[167] Tomlinson and Czlonka argue that "[a]ttempted resuscitation is futile when it provides no meaningful possibility of extended life or other benefit for the patient."[168] But what possibility is "meaningful"? Certain scholars believe that a provider must offer even a chance of "'1 in a million.'"[169] Setting a threshold of probability not worth pursuing is a value judgment.[170] Moreover, it is a value judgment about which there is considerable variability.[171]

26:2H-67(a)(1) (West 2007) (emphasis added). The "likeliness" of these conditions occurring is determined by the health care provider.

167. WRONG MEDICINE, *supra* note 37, at 97; *cf.* Morgan County Dep't Human Servs. v. Yeager, 93 P.3d 589, 591 (Colo. Ct. App. 2004) (commenting on physician's testimony that "the likelihood of resuscitating [the patient] would be approximately one out of a hundred" and thus justified DNR order).

168. Tomlinson & Czlonka, *supra* note 59, at 33. Setting the percentage threshold also requires determining what constitutes a benefit. In *Causey,* the defendant physician "agreed that with dialysis and a ventilator Mrs. Causey could live for another two years . . . [but] that she would have only a slight (1% to 5%) chance of regaining consciousness." Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1073 (La. Ct. App. 1998).

169. GREGORY E. PENCE, CLASSIC CASES IN MEDICAL ETHICS: ACCOUNTS OF CASES THAT HAVE SHAPED MEDICAL ETHICS, WITH PHILOSOPHICAL, LEGAL, AND HISTORICAL BACKGROUNDS 11 (2d ed. 1995).

170. *See, e.g.*, Wendland v. Sparks, 574 N.W.2d 327, 332 (Iowa 1998) (refusing to dismiss medical malpractice and lost chance action based on physician's unilateral DNR order, explaining "even a small chance of survival is worth *something*"); *Causey*, 719 So. 2d at 1074 ("Placement of statistical cut-off points for futile treatment involves subjective value judgments. The difference in opinion as to whether a 2% or 9% probability of success is the critical point for determining futility can be explained in terms of personal values, not in terms of medical science."); Ferguson, *supra* note 16, at 1229 ("It appears to be a technical assessment of the limits of our technology, but these limits often become confused with the moral propriety of applying a particular technology."); *id.* at 1234 ("Simply because a treatment *is* only of marginal success does not mean that it *ought* not be pursued. Such reasoning belies a moral decision

Furthermore, even if policymakers could settle on a percentage threshold definition of medical inappropriateness, it would be difficult to employ with sufficient precision because "[p]rognostication is difficult on a case-by-case basis."[172] Thus, as applied to any particular patient, available measures from scholarly studies are very imprecise.[173]

D. Qualitative Futility

When applying either a physiological futility or a quantitative futility standard of medical inappropriateness, the provider starts with the patient's own goals and determines whether those goals are sufficiently achievable.[174] However, when applying a qualitative futility standard of medical inappropriateness, the provider questions whether the patient's goals themselves are worthwhile.[175] For example, LSMT for a patient in a persistent vegetative state can sometimes sustain the patient's life for a very long time.[176]

---

being made about the value of percentages and scientific assessments of success . . . ."); *cf.* Bouvia v. Superior Court, 225 Cal. Rptr. 297, 305 (Cal. App. 1986) ("Who shall say what the minimum amount of available life must be? Does it matter if it be 15 to 20 years, 15 to 20 months, or 15 to 20 days . . . ?").

171. *See* Lee, *supra* note 31, at 482; Karen Trotochaud, *"Medically Futile" Treatments Require More than Going to Court*, CASE MANAGER, May–June 2006, at 60, 61 ("Although most physicians believed a roughly 5% chance of survival equated to futility, the range was from 0% to 60% . . . ."). Of course, the likelihood for a specific patient can be clarified through a time-limited trial.

172. Lee, *supra* note 31, at 482; *see* James F. Blumstein, *Medical Malpractice Standard-Setting: Developing Malpractice "Safe Harbors" As a New Role for QIOs?*, 59 VAND. L. REV. 1017, 1027 (2006) (discussing the "widespread existence of clinical uncertainty").

173. *See* Arato v. Avedon, 858 P.2d 598, 601 (Cal. 1993) (observing that "statistical life expectancy data had little predictive value when applied to a particular patient with individualized symptoms, medical history, character traits and other variables"); Gampel, *supra* note 59, at 94 ("It is rare in clinical practice to have reliable numbers based on scholarly studies."); Tomlinson & Czlonka, *supra* note 59, at 31 (arguing that quantitative futility creates "the illusion of specificity" because it fails to consider "individual clinical circumstances"); Trotochaud, *supra* note 171, at 61 ("Although [scoring] systems can be helpful in predicting outcomes of populations of patients, they fail to be specific enough to be of significant help in predicting outcomes for an individual patient."); *see also* BERNARD LO, RESOLVING ETHICAL DILEMMAS: A GUIDE FOR CLINICIANS 75-76 (2d ed. 2000) (noting the likelihood of successful CPR is often mistaken); Louise Swig et al., *Physician Responses to a Hospital Policy Allowing Them to Not Offer Cardiopulmonary Resuscitation*, 44 J. AM. GERIATRICS SOC'Y 1215, 1217 (1996) (reporting 58% of those patients considered by their physicians to be unlikely to benefit from CPR were later discharged); Robert D. Truog et al., *The Problem with Futility*, 326 NEW ENG. J. MED. 1560, 1561 (1992) ("[P]hysicians are often highly unreliable in estimating the likelihood of success of a therapeutic intervention.").

174. *See supra* notes 141-46, 157-66 and accompanying text.

175. *See* WRONG MEDICINE, *supra* note 37, at 9 (questioning whether a patient's request coincides with the goal of medicine).

176. *See, e.g.*, Bush v. Schiavo, 885 So. 2d 321, 324 (Fla. 2004) (noting Theresa Schiavo

Assuming that life itself is the goal, LSMT is neither physiologically nor quantitatively futile for a patient in a persistent vegetative state because providing LSMT really will achieve this goal. In contrast, LSMT for a patient in a persistent vegetative state might be qualitatively futile because the life sustained is not "worth" sustaining.[177]

Qualitative futility has three distinct forms: (1) LSMT is inappropriate when its prospective benefits are outweighed by its associated burdens to the patient, (2) LSMT is inappropriate when its prospective benefits are not worth the required health care resources, or (3) LSMT is inappropriate when it simply cannot provide the patient a quality of life worth living.

1. Burdens Outweigh the Benefits

The first form of qualitative futility asserts that LSMT is medically inappropriate where the prospective benefits of treatment are outweighed by their associated burdens.[178] For example, for a patient that is unable to derive any pleasure, emotional enjoyment, or other satisfaction from life, the benefits of prolonged life may be outweighed by pain and suffering.[179]

Since this standard has enormous intuitive appeal, providers employ it with some regularity.[180] For example, Seattle providers were unwilling to provide

---

was in a persistent vegetative state for nearly fifteen years).

177. *See In re* Finn, 625 N.Y.S.2d 809, 812–13 (N.Y. Sup. Ct. 1995) (noting doctor unilaterally entered DNR order for a patient on grounds that CPR would be medically futile because the patient was "profoundly retarded and would likely be more severely retarded after the administration of CPR" and therefore that patient's life would not be "worth living").

178. *See* Pellegrino, *supra* note 2, at 3 ("[W]hen the capabilities of medicine to cure, ameliorate, or reverse a disease process have been exhausted[,] . . . continuance of treatment under those circumstances may impose further suffering and other burdens on the patient—physical, emotional, and fiscal."); Linda B. Siegel, *When Staff and Parents Disagree: Decision Making for a Baby with Trisomy 13*, 73 MOUNT SINAI J. MED. 590, 591 (2006) (describing baby who "was suffering significantly" and "did not appear to get any pleasure from life"); Tomlinson & Czlonka, *supra* note 59, at 33 (defining attempted resuscitation as "harmful" where "harm inflicted on the patient is grossly disproportionate to any possibility of benefit"); *see also* Morreim, *supra* note 124, at 898. Under the circumstances, a compelling case can be made that a surrogate demanding such continued aggressive treatment should be stripped of decision making authority. *See supra* notes 101-06.

179. Some have referred to this qualitative standard as the "unbearable situation." ROYAL COLLEGE OF PAEDIATRICS AND CHILD HEALTH, WITHHOLDING OR WITHDRAWING LIFE SUSTAINING TREATMENT IN CHILDREN: A FRAMEWORK FOR PRACTICE 29 (2d ed. 2004). Others have defined the treatment to be "inhumane." 45 C.F.R. § 1340.15(b)(2)(iii) (1990).

180. An "objective" or "best interests" standard is well-established for proxy decision makers in circumstances where they have little or no evidence of the patient's preferences. *See, e.g.*, OR. REV. STAT. § 127.580(1)(b) (West 2005) (providing an exception to the administration of nutrition or hydration if it causes "severe, intractable or long-lasting pain"); *id.* § 127.635(1)(c) (allowing withdrawal of LSMT if it creates no benefit to patient's condition or causes "permanent and severe pain"); S.D. CODIFIED LAWS § 59-7-2.7 (2004) (allowing

long-term dialysis to Ryan Nguyen, "since it would prolong agony with 'no likelihood of a good outcome.'"[181] Baylor Regional Medical Center at Plano withdrew LSMT from Tirhas Habtegiris because the care was "disproportionately burdensome"[182] and was only "increasing her suffering."[183] Similarly, D.C. Children's Hospital wanted to withdraw LSMT from Baby Rena because she had no prospect for recovery or positive interaction with her environment and had to be "constantly sedated" to soothe her continuous pain.[184]

In one of the earliest reported futility cases, providers argued that further intervention for "Baby L" would be inhumane and that continued LSMT "would only add to her pain, without helping."[185] Because Baby L was blind, deaf, quadriplegic, and could not otherwise interact with her environment, maintaining her on a respirator provided no opportunity for improvement or cure, but only more seizures, infections, and cardiac arrests.[186] Baby L "could experience nothing but pain."[187]

2. Resources Outweigh the Benefits

The second form of qualitative futility also weighs the prospective benefits of treatment. Yet, unlike the first form, which balances the benefits against the burdens of treatment for the patient, the second form balances the benefits against the health care resources used to provide the treatment.[188] Under this theory, LSMT is medically inappropriate where it is not worth the requisite resources that are better spent elsewhere.[189]

---

withdrawal or withholding of artificial nutrition or hydration if "the burden of providing [it]... outweighs its benefit, provided that the determination of burden refers to the provision of artificial nutrition or hydration itself and not to the quality of the continued life of the principal"). As a standard, qualitative futility has been employed not only to patients without subjective preferences but also to patients who have exercised preferences for continued LSMT.

181. Alexander M. Capron, *Baby Ryan and Virtual Futility*, 25 HASTINGS CTR. REP., Mar.-Apr. 1995, at 20.

182. Baylor Response, *supra* note 144.

183. Baylor Health Care System, Tirhas Habtegiris Case: Medical History, http://www.baylorhealth.com/articles/habtegiris/history.htm (last visited Oct. 15, 2007).

184. Weiser, *supra* note 45, at A1.

185. Joan Beck, *Use Medical Technology to Save Every Damaged Baby?*, ORLANDO SENTINEL, May 18, 1990, at A13; *see also* John J. Paris et al., *Physicians' Refusal of Requested Treatment: The Case of Baby L*, 322 NEW ENG. J. MED. 1012, 1013 (1990) (reporting conclusion of ethics committee meeting that because Baby L "could experience only pain[,]" further LSMT was "not in the best interest of the patient").

186. Beck, *supra* note 185, at A13.

187. *Id.*

188. *See* AMA Council, *supra* note 53, at 938 ("Another context in which futility questions come up is resource allocation. Some commentators argue that elimination of futile care is good for both patients and allocation of resources.").

189. *See* Tomlinson & Czlonka, *supra* note 59, at 32 ("Many interventions are not

Commentators have articulated both a modest and a robust version of resource-focused qualitative futility. The modest version focuses on hard resources like ICU beds.[190] When these resources are needed by other patients with better prospects, then it is inappropriate to give those resources to the patient with the poorer prospects.[191] This modest version of resource-focused qualitative futility is similar to the concept of "triage" where emergency room providers do "not work on a first come, first serve basis," but serve the most urgent or severe yet treatable injuries and illnesses first "to avoid [any] delay in treatment."[192] The modest version of resource-focused qualitative futility is employed in a few states.[193]

While the modest version of resource-based qualitative futility is well-grounded, the robust version of resource-focused qualitative futility is more controversial. Rather than looking to the allocation of hard resources, the robust version examines the allocation of soft resources like health care dollars.[194] In many cases, families allege that providers make unilateral

---

costworthy because they consume too many resources *relative* to their benefit, not because they offer no benefits at all."). Some commentators have referred to this as "[t]herapeutic extravagance . . . mean[ing] the provision of high-cost treatments that offer little or no benefit." Tovino & Winslade, *supra* note 41, at 2-3 n.5.

190. *See, e.g.*, Amy Iggulden, *Premature Babies Are Blocking Beds, Says Royal Medical College*, TELEGRAPH, Mar. 27, 2006, *available at* http://www.telegraph.co.uk/news/main.jhtml?xml=/news/2006/03/27/nprem27.xml (reporting that the Royal College of Obstetrics and Gynaecology felt that very premature babies were "'bed blocking'" by "taking up intensive care space that could be used by healthier babies"); Roy Lilley, *A Bad Time to Be Very Young or Old*, TELEGRAPH, Mar. 28, 2006, *available at* http://www.telegraph.co.uk/opinion/main.jhtml?xml=/opinion/2006/03/28/do2802.xml ("[T]he Royal College of Obstetricians and Gynaecologists . . .[suggests that] [b]abies born at 25 weeks . . . should be left to die . . . [because] more weight [should] be given to 'economic considerations.'").

191. *See* sources cited *supra* note 190.

192. *See* STEVEN E. PEGALIS, AMERICAN LAW OF MEDICAL MALPRACTICE § 6.18 (3d ed. 2006) (describing the process of "triage"); JOINT COMM'N ON THE ACCREDITATION HEALTHCARE ORGANIZATIONS, MANAGING PATIENT FLOW: STRATEGIES AND SOLUTIONS FOR ADDRESSING HOSPITAL OVERCROWDING 120–29 (2004).

193. *See, e.g.*, OKLA. STAT. ANN. tit. 63, § 3101.9 (West 1998) ("Nothing in this section shall require the provision of treatment if the physician or other health care provider is physically or legally unable to provide or is physically or legally unable to provide without thereby denying the same treatment to another patient."); VA. CODE ANN. § 54.1-2990(C) (1992) ("Nothing in this section shall require the provision of treatment that the physician is physically or legally unable to provide, or treatment that the physician is physically or legally unable to provide without thereby denying the same treatment to another patient.").

194. *See, e.g.*, Mary Ann Roser, *Austin Doctors Want to Withdraw Care from Vegetative Patient (Terri Schiavo Type Situation)—Family Objects and Says Woman Is Still Aware; Seeking Transfer to Another Facility in Texas*, AUSTIN AM.-STATESMAN, Apr. 28, 2006, *available at* http://www.freerepublic.com/focus/f-chat/1623122/posts (reporting in the case of Lang Yen Thi Vo that the patient's daughter "sees a financial reason behind the decision [to withdraw care] . . . . Her mother will soon exhaust her Medicare and Medicaid benefits.").

withdrawal decisions based on financial reasons.[195] Providers, on the other hand, almost always deny that money is a relevant factor.[196]

Whether or not providers determine LSMT to be inappropriate based in whole or in part on its cost, most commentators agree that neither resource consumption nor rationing is a legitimate ground for making life-and-death decisions for individual patients.[197] For example, a treatment that has a 2%

---

195. *See, e.g., id.*

196. *See, e.g., In re* Baby K, 832 F. Supp. 1022, 1026 (E.D. Va. 1993) ("The Hospital has stipulated that it is not proposing to deny ventilator treatment to Baby K because of any lack of adequate resources or any inability of Ms. H to pay for the treatment."); Burke v. Gen. Med. Council [2004] EWHC 1879 (Admin), 2 W.L.R. 431, 444–45 (2005) (explaining that the case was not "about the prioritisation [sic] or allocation of resources" or concern over "significant cost implications"); WRONG MEDICINE, *supra* note 37, at 53 (noting that Wanglie's providers "avoided seeking court permission to withdraw treatment on another patient who happened to be in the hospital at the same time in a similar condition—but who happened to be on welfare"); Schwartz, *supra* note 105, at 161 ("[Helga Wanglie's] hospitalization cost nearly 1 million dollars, which was paid by Medicare and her private medigap insurance carrier. Neither objected to the care for financial or cost-benefit reasons, and the cost properly did not enter into the judicial analysis of the case."); Frank Bruni, *Care vs. Cost: Suit Against Pa. Hospital on Life Support Raises Questions*, PITTSBURGH POST-GAZETTE, Mar. 10, 1996, at A1 (reporting that the CEO of Hershey Medical Center denied that the financial cost of caring for Brianna [Rideout] or the fact that her insurance was running out influenced the decision to remove her ventilator); Kowalczyk, *Mortal Differences*, *supra* note 66 (reporting Massachusetts General Hospital executives denied that they were motivated to stop Barbara Howe's LSMT because Blue Cross stopped paying); Roser, *supra* note 194 (reporting in the case of Lang Yen Thi Vo that the hospital "had no idea of Vo's financial status and that it was not a factor"); Baylor Response, *supra* note 144 ("The hospital did not stop treatment because of economic considerations. . . . The same course of action followed in this case has in the past been followed with privately insured patients . . . ."). While costs may not be the basis of the unilateral decision, they may be the reason other institutions refuse to accept transfer of the patient. *See, e.g.*, Murphy, *supra* note 43, at A37 (reporting that while the family of Joseph Ndiyob eventually found a Los Angeles hospital willing to accept him, the hospital "recanted when it learned he lacked health insurance"); Smith, *supra* note 26 ("[P]atients who would be refused care under futility protocols would usually be the most expensive to care for and thus, given the economics of managed care, probably unwelcome in another institution."). For some, it is unnecessary to even consider the legitimacy of cost-based inappropriateness, because other more acceptable standards are available. *See* Hudson, *supra* note 66, at 32 ("Some treatments—such as keeping a patient in a persistent vegetative state alive, even if it costs only 10 cents a day—are not what medicine is about." (quoting Lawrence Schneiderman)).

197. AMA Council, *supra* note 53, at 938 ("Efforts to understand futility should not make use of resource-saving criteria, and rationing needs should not motivate declarations of futility."); Dubler, *supra* note 26, at 297 ("[F]inancial disincentives . . . must not be permitted to contaminate decisions about death."); S.Y. Tan et al., *Creating a Medical Futility Policy*, HEALTH PROGRESS, July–Aug. 2003, *available at* http://findarticles.com/p/articles/mi_qa3859/is_200307/ai_n9263834/pg_1 ("Resource consumption, inability to pay, or rationing are not legitimate criteria to be used in defining medical futility."); Tomlinson & Czlonka, *supra* note 59, at 32 (relying on costs to define, rather than just to prompt consideration of medical inappropriateness, will poison communication, credibility, and trust).

chance of extending a patient's life ten days at a cost of $1 million may be too expensive. However, the consideration of cost, alone, would not make the treatment medically inappropriate.

3. Treatment Cannot Provide a Worthwhile Quality of Life

The third form of qualitative futility does not weigh the prospective benefits of treatment against either the prospective burdens or the required resources.[198] Instead, providers determine that the expected outcome of the requested treatment is of no value, without regard to either burdens or resources.[199] The provider judges the expected outcome to be of no value because of the patient's extremely poor condition or prognosis.[200]

The most notable situations, in which providers consider continued LSMT to be qualitatively inappropriate, exist when a patient is permanently unconscious, totally dependent on intensive medical care, or both.[201] Permanent unconsciousness means a condition that, to a high degree of medical certainty, will last permanently without improvement. In this condition, patients have no thought, sensation, purposeful action, social interaction, awareness of self, or awareness of their environment.[202]

But cf. HALL ET AL., *supra* note 78, at 600 (suggesting that futility may be a mask for rationing and driven by the concern about scarce health care resources); Lantos, *supra* note 12, at 589 ("[T]he only downside to trying a treatment that is unlikely to work is economic. It will be a wasted expenditure. To the extent that this is the case, futility determinations collapse into rationing decisions."); Mildred Z. Solomon, *How Physicians Talk about Futility: Making Words Mean Too Many Things*, 21 J. L. MED. & ETHICS 231, 232-33 (1993) (explaining that medical futility denotes "both efficacy *and* evaluative judgments about the wisdom of pursuing further treatment").

198. *See* AMA Council, *supra* note 53, at 937 (without considering benefits or resources, the physician simply "sees dying as inevitable and wishes to pursue the goal of comfort care").

199. *See id.* at 938 (examining the qualitative approach of the "worth-the-effort quality of life" standard).

200. *See id.* at 937 (providers may decline intervention as futile if the intent is only to prolong dying).

201. *See, e.g.*, GA. CODE ANN. § 31-39-2(4)(B) (Supp. 2007) ("'Candidate for nonresuscitation' means a patient . . . in a noncognitive state with no reasonable possibility of regaining cognitive functions."); N.C. GEN. STAT. § 90-322(a) to (b) (2006) (permitting providers, in the absence of a contrary patient or surrogate request, to unilaterally stop LSMT for a patient who is in a persistent vegetative state or terminal, incurable and comatose or mentally incapacitated); OR. REV. STAT. § 127.580 (2005). This provision similarly permits providers, in the absence of a contrary request from the patient or surrogate, to unilaterally stop LSMT for a patient who is "permanently unconscious" or who "has a progressive illness that will be fatal and is in an advanced stage," if the patient "is consistently and permanently unable to communicate by any means, swallow food and water safely, care for the person's self and recognize the person's family and other people, and it is very unlikely that the person's condition will substantially improve." *Id.*

202. *See, e.g.*, OR. REV. STAT. § 127.505(18) (2005) ("'Permanently unconscious' means completely lacking an awareness of self and external environment, with no reasonable

Withholding LSMT as medically inappropriate based on a quality of life assessment is a heavily criticized standard.[203] While some accept that an individual may make a personal choice to forgo LSMT, it is highly controversial for a health care provider to make this decision on the patient's behalf.[204] The controversy arises because health care providers can be poor predictors of a patient's quality of life.[205] The point at which life becomes "worthless" is not known to the patient's health care provider "any better than [it is] known to nine people picked at random from the Kansas City telephone directory."[206] A health care provider may judge the patient's quality of life to be far less than the patient would.[207] Nonetheless, "people with physical, sensory, and cognitive impairments can and do obtain many satisfactions and rewards in their lives."[208] For this reason, Professor Felicia Ackerman rejects quality of life determinations:

---

possibility of a return to a conscious state, and that condition has been medically confirmed by a neurological specialist who is an expert in the examination of unresponsive individuals.")

203. *Cf.* Superintendent of Belchertown State Sch. v. Saikewicz, 370 N.E.2d 417, 432 (Mass. 1977) ("To the extent that this formulation equates the value of life with any measure of the quality of life, we firmly reject it.").

204. *See* Adrienne Asch, *Recognizing Death While Affirming Life: Can End of Life Reform Uphold a Disabled Person's Interest in Continued Life?*, 35 HASTINGS CTR. REP. (SPECIAL REPORT), Nov.–Dec. 2005, at S31. Dr. Asch questions the autonomy of the patient's choice and argues that "clinicians and policymakers [should be prompted] to question how truly autonomous is anyone's wish to die when living with changed, feared, and uncertain physical impairments . . . ." *Id.* at S33; *see also* ROBERT L. BURGDORF JR., NATIONAL COUNCIL ON DISABILITY, ASSISTED SUICIDE: A DISABILITY PERSPECTIVE 48 (1997) ("The pressures upon people with disabilities to choose to end their lives . . . are already way too common in our society. These pressures are increasing and will continue to grow . . . .").

205. *See* Susan Dorr Goold et al., *Conflicts Regarding Decisions to Limit Treatment: A Differential Diagnosis*, 283 JAMA 909, 912 (2000) (noting the uncertainty of prognostications made by doctors).

206. Cruzan v. Dir., Mo. Dep't of Health, 497 U.S. 261, 293 (1990) (Scalia, J., concurring).

207. *See* Asch, *supra* note 204, at S35 (questioning the basis of a provider's decision to end LSMT contrary to the patient's and the patient's family's wishes because the provider felt that continued treatment was "inhumane").

208. *Id.* at S32; *see also In re* Finn, 625 N.Y.S.2d 809, 813 (N.Y. Sup. Ct. 1995) ("Although Leonard's life as a developmentally disabled person may seem a small possession from the perspective of some, it remains his possession and 'no person or court should substitute its judgment as to what would be an acceptable quality of life for another.'" (quoting *In re* Westchester County Med. Ctr. *ex rel* O'Connor, 531 N.E.2d 607, 613 (N.Y. 1988))); A Nat'l Health Serv. Trust v. D, [2000] EWHC FD 00P10551 (Fam), [2000] 2 FLR 677, 687 (Eng.) (describing child with terminal illness who "has a delightful smile and can indicate pleasure and displeasure"); Lewis Smith, *Victory for Dying Boy's Family*, THE TIMES, Mar. 16, 2006, at 4 (reporting High Court in London refused application to withdraw ventilator from a 18-month old baby with spinal muscular atrophy because even though the baby was paralyzed, he could still experience pleasure from sight, touch, and sound).

> It is as presumptuous and ethically inappropriate for doctors to suppose that their professional expertise qualifies them to know what kind of life is worth prolonging as it would be for meteorologists to suppose their professional expertise qualifies them to know what kind of destination is worth a long drive in the rain.[209]

Some commentators refer to this as the problem of "therapeutic illusion" because providers may not recognize possible benefits of treatment.[210] Furthermore, a qualitative standard of inappropriateness, unmoored from any demonstrable weighing of benefits and burdens, is obviously subject to abuse.[211]

E. Summary of Definitions of "Medically Inappropriate"

Despite an exhaustive debate over the past fifteen years, only brain death and physiological futility are fully supported by a consensus in the medical, legal, and bioethical communities as acceptable definitions of medical inappropriateness.[212] However, these are not the relevant conditions in the vast majority of futility disputes. The typical case involves a living patient for whom LSMT can produce *some* effect.[213]

In order to define a treatment as medically inappropriate, a health care provider typically must question whether the expected effect on the patient is beneficial and worthwhile. There is no consensus about this.[214] Many

---

209. Felicia Ackerman, *The Significance of a Wish*, 21 HASTINGS CTR. REP., July-Aug. 1991, at 27, 28 (emphasis omitted). But while people may find satisfaction despite severe physical or mental handicaps, this is not possible where they are irreversibly unconscious. *See* WRONG MEDICINE, *supra* note 37, at 18.

210. Tovino & Winslade, *supra* note 41, at 2 n.5.

211. *Cf.* ROBERT JAY LIFTON, THE NAZI DOCTORS: MEDICAL KILLING AND THE PSYCHOLOGY OF GENOCIDE 45 (1986) (describing the Nazi Germany program whereby the disabled, labeled as "life unworthy of life," were euthanized).

212. *Cf.* Robert S. Chabon, *The Case of Baby K*, 311 NEW ENG. J. MED. 1383, 1383 (1994) ("Within the medical profession itself there appear to be disputes about whether physicians must provide medically inappropriate interventions on a patient's or surrogate's request."); Fletcher, *supra* note 10, at S:232 (comparing the narrow, objective definition of futility as physiologically ineffective with the broad, subjective definition of "nonbeneficial"); William Prip & Anna Moretti, *Medical Futility: A Legal Perspective*, *in* MEDICAL FUTILITY AND THE EVALUATION OF LIFE-SUSTAINING INTERVENTIONS 136, 137 n.2 (Marjorie B. Zucker & Howard D. Zucker eds., 1997) (examining roots of notion that brain death defines actual death).

213. *Cf.* Prip & Moretti, *supra* note 212, at 137 (describing the progression from patients demanding assisted suicide to patients challenging the physician's decision to stop LSMT).

214. *See* Jeffrey P. Burns & Robert D. Truog, *Futility: A Concept in Evolution*, 132 CHEST 1987, 1988-89 (2007) (reviewing "problems inherent to definitional approaches to futility"); Bryan Rowland, *Communicating Past the Conflict: Solving the Medical Futility Controversy with Process-Based Approaches*, 14 U. MIAMI INT'L & COMP. L. REV. 271, 284 (2006) ("Even those who accept the concepts of quantitative and qualitative futility disagree on how to draw the dividing line between futile and non-futile care.").

providers are unable to reduce medical inappropriateness to an algorithm "contained within the four corners of a formula."[215] Consequently, medical inappropriateness can only be identified the way beauty is perceived, "in the eye of the beholder,"[216] or the way pornography is identified—we know it when we see it.[217]

III. LEGAL CONSTRAINTS ON THE UNILATERAL TERMINATION OF LSMT

Employing these ad hoc definitions of medical inappropriateness, providers often want to stop LSMT unilaterally when they are unable to secure surrogate consent; however, the unilateral withholding and withdrawing of LSMT is remarkable in three important respects.[218] First, it typically results in the patient's death.[219] Second, it is rare and unusual.[220] Third, and most significantly, it devalues patient autonomy.[221]

Before the 1970's, this devaluation of patient autonomy did not seem so remarkable.[222] Historically, it did not matter so much what the patient wanted because health care providers just provided the treatment that they thought was right.[223] But today, providers generally must comply with treatment requests made by or on behalf of their patients.[224] Autonomy has become the touchstone.

This nonconsensual aspect of unilateral termination is the most distinctive.[225] Both without patient or surrogate consent and typically even over

---

215. Tunkl v. Regents of Univ. of Cal., 383 P.2d 441, 444 (Cal. 1963).

216. *Cf.* Fla. Power & Light Co. v. Lorion, 470 U.S. 729, 737 (1985) (discussing the subjectivity of the "plain meaning" of a statute granting judicial review of a lower court's decision).

217. *Cf.* Jeffrey Burns, *Does Anyone Actually Invoke Their Hospital Futility Policy?*, 12 LAHEY CLINIC MED. ETHICS J. 3, 3 (2005) (comparing futility to Potter Stewart's remark of pornography: "I know it when I see it" (quoting Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring))).

218. *See* Diane E. Hoffmann & Jack Schwartz, *Who Decides Whether a Patient Lives or Dies?*, TRIAL, Oct. 2006, at 30, 32 (revealing these three aspects through the last days of a two-year-old patient).

219. *See id.* at 32, 34.

220. *See id.* at 32.

221. *See id.*

222. *See* RUTH R. FADEN & TOM L. BEAUCHAMP, A HISTORY AND THEORY OF INFORMED CONSENT 76-101 (1986).

223. *See id.*

224. *See, e.g.*, CAL. PROB. CODE § 4650(a) (West Supp. 2007) ("[T]he law recognizes that an adult has the fundamental right to control the decisions relating to his or her own health care . . . ."); *id.* § 4733 (stating that health care providers are required to comply with the requests of their patients or surrogates); Ferguson, *supra* note 16, at 1237 ("[T]wenty-five years of patients' rights development indicate that unilateral actions are *not* the standard. The unilateral withdrawal of care . . . violates our sense of patient autonomy . . . .").

225. MASON & LAURIE, *supra* note 77, at 601 ("An act of involuntary euthanasia involves

vehement opposition, the provider causes the patient's death.[226] Consequently, taking unilateral action can expose the health care provider to civil, criminal, and disciplinary sanctions.[227]

A. Civil Sanctions

Health care providers who make unilateral decisions to stop LSMT may be subject to a wide array of civil sanctions. Reported cases show claims for patient neglect and abuse,[228] infliction of emotional distress,[229] and breach of contract.[230] However, the causes of action most often utilized in response to unilateral decisions are the following: (1) lack of informed consent, (2) medical malpractice, and (3) wrongful death.[231]

---

ending the patient's life in the absence of either a personal or proxy invitation to do so."); Brenda Fastabend, *Virginia's Involuntary Euthanasia Problem*, VSHL LIFESAVER, Aug. 1999, *available at* http://www.vshl.org/education/euthanasia/5_4/5_4_4_Virginia_Involuntary_Euthanasia_Problem.shtml (referring to "medical futility" as "involuntary [passive] euthanasia"). Where patients decline LSMT through contemporaneous decisions, advance directives, or surrogates, this is known as "voluntary passive euthanasia." Medical futility is characterized as "passive" where providers withhold or withdraw LSMT, but take no affirmative action such as a lethal injection. Medical futility becomes "involuntary" when LSMT is stopped *without* the patient's or surrogate's consent.

226. *See* Hoffman & Schwartz, *supra* note 218, at 32.

227. *See infra* notes 228-75 and accompanying text. The following discussion is qualified in three respects. First, this Article does not distinguish the liability of the individual provider from that of the institutional provider. Second, while the Article assumes that the provider has already implemented the unilateral decision, in addition to these *ex post* sanctions, the patient or surrogate may seek injunctive relief. Third, the Article focuses here on state law. For discussion of federal law constraints, see *infra* notes 439 to 464 and accompanying text. For a thorough analysis of futility disputes in court, see Pope, *supra* note 104.

228. *See, e.g.*, *In re* Estate of Greenspan, 558 N.E.2d 1194, 1200 (Ill. 1990).

229. *See, e.g.*, Manning v. Twin Falls Clinic & Hosp., 830 P.2d 1185, 1187 (Idaho 1992); Morgan v. Olds, 417 N.W.2d 232, 236 (Iowa Ct. App. 1987); Rideout v. Hershey Med. Ctr., 30 Pa. D. & C.4th 57, 62 (Dauphin County Ct. C.P. Dec. 29, 1995) (No. 872S1995), 1995 WL 924561; Strickland v. Deaconess Hosp., 735 P.2d 74, 75 (Wash. Ct. App. 1987); Capron, *supra* note 47, at 28 (discussing *Gilgunn v. Massachusetts Gen. Hosp.*, No. CIV.A.92-4820 (Suffolk Super. Ct. Apr. 21, 1995)); Hoffman & Schwartz, *supra* note 218, at 30, 32 (citing Bland v. Cigna Healthplan of Tex., No. 93-52630 (Harris Cty., Tex. Dist. Ct. Apr. 25, 1995)).

230. *See, e.g.*, Gamble v. Perra, No. E2006-00229-COA-R3-CV, slip. op. at 2 (Tenn. Ct. App. Feb. 22, 2007). There is also potential exposure for providers under state disability laws. *See, e.g.*, ALASKA STAT. § 13.52.135 (2006) ("When determining the best interest of a patient under this chapter, health care treatment may not be denied to a patient because the patient has a disability or is expected to have a disability.").

231. *See infra* notes 232–54 and accompanying text. At least one court has suggested that unilateral decisions to terminate would constitute tortuous abandonment. Bryan v. Rectors & Visitors of the Univ. of Va., 95 F.3d 349 (4th Cir. 1996) ("Such reprehensible disregard for one's patient . . . would . . . constitute . . . the well established tort of abandonment."). However, abandonment claims are weak for two reasons. First, it is unlikely that the physician-

1. Lack of Informed Consent

Patients and surrogates have brought informed consent actions against health care providers that implemented unilateral decisions to stop LSMT.[232] For example, in *Rideout v. Hershey Medical Center,* the hospital withdrew a ventilator from a two-year-old girl, not only without her parent's consent, but also "against their vehement and desperate opposition."[233] The court overruled the hospital's motion to dismiss the parent's informed consent cause of action.[234] The case subsequently settled for an undisclosed sum.[235]

The doctrine of informed consent requires health care providers to obtain consent to discontinue a patient's treatment.[236] In a typical futility dispute, the

---

patient relationship would be terminated completely and unilaterally. *See* THE RIGHT TO DIE, *supra* note 17, at § 11.03[d]; Prip & Moretti, *supra* note 212, at 142. While a provider may decline to continue LSMT, the provider must continue comfort care. *See* sources cited *supra* note 79 and accompanying text. Second, if treatment were medically inappropriate, then the treatment relationship would have already ended because the provider's services were no longer necessary. *See* Levine, *supra* note 10, at 88 (arguing that if the treatment is medically inappropriate, then physician's services were no longer necessary).

232. *See, e.g.*, *Morgan*, 417 N.W.2d at 235 (DNR order without patient's consent); Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1075-76 (La. Ct. App. 1998) (discussing that while the physician explained the situation to the patient's family, he withdrew the treatment "despite the lack of any consent"); *Strickland,* 735 P.2d at 75 (patient removed from respirator without consent); Preston v. Meriter Hosp., Inc., 678 N.W.2d 347, 352 (Wis. Ct. App. 2004) (dismissing informed consent claim against hospital because it had no independent duty to obtain consent; only doctors are required to obtain informed consent), *rev'd on other grounds*, 700 N.W. 2d 158 (Wis. 2005); Belcher v. Charleston Area Med. Ctr., 422 S.E.2d 827, 838 (W. Va. 1992) (remanding case for trial on whether doctors should have sought parental consent for DNR order from a patient just a few weeks shy of 18).

233. *Rideout,* 30 Pa. D. & C.4th at 69-70.

234. *Id.* at 73.

235. *See* Email from Thomas W. Hall to Thaddeus M. Pope (May 4, 2007) (on file with the Tennessee Law Review).

236. Informed consent also requires health care providers to disclose information about the treatment and its alternatives. However, providers probably have no duty to advise the patient or surrogate of the option to continue treatment that the provider considers inappropriate. Physicians need not disclose information about unreasonable options. They need not disclose information about procedures and interventions that are not within the medical standard of care. *See* FAY A. ROZOVSKY, CONSENT TO TREATMENT: A PRACTICAL GUIDE § 1.02 (4th ed. 2007) (describing the characteristics of a "valid consent" and what disclosure is required by the provider); *see also* Peter D. Jacobson & C. John Rosenquist, *The Introduction of Low-Osmolar Agents in Radiology: Medical, Economic, Legal, and Public Policy Issues*, 260 JAMA 1586, 1588-89 (1988) (discussing the requirements of informed consent for radiologists implementing a new contrast media); Paris, *supra* note 185, at 1013 ("[A] physician who merely spreads an array of vendibles in front of the patient [or family] and then says, 'Go ahead and choose, it's your life,' is guilty of shirking his duty, if not of malpractice." (quoting F.J. Ingelfinger, *Arrogance*, 303 NEW ENG. J. MED. 1507 (1980))). Furthermore, in a futility conflict, the patient's surrogates are typically already aware of the treatment options that the health care

surrogate demands the continuation of treatment in opposition to the provider's wishes.[237] Therefore, the provider who unilaterally discontinues treatment fails to obtain consent *and* overrides the surrogate's explicit opposition.[238]

2. Medical Malpractice and Negligence

In addition to causes of action for lack of informed consent, patients and surrogates have brought medical malpractice and negligence actions against health care providers that made unilateral decisions to stop LSMT.[239] For example, in *Causey v. St. Francis Medical Center,* a physician and hospital withheld LSMT from a 31-year-old quadriplegic, comatose patient with kidney failure over the strongly expressed objections of her family.[240] While the family members pleaded an intentional tort cause of action, the court allowed the case to proceed as a medical malpractice case.[241]

The heart of a medical malpractice claim is that the provider failed to administer the care and skill ordinarily exercised by members of their profession practicing in the same or similar location under similar circumstances.[242] Therefore, providers should not be exposed to malpractice liability if stopping LSMT really is the standard of care.[243] Because the medical

---

provider judges medically inappropriate. *See, e.g., In re* Baby K, 16 F.3d 590, 592 (4th Cir. 1994); *Causey*, 719 So. 2d at 1075-76.

237. *See, e.g., Baby K*, 16 F.3d at 593; *Causey*, 719 So. 2d at 1075-76.

238. *See, e.g., Baby K*, 16 F.3d at 593; *Causey*, 719 So. 2d at 1075-76.

239. *See, e.g.*, Wendland v. Sparks, 574 N.W.2d 327, 328 (Iowa 1998) (allowing husband to proceed with malpractice action against physician and hospital that unilaterally decided not to attempt CPR on his wife); *Causey*, 719 So. 2d at 1072; Kelly v. St. Peter's Hospice, 553 N.Y.S.2d 906, 907 (N.Y. App. Div. 1990) (reviewing plaintiff's medical malpractice claim alleging that physician and facility failed to provide "sufficiently aggressive" treatment to patient with metastatic breast cancer); Strickland v. Deaconess Hosp., 735 P.2d 74, 75 (Wash. Ct. App. 1987); Preston v. Meriter Hosp., Inc., 678 N.W.2d 347, 351 (Wis. Ct. App. 2004) (dismissing medical malpractice against hospital that intentionally refused to treat premature infant because plaintiff failed to identify an expert); *see also* Litz v. Robinson, 955 P.2d 113, 114 (Idaho Ct. App. 1998) (hearing "alleg[ation] that the defendants breached their duties as physicians by wrongfully withholding life sustaining procedures"); King v. Crowell Mem'l Home, 622 N.W.2d 588, 591-92 (Neb. 2001) (hearing medical malpractice action alleging that the defendant nursing home classified the decedent as a DNR patient even though their instructions were to use "'any and all medical measures'"); Gamble v. Perra, No. E2006-00229-COA-R3-CV, slip op. at 2 (Tenn. Ct. App. 2007) (wife "alleged that her husband could have lived longer, but for his lack of treatment by defendants").

240. *Causey*, 714 So. 2d at 1073.

241. *Id.* The trial court found that, as a medical malpractice action, the claim must first be presented to a medical review panel. *Id.* As a result, the court dismissed the plaintiff's action as premature. *Id.*

242. BARRY R. FURROW ET AL., HEALTH LAW § 6-2, at 264–65 (2d ed. 2000).

243. *See id.* § 6-2 at 269 (describing how guidelines establish the standard of care and therefore provide a shield against liability); *see also id.* § 16-77 at 905 ("[H]ealth care providers must offer patients only that range of treatments that is medically indicated under the

standard of care is custom-based, malpractice liability would not seem to present an obstacle to unilaterally stopping LSMT.[244] Although providers do in fact collectively set the standard, three implementation realities dispel this notion.[245]

First, "the practical difficulties of proving just what is the prevailing medical custom break down this protective theory in the real world."[246] Second, to the extent the standard of care is ascertainable, unilaterally stopping LSMT is not now the standard of care.[247] As Justice Brennan observed, "[c]urrent medical practice recommends use of heroic measures if there is a scintilla of a chance that the patient will recover . . . ."[248] Third, by continuing to give such care, providers are creating and perpetuating the very standard with which they do not want to comply.[249]

3. Wrongful Death

In addition to informed consent and medical malpractice actions, patients and surrogates have brought wrongful death suits against health care providers that made unilateral decisions to stop LSMT.[250] In *Velez v. Bethune,* the physician unilaterally terminated the life support of a severely impaired infant.[251] The court held that the parents had a valid claim for wrongful death.[252] The court stated that "Dr. Velez had no right to decide, unilaterally, to discontinue medical treatment even if, as the record in this case reflects, the child was terminally ill and in the process of dying. That decision must be

---

circumstances."); Laurence J. Schneiderman & Alexander Morgan Capron, *How Can Hospital Futility Policies Contribute to Establishing Standards of Practice?*, 9 CAMBRIDGE Q. HEALTHCARE ETHICS 524, 529 (2000) (arguing that any one of various standards is sufficient if a "'respectable minority'" of physicians would stop LSMT); *cf. Kelly*, 533 N.Y.S.2d at 907-08 (patient's husband failed to present evidence that treatment departed from acceptable medical practice).

244. FURROW, *supra* note 242, § 6-2 at 265, § 16-77 at 905.

245. *Id.* § 6-2 at 265.

246. Mark A. Hall, *The Defensive Effect of Medical Practice Policies in Malpractice Litigation*, 54 LAW & CONTEMP. PROBS. 119, 127 (1991).

247. FURROW, *supra* note 242, § 16-77 at 906.

248. Cruzan v. Dir., Mo. Dep't of Health, 497 U.S. 261, 314 (1940) (Brennan, J., dissenting); *see also* Middleditch & Trotter, *supra* note 26, at 399-400 (finding that use of mechanical feeding and breathing devices for patients in a persistent vegetative state as the new custom).

249. *See* FURROW, *supra* note 242, at § 6-2 at 269.

250. *See, e.g.*, Kranson v. Valley Crest Nursing Home, 755 F.2d 46, 48 (3d Cir. 1985); Velez v. Bethune, 466 S.E.2d 627, 628 (Ga. Ct. App. 1995); Manning v. Twin Falls Clinic & Hosp., Inc., 830 P.2d 1185, 1187 (Idaho 1992) (affirming award of compensatory and punitive damages); Wendland v. Sparks, 574 N.W.2d 327, 328-29 (Iowa 1998); Belcher v. Charleston Area Med. Ctr., 422 S.E.2d 827, 830 (W. Va. 1992).

251. *Velez*, 466 S.E.2d at 628.

252. *Id.*

made *with* the consent of the parents."[253] While both the imminence and inevitability of the infant's death may have been relevant to the amount of damages, neither properly factor into whether the physician had committed an intentional tort.[254]

B. Criminal and Regulatory Sanctions

In addition to civil sanctions, health care providers that make unilateral decisions to stop LSMT may be subject to an array of criminal and regulatory sanctions, including charges for patient neglect,[255] adverse peer review,[256] and even murder.[257]

1. Murder

For health care providers, withholding or withdrawing LSMT, even with consent, and thereby facilitating death, was once considered a serious crime.[258] A health care provider's omission to continue treatment fits the literal definition of murder: an intentional act done with the knowledge that the patient would die.[259] However, that concept was eventually rejected in both cases and statutes.[260]

---

253. *Id.* at 629 (emphasis added).

254. *Id.*; *see also Wendland*, 574 N.W.2d at 331 ("That a terminally ill victim would have died on Tuesday, the next day, does not prevent the defendant's conduct from being a cause of his death on Monday, but would obviously be quite relevant to the question of damages.").

255. *See., e.g.*, Preston v. Meriter Hosp., Inc., 700 N.W.2d 158, 163 (Wis. 2005) (citing WIS. STAT. § 940.295(1)(J) (1997-1998)).

256. *See, e.g.*, Warthen v. Toms River Comty. Mem'l Hosp., 488 A.2d 229, 230 (N.J. Super. Ct. App. Div. 1985) (reviewing termination of nurse's employment for refusing to administer dialysis to terminally ill patient); Irene Hurst, *The Legal Landscape at the Threshold of Viability for Extremely Premature Infants: A Nursing Perspective, Part I*, 19 J. PERINATAL & NEONATAL NURSING 155, 162 (2005) ("Hospital administrators warned Dr. Jacob that he should reconsider his recommendation [not to resuscitate, contrary to hospital policy] or lose his privileges at the Hospital and be subject to a peer review."); Arthur E. Kopelman et al., *The Benefits of a North Carolina Policy for Determining Inappropriate or Futile Medical Care*, 66 N.C. MED. J. 392, 394 (2005) ("[T]he legislation . . . [gives] assurance that they are not making a decision that will be questioned by their colleagues or other healthcare peers."); Mildred Z. Solomon et al., *Decisions Near the End of Life: Professional Views on Life-Sustaining Treatments*, 83 AM. J. PUB. HEALTH 14, 19 (1993) (describing health care providers' "fear of sanction from peer review boards").

257. DWORKIN, *supra* note 124, at 112; *see* Barber v. Super. Ct., 195 Cal. Rptr. 484, 486 (Cal. Ct. App. 1983).

258. *See* DWORKIN, *supra* note 124, at 112.

259. WAYNE R. LAFAVE, 1 SUBSTANTIVE CRIMINAL LAW, § 6.2 at 435-36 (2d ed. 2003); WAYNE R. LAFAVE, 2 SUBSTANTIVE CRIMINAL LAW, § 14.2 at 428 (2d ed. 2003).

260. *See, e.g.*, UNIFORM HEALTH-CARE DECISIONS ACT (UHCDA) § 13(b) (1993) ("Death resulting from the withholding or withdrawal of health care in accordance with this [Act] does

In *Barber v. Superior Court,* for example, physicians withdrew LSMT, at the family's request, from a patient in a vegetative state likely to be permanent.[261] The Los Angeles District Attorney prosecuted the physicians for murder, but the appellate court rejected the charges because the physicians stopped LSMT *with* the consent of the authorized decision maker.[262]

Cases like *Barber* differ from the futility context in two material respects. First, physicians do not have patient or surrogate consent to cease LSMT.[263] The *Barber* court's holding—that the providers were under no duty to continue ineffective treatment—meant only that the authorized decision maker was under no duty to request such treatment.[264] The court's ruling did not mean that the health care provider had no duty to provide LSMT when requested.[265] Second, in contrast to the *Barber* situation where the surrogates and providers were in agreement, somebody will always be angry enough to complain to the authorities in a futility case.[266]

Unilateral decisions to stop LSMT have thus led to homicide charges[267] and at least one conviction.[268] Admittedly, health care providers are rarely convicted.[269] Yet, they must still expend considerable time and resources in the investigation and litigation process.[270]

---

not for any purpose constitute a suicide or homicide . . . ."); *Barber*, 195 Cal. Rptr. at 493.

261. *Barber*, 195 Cal Rptr. at 486.

262. *Id.* at 486, 493.

263. *See supra* notes 225-26 and accompanying text.

264. *See Barber*, 195 Cal. Rptr. at 486, 493.

265. *Cf.* Marcia Angell, *The Supreme Court and Physician-Assisted Suicide—The Ultimate Right*, 336 NEW ENG. J. MED. 50, 51 (1997) ("[S]witching off the ventilator of a patient dependent on it . . . would be considered homicide if done without the consent of the patient or a proxy.").

266. *See* Ann Alpers, *Criminal Act or Palliative Care? Prosecutions Involving the Care of the Dying*, 26 J. L. MED. & ETHICS 308, 311, 320 (1998) (noting how state authorities typically react only in response to complaints).

267. MASON & LAURIE, *supra* note 77, at 545-47, 582; Fletcher, *supra* note 10, at S:229 (noting one unilateral decision in Virginia led to a charge of homicide and an investigation by the State Board of Medicine).

268. *See* State v. Naramore, 965 P.2d 211, 213, 224 (Kan. Ct. App. 1998) (reversing convictions of murder for failing to resuscitate Mr. Wilt and of attempted murder for over-prescribing pain medication for Ms. Leach). At least one district attorney in Milwaukee, Wisconsin has announced that he will investigate and prosecute deaths caused by the unilateral withdrawal of LSMT. Telephone interview with Dr. Michael Katzoff, Medical Director, Sleep Disorder Center, St. Luke's Medical Center.

269. Joseph P. Pestaner, *End-of-Life Care: Forensic Medicine v. Palliative Medicine*, 31 J.L. MED. & ETHICS. 365, 366 (2003) ("[T]o criminally convict a palliative care provider of a homicidal act essentially requires in admission of guilt.").

270. *See* MARSHA GARRISON & CARL E. SCHNEIDER, THE LAW OF BIOETHICS: INDIVIDUAL AUTONOMY AND SOCIAL REGULATION (TEACHER'S MANUAL) 112-19 (2003). Dr. Naramore, for example, got his conviction reversed on appeal. *Id.* at 118-19. Nevertheless, he suffered a host of adversities, including: (1) losing his staff privileges, (2) losing his medical license, (3) losing his reputation, (4) incarceration pending trial, and (5) difficulty getting another job. *Id.* at 115-

2. Statutory Damages

Statutory damages are far less serious than murder charges, but they are nevertheless significant. State health care decision statutes normally require compliance with a patient's or surrogate's decision.[271] Many states allow for statutory damages and attorney's fees when intentional statutory violations occur.[272]

If a unilateral decision to stop LSMT is intentionally made to interfere with the patient's autonomy in making health care decisions, then that unilateral decision can constitute a statutory violation resulting in fines,[273] disciplinary action, or both.[274] In one case, the patient's estate brought a $2.5 million civil action based on violation of the state Health Care Decisions Act when the University of Virginia Hospital entered a unilateral DNR order.[275]

C. The Chilling Effect of Legal Constraints

While these legal sanctions may not be very probable, they exert a substantial chilling effect on extremely risk averse health providers.[276] As put

---

16, 119.

271. *See, e.g.*, TENN. CODE ANN. § 68-11-1808(b) (2006).

272. *See, e.g.*, HAW. REV. STAT. § 327E-10(a) (Supp. 2005); ME. REV. STAT. ANN. tit. 18-A, § 5-810(a) (1995); MISS. CODE ANN. § 41-41-221(1) (2005); N.J. STAT. ANN. § 26:2H-78(b) (West 1992); N.M. STAT. § 24-7A-10(A) (2006); WYO. STAT. ANN. § 35-22-411(a) (2007); UNIFORM HEALTH-CARE DECISIONS ACT §10 (1993).

273. *See* Marah Stith, *The Semblance of Autonomy: Treatment of Persons with Disabilities under the Uniform Health-Care Decisions Act*, 22 ISSUES L. & MED. 39, 71 (2006) (stating that the civil penalties imposed are "extremely low").

274. *See, e.g.*, IND. CODE ANN. § 16-36-4-21 (LexisNexis 1993) ("A physician who knowingly violates this chapter is subject to disciplinary sanctions . . . as if the physician had knowingly violated a rule adopted by the medical licensing board . . . ."); KAN. STAT. ANN. § 65-28,107(a) (1992); N.J. STAT. ANN. § 26:2H-78(a) (West 1992); Fletcher, *supra* note 10, at S:229 (reporting one unilateral decision led to a an investigation by the State Board of Medicine). On the other hand, some have suggested that if providers follow a process and are in accord with professional guidelines, it is unlikely they will be found to have committed a disciplinary offense. *See* WRONG MEDICINE, *supra* note 37, at 89–94 (suggesting that it is a legal myth that providers will always be subject to legal liability for stopping LSMT).

275. Amended Motion for Judgment at 1-3, Bryan v. Rectors & Visitors of the Univ. of Va., No. CL95-060 (Fauquier County, Va. Cir. Ct. Nov. 27, 1995).

276. *See* Pope & Waldman, *supra* note 9, at 170-85; *see also* GARRISON & SCHNEIDER, *supra* note 270, at 70 ("Doctors egregiously over-estimate the risks of being sued by their patients."); SCHNEIDERMAN, *supra* note 57, at 126-28; ZUSSMAN, *supra* note 51, at 181 ("[U]nfortunately, because of a fear of being sued at a later date, most physicians really are willing to provide every available technology to a patient . . . ."); Kapp, *supra* note 122, at 232 ("[L]aw-related anxieties . . . are palpable, powerful influences on . . . medical care . . . ."); Rowland, *supra* note 214, at 307 ("Legal considerations are of paramount concern when discussing the discontinuation of care."); Carl E. Schneider, *Regulating Doctors*, 29 HASTINGS CENTER REP., July-Aug. 1999, at 21; Connie Zuckerman, Milbank Memorial Fund, *End-of-Life*

by Professors Robert Weir and Larry Gostin, "Because the professional responsibility of hospital attorneys is to protect the hospital's legal and financial interests, they are frequently inclined to give advice on cases that is unduly conservative . . . ."[277] This ultra-cautious approach is no less true in the context of futility disputes.[278] In 1993, the National Center for State Courts observed that there was "no consensus . . . on the legal ramifications associated with [futility]."[279] Before statutory authorization for unilateral decision making in the mid-1990s,[280] legal uncertainty was rampant and the fear of liability discouraged most institutions from adopting futility policies.[281]

Nonetheless, by the early 1990s, a few hospitals had formally adopted futility policies.[282] Yet even these hospitals never fully implemented the

---

Care and Hospital Legal Counsel: Current Involvement and Opportunities for the Future 3 (1999), *available at* http://www.milbank.org/end.html ("Legal considerations . . . strongly influence how clinicians think about end-of-life care.").

277. Robert F. Weir & Larry Gostin, *Decisions to Abate Life-Sustaining Treatment for Nonautonomous Patients: Ethical Standards and Legal Liability for Physicians After* Cruzan, 264 JAMA 1846, 1846 (1990); *see also* Alan J. Weisbard, *Defensive Law: A New Perspective on Informed Consent*, 146 ARCHIVES INTERN. MED. 860, 860 (1986) ("[T]he lawyer's . . . advice is likely to become ultracautious and may tend to conflict with the responsible practice of medicine . . . .").

278. *See* COORDINATING COUNCIL ON LIFE-SUSTAINING MEDICAL TREATMENT DECISION MAKING BY THE COURTS, GUIDELINES FOR STATE COURT DECISION MAKING IN LIFE-SUSTAINING MEDICAL TREATMENT CASES 147 (2d ed. 1993).

279. *Id.*

280. *See infra* Part IV.

281. *See, e.g.*, WRONG MEDICINE, *supra* note 37, at 32 ("Physicians often . . . fear the legal consequences of forgoing treatment . . . ."); Fletcher, *supra* note 10, at S:229. Professor Fletcher recalls, "On coming to the University of Virginia in 1987, I observed many clinicians overtreating hopelessly ill patients primarily due to fears of legal liability. Also, clinicians were acutely aware of the lack of legal backing if they refused to acquiesce . . . ." *Id.*; *see* Moldow, *supra* note 12, at 3 ("Fear of legal action has previously discouraged many institutions from adopting policies in the area of medical futility . . . ."); Sibbald et al., *supra* note 44, at 1203 (reporting from a survey of ICUs: "When participants were asked why they followed the instructions of families or substitute decision makers instead of doing what they feel is appropriate, almost all cited a lack of legal support."); Weiser, *supra* note 45, at A1 (describing how physicians' wanted to unilaterally withdraw LSMT from severely ill infant, but were prevented by hospital's rules); *see also* ZUSSMAN, *supra* note 51, at 178 ("'I wish,' she concluded, 'the family didn't have the final say. But in 1987 they do . . . .'"). Marshall Kapp argues that the legal risks in the early 1990s were not serious, yet concedes that physicians had "overblown anxiety." Kapp, *supra* note 122, at 175; *see also* Hall, *supra* note 246, at 119 ("[T]o the extent that a crisis is in fact widely perceived, it has the quality of a self-fulfilling prophecy . . . .").

282. *See, e.g.*, Fine, *supra* note 92, at 62 (noting early futility policies); Fletcher, *supra* note 10, at S:228 ("Massachusetts General Hospital (MGH) was the first to experiment with an approach to futility disputes . . . that gave institutional backing to physicians to write a DNR order over the objections of a surrogate . . . ."); Hudson, *supra* note 66, at 26 (noting Santa Monica adopted a policy in 1991); Schneiderman & Capron, *supra* note 243, at 526

policies by actually taking unilateral action to stop LSMT requested by a patient or surrogate.[283] Providers understood that an institutional policy did little to alleviate uncertainty about the legal implications of unilaterally stopping LSMT.[284]

(referencing meeting in 1998 to discuss futility policies of twenty-six hospitals).

283. Fine et al., *supra* note 100, at 1221 (describing that before the Texas statute, "[i]n ~80% of such [futility] cases, the ethics consultants were able to persuade families . . . . However, in the other 20% of cases, families insisted on continued [LSMT], and physicians complied, being unwilling to subject themselves to legal jeopardy by overruling the family/surrogate"); Fine & Mayo, *supra* note 84, at 744 ("It is unclear how effective such guidelines could be in the face of legal uncertainty. Even when ethics committees agreed that treatment was futile, treating physicians were generally unwilling to withdraw life-sustaining treatment . . . ."); Amir Halevy & Amy L. McGuire, *The History, Successes and Controversies of the Texas "Futility" Policy*, HOUSTON LAW., May-June 2006, at 38, *available at* http://www.thehoustonlawyer.com/aa_may06/page38.htm ("In spite of its adoption as hospital policy . . . no cases went through the entire process . . . . The most likely explanation is that residual legal uncertainty regarding the policy still lingered."); Hudson, *supra* note 66, at 26 (noting that the hospital had "never reached the last two steps in the [futility] process"); Rivin, *supra* note 30, at 390 ("Despite the recommendations of the physicians and the ethics committee, the [Santa Monica] hospital refused to discontinue life support for fear of lawsuit."); Anna V. Schlotzhauer & Bryan A. Liang, *Definitions of Death*, *in* HEALTH LAW AND POLICY: A SURVIVAL GUIDE TO MEDICOLEGAL ISSUES FOR PRACTITIONERS 287, 291 (2000) ("[N]othing can be done in cases where families of PVS patients seek to continue treatment indefinitely . . . ."); Swig et al., *supra* note 173, at 1218 ("[D]espite a policy that allowed them to do otherwise, . . . physicians at San Francisco General Hospital usually offered CPR to patients who they thought were unlikely to benefit.").

284. *See generally* Cerminara, *supra* note 48, at 327 ("[G]ood process . . . will not insulate a decision maker from being overturned in court . . . ."); Fine, *supra* note 92, at 63 ("Guidelines in the face of legal uncertainty, however, were not particularly effective. . . . [F]ew physicians were willing to limit such treatment in the face of potential lawsuits from families who disagreed."); Fine, *supra* note 100, at 1221 (noting when families insisted on continued LSMT "physicians complied, being unwilling to subject themselves to legal jeopardy by overruling the family/surrogate"); Flamm, *supra* note 10, at 4 ("[T]he previous ambiguity of legal consequences often prevented clinicians from fulfilling ethical obligations against providing medically inappropriate care."); Halevy & McGuire, *supra* note 283, at 38 ("[R]esidual legal uncertainty regarding the policy still lingered."); Kopelman et al., *supra* note 256, at 393 ("Uncertainty about the legal implications of acting against the patient's or surrogate's wishes often prevents physicians from taking that [unilateral] step, despite agreement among all or almost all clinicians."); Rivin, *supra* note 30, at 393 (noting that even when physicians thought a case was futile, they were unwilling to invoke the futile care policy for "fear of a lawsuit"); Solomon et al., *supra* note 256, at 19 (reporting physician uncertainty about legal standards for withdrawing treatment); Swig et al., *supra* note 173, at 1218 (citing "legal considerations" as a possible explanation for why physicians did not utilize their hospital's futility policy); Belluck, *supra* note 38, at 22 ("In the absence of laws like Texas's, hospitals often accede to a family's wishes because they fear being sued."); Burling, *supra* note 66, at A1 ("The weak point of virtually all policies is that hospital leaders fear they would lose a lawsuit if they denied care demanded by a family."); *cf.* COMM. ON PALLIATIVE AND END-OF-LIFE CARE FOR CHILDREN AND THEIR FAMILIES, BD. ON HEALTH SCI. POL'Y, WHEN CHILDREN DIE: IMPROVING PALLIATIVE AND

To alleviate this uncertainty, some hospitals sought judicial permission to implement their futility policies.[285] Declaratory judgments were designed to address such cases of uncertainty.[286] But this judicial approach suffered from two serious drawbacks. First, given the time and resources required, it was perceived as generally unworkable.[287] Second, even if hospitals were willing to invest the time and resources, courts have consistently declined to authorize providers to implement their futility policies.[288]

Consequently, providers complied with requests for treatment that they considered inappropriate, because they recognized that surrogates had a veto authority over their judgment.[289] In light of all the legal constraints and risks, providers wanted legal protection before taking any unilateral action.[290]

END-OF-LIFE CARE FOR CHILDREN AND THEIR FAMILIES 322 (Marilyn J. Field & Richard E. Behrman, eds., 2003) [hereinafter WHEN CHILDREN DIE] ("[T]he findings of an ethics committee have no legal standing and cannot be used alone as the basis for termination of life support."); Brett, *supra* note 70, at 289 (noting the "pragmatic problem with policies that confer no legal protection"); Schneiderman & Capron, *supra* note 243, at 525 ("[T]he *Baby K* decision . . . had a chilling effect on hospitals' willingness to implement futility policies.").

285. *See, e.g.*, *In re* Farrell, 529 A.2d 404, 406-07 (N.J. 1987); *In re* Quinlan, 355 A.2d 647, 669 (N.J. 1976).

286. JAMES WM. MOORE, 12 MOORE'S FEDERAL PRACTICE § 57 (3d ed. 2007).

287. *Cf. Farrell*, 529 A.2d at 415 (resolving end-of-life disputes through a judicial process will "take too long"); *Quinlan*, 355 A.2d at 669 ("[A] practice of applying to a court to confirm such decisions would generally be inappropriate . . . because that would be a gratuitous encroachment on the medical profession's field of competence . . . [and] impossibly cumbersome.").

288. *See, e.g.*, *In re* Baby K, 16 F.3d 590, 592 (4th Cir. 1994) (denying motion by Fairfax Hospital seeking declaratory judgment to withdraw treatment from anencephalic infant); *Judge Affirms Husband's Right to Continue Wife's Treatment*, 53 BIOLAW §12-6, at U:2161 (Aug.-Sept. 1991) (noting that "a county court judge . . . refused the doctors' request to appoint an independent conservator to decide the patient's fate"); Frank Bruni, *A Fight over Baby's Dignity and Death: Parents Sue Hospital Over Shutdown of Life Support Equipment*, N.Y. TIMES, Mar. 9, 1996, at A6 ("[W]hen hospitals go to court for permission to terminate treatment of a patient over the objections of family, courts seldom give consent."); *cf.* Hoffman & Schwartz, *supra* note 218, at 37 (noting that some court have decided that futility issues "should be addressed by the legislative rather than the judiciary").

289. *See* sources cited *supra* note 284.

290. *See* Fletcher, *supra* note 10, at S:231 ("The framers of such futility guidelines would also be well-advised to seek amendments to existing health care legislation that strengthen the authority of clinicians and health care organizations to resolve such disputes."); *id.* at S:229 ("[A]ction was necessary in the Virginia legislature to assure physicians of legal backing if they refused, in certain circumstances, to acquiesce to demands for overtreatment."); Carol Isackson, *Futile Treatment: The Need for Legislation and Uniform Policies*, 9 HEALTH CARE L. MONTHLY, 7, 10 (Oct. 1994) ("In order to protect providers from arbitrary decisions . . . legislation should be enacted . . . ."); Halevy & McGuire, *supra* note 283, at 38 ("Many institutions were interested in pursuing policies that would allow physicians to refuse . . . [but] the legal and ethical uncertainties . . . discouraged institutions from proceeding alone."); Susan Jacoby, *The Schiavo Factor: Now the States Are Rushing to Decide Who Decides*, AARP BULLETIN, May

IV. UNILATERAL DECISION STATUTES

Providers soon got the legal protection for unilateral decision making that they were seeking. Beginning in the early 1990s, a significant number of states began enacting legislation permitting health care providers to unilaterally refuse to provide LSMT that they considered to be medically inappropriate.[291]

A. The Uniform Health-Care Decisions Act

Most notable among the unilateral decision statutes is the Uniform Health-Care Decisions Act (UHCDA).[292] The UHCDA is notable for three reasons. First, it has a significant and growing prevalence.[293] It has now been adopted in ten states, more than any other unilateral decision statute.[294] Second, the

---

2005, *available at* http://www.aarp.org/bulletin/yourhealth/the_schiavo_factor.html ("In states without such [futile care] laws, doctors frequently comply with the family's wishes for fear of being sued . . . ."); Weiser, *supra* note 45, at A1 (reporting how a doctor in the Baby Rena case "felt the time had come to change the rules" to give doctors the authority they need in futile cases).

291. *See infra* notes 298-307; *see also* Maggie Datiles, *The Rising Role of Advance Directives in Protecting the Sanctity of Human Life, in* AMERICANS UNITED FOR LIFE, DEFENDING LIFE 2008: A STATE-BY-STATE LEGAL GUIDE TO ABORTION, BIOETHICS AND THE END OF LIFE 511, 512 (2008) ("The majority of states provide that physicians and healthcare facilities may decline to comply [with requests for LSMT]"); Patrick Moore, *An End-of-Life Quandary in Need of a Statutory Response: When Patients Demand Life-Sustaining Treatment that Physicians are Unwilling to Provide,* 48 B.C. L. REV. 433 (2007); Monica Sethi, *A Patient's Right to Direct Own Health Care vs. a Physician's Right to Decline to Provide Treatment,* 29 BIFOCAL, Dec. 2007, at 21 (examining "provisions from all 50 states regarding the various reasons for which a health care provider may refuse to comply with a patient's demand for treatment").

292. *Uniform Health-Care Decisions Act: National Conference of Commissioners on Uniform State Laws,* 22 ISSUES L. & MED. 83, 83-97 (2006) [hereinafter *Uniform Act*].

293. *Id.* at 83.

294. The National Conference of Commissioners on Uniform State Laws' (NCCUSL) website identifies only eight states as having adopted the UHCDA. A Few Facts About the Uniform Health-Care Decisions Act, http://www.nccusl.org/nccusl/uniformact_factsheets/uniformacts-fs-uhcda.asp (last visited Oct. 21, 2007) [hereinafter NCCUSL]. However, the legislative history of both the California and Tennessee statutes confirms that they were largely derived from the UHCDA as well. *See Health Care Decisions for Adults Without Decisionmaking Capacity, Bill Analysis of A.B. 891 Before the Assem. Comm. on the Judiciary,* at 5 (Apr. 15, 1999) (noting that the bill is "[d]rawing heavily on the [UHCDA]."); *Health Care Decisions: Durable Power of Attorney, Bill Analysis of A.B. 891 Before the S. Comm. on the Judiciary,* at 2 (July 13, 1999) ("The provisions of the proposed Health Care Decisions Law (HCDL) are drawn heavily from the Uniform Health Care Decisions Act (1993), and implement major parts of the Commission's recommendation[s]."); Charles M. Key & Gary D. Miller, *The Tennessee Health Care Decisions Act A Major Advance in the Law of Critical Care Decision Making,* 40 TENN. B.J. 25, 28 (2004).

UHCDA has provisions specifically designed to handle futility disputes.[295] Third, the UHCDA is a reasonably comprehensive statute, broadly authorizing health care providers to take unilateral action in all types of futility disputes.[296]

1. Prevalence of the UHCDA

The National Conference of Commissioners on Uniform State Laws completed drafting the UHCDA in 1993.[297] Over the next twelve years, it was adopted in the following ten states: New Mexico (1995),[298] Maine (1995),[299] Delaware (1996),[300] Alabama (1997),[301] Mississippi (1998),[302] California (1999),[303] Hawaii (1999),[304] Tennessee (2004),[305] Alaska (2004),[306] and Wyoming (2005).[307] Together, these ten states comprise about one-fifth of the U.S. population.[308]

Several other states have recently considered adopting the UHCDA, including its unilateral decisions provisions.[309] It is likely that the UHCDA will continue to be adopted or will otherwise influence health care decision making law in other states.[310]

---

295. *See Uniform Act*, *supra* note 292, at 84.

296. *Id.*

297. *Id.* at 83.

298. N.M. STAT. §§ 24-7A-1 to -18 (2000).

299. ME. REV. STAT. ANN. tit. 18-A, §§ 5-801 to -817 (1995).

300. DEL. CODE ANN. tit. 16, §§ 2501–2518 (2003).

301. ALA. CODE §§ 22-8A-1 to -14 (LexisNexis 2006).

302. MISS. CODE ANN. §§ 41-41-201 to -229 (2005).

303. CAL. PROB. CODE §§ 4600-4806 (West Supp. 2007).

304. HAW. REV. STAT. §§ 327E-1 to -16 (Supp. 2005).

305. TENN. CODE ANN. §§ 68-11-1801 to -1815 (2006).

306. ALASKA STAT. §§ 13.52.010 to .395 (2006).

307. WYO. STAT. ANN. §§ 35-22-401 to -416 (2007).

308. U.S. Census Bureau, http://factfinder.census.gov (last visited Oct. 21, 2007) (extrapolating total population from 2006 estimates for each of the ten states).

309. *See, e.g.*, Utah S.B. 75 (effective Jan. 1, 2008) (to be codified at UTAH CODE ANN. §§ 75-2a-1103(6)(b) & 75-2a-1114) (based on the UHCDA); *Uniform Health Care Decisions Act: Hearing on S.B. 229 Before the S. Comm. on the Judiciary*, 60th Legis. (Mont. 2007). Unfortunately, the Montana bill died in standing committee on April 27, 2007. *See* http://laws.leg.mt.gov/pls/laws07/law0203w$.startup (search "Bill Type and Number" for "S.B. 229").

310. *See, e.g.*, REP. TO VERMONT ATTORNEY GENERAL WILLIAM H. SORRELL FROM THE COMMS. OF THE ATTORNEY GENERAL'S INITIATIVE ON END OF LIFE CARE 15 (2005) (recording recommendations of committees reached by reviewing UHCDA); ADVANCE DIRECTIVES IN NEW HAMPSHIRE: A STATUTORY REVIEW & SURVEY OF CURRENT ISSUES 1 (2000) (considering advance care planning); *see also* David M. English & Alan Meisel, *Uniform Health-Care Decisions Act Gives New Guidance*, 21 EST. PLAN. 355, 357 (1994) ("It is likely that the Act will serve as an influential model for many years to come.").

2. Purpose of the UHCDA

Some have suggested that the UHCDA's unilateral decision provisions were not written in contemplation of futility disputes, but rather exclusively "in contemplation of the opposite situation" in which the family wants to reject treatment but the health care provider wants to continue.[311] Indeed, the UHCDA does focus on patient autonomy and the empowerment of patients and surrogates.[312]

Nevertheless, the legislative history of the Uniform Act clearly shows this charge to be untrue.[313] The UHCDA commissioners specifically contemplated and sought to relieve health care providers of any obligation to provide inappropriate treatment.[314] Moreover, the very logic of the UHCDA compels an interpretation that authorizes providers to unilaterally terminate LSMT.[315]

---

311. *See, e.g.*, ROBERT POWELL CENTER FOR MEDICAL ETHICS, NATIONAL RIGHT TO LIFE COMMITTEE, WILL YOUR ADVANCE DIRECTIVE BE FOLLOWED?, at 8 n.* (Apr. 15, 2005), *available at* http://www.nrlc.org/euthanasia/AdvancedDirectives/ReportRevised2007.pdf. Indeed, some laws do allow only unilateral decisions to *provide* treatment. For example, until this year, Pennsylvania provided immunity only for *provision* of treatment contrary to a patient's living will. *Compare* 20 PA. STAT. ANN. § 5409(c) ("[T]he provision of life-sustaining treatment to a declarant shall not subject a health care provider to criminal or civil liability or administrative sanction for failure to carry out the provisions of a declaration."), *with* 20 PA. STAT. ANN. § 5431(a)(6) (stating that providers will not be subject to criminal or civil liability, or administrative sanctions for "[r]efusing to comply with a direction or decision of an individual [if] based on a good faith belief that compliance with the direction or decision would be unethical" or would result in baseless medical treatment); *see also* MINN. STAT. ANN. § 145C.11(2) (West 1998) (addressing specifically the provision of treatment, but not addressing a provider's refusal of treatment contrary to decision of the agent).

312. *See Uniform Act*, *supra* note 292, at 83.

313. *See, e.g.*, Nat'l Conference of Comm'rs on Uniform State Laws, Proceedings in Comm. of the Whole, Uniform Health-Care Decisions Act, July 30, 1993, at 33 (statement of Comm'r David M. English) ("[T]hey are not obligated to provide me with the type of state-of-the-art, all-out care . . . [because] if a competent patient couldn't ask for it, then an agent couldn't ask for it either."); *id.* at 183 (statement of Comm'r M. King Hill) ("We do not want to impose upon physicians or other health-care providers . . . the obligation to provide treatment that will not be effective."); Nat'l Conference of Comm'rs on Uniform State Laws, Proceedings in Comm. of the Whole, Uniform Health-Care Decisions Act, Aug. 2, 1993, at 268–69 (statement of Comm'r Richard V. Wellman) ("[Provision to decline treatment] is here as a . . . needed qualification of duties imposed on health-care providers to follow instructions and directions by surrogates and others."); *id.* at 269-70 (statement of Comm'r M. King Hill) ("medically ineffective" refers to costs—"This says to the physician that you don't have to institute some new radical $200,000 procedure if it's only going to keep the patient alive for two or three months, even though there may be many articles in the journals that say that's an accepted health-care standard for a [twenty-two] year old.").

314. *See* sources cited *supra* note 313.

315. *See* Stith, *supra* note 273, at 62 (arguing that the UHCDA gives physicians the right to "ignore desired but 'medically ineffective' treatment" and also contains "normative aspects that cause it to favor death-hastening physician judgments: Only continuance of care can be

3. Comprehensiveness of the UHCDA

The UHCDA requires that health care providers generally comply with patient and surrogate health care decisions.[316] But it also makes clear that a health care provider's obligation to comply with a treatment request "is not absolute."[317] A health care provider or health care institution may decline to comply with an individual instruction that requires "medically ineffective health care" or "health care contrary to generally accepted health-care standards."[318] A health care provider may also decline to comply for "reasons of conscience."[319]

---

ineffective."); *id.* at 63 ("The UHCDA's preference for the ability to discontinue care could not be clearer.").

316. UNIFORM HEALTH-CARE DECISIONS ACT § 7(d) (1993).

317. *Id.* at Prefatory Note ("The obligation to comply is not absolute, however. A health-care provider or institution may decline to honor an instruction or decision for reasons of conscience or if the instruction or decision requires the provision of medically ineffective care or care contrary to applicable health-care standards."); *id.* at § 4 cmt. ("[H]ealth-care instructions . . . are binding . . . subject to exceptions specified in Section 7(e)–(f), on the individual's health-care providers."); *id.* § 7 cmt. ("Not all instructions or decisions must be honored, however.").

318. *Id.* §§ 7(f), 13(d); *accord* ALA. CODE § 22-8A-8(a) (LexisNexis 2006); ALASKA STAT. § 13.52.060(f) (2006); CAL. PROB. CODE §§ 4654, 4735 (West Supp. 2007); DEL. CODE ANN. tit. 16, § 2508(f) (2003); HAW. REV. STAT. § 327E-7(f) (Supp. 2005); ME. REV. STAT. ANN. tit. 18-A, § 5-807(f) (1995); MISS. CODE ANN. § 41-41-215(6) (2005); N.M. STAT. §§ 24-7A-7(F), 24-7A-13(D) (2000); TENN. CODE ANN. § 68-11-1808(e) (2006); WYO. STAT. ANN. §§ 35-22-408(f), 35-22-414(d) (2007).

319. *See* UNIFORM HEALTH-CARE DECISIONS ACT § 7(e); *accord* CAL. PROB. CODE § 4734(a) (West Supp. 2007) ("A health care provider may decline to comply with an individual health care instruction or health care decision for reasons of conscience."). The conscience exception is well established in the reverse situation, permitting providers to refuse to comply with patient or surrogate requests to stop treatment. *See, e.g.*, Morrison v. Abramovice, 253 Cal. Rptr. 530, 534 (Cal. Ct. App. 1988) ("The prevailing viewpoint among medical ethicists appears to be that a physician has the right to refuse on personal moral grounds to follow a conservator's direction to withhold life-sustaining treatment . . . ."); Brophy v. New Eng. Sinai Hosp., 497 N.E.2d 626, 639 (Mass. 1986) (stating providers should not feel compelled "to take active measures which are contrary to their view of their ethical duty toward their patients"). *But see* Gray v. Romeo, 697 F. Supp. 580, 591 (D.R.I. 1988) (finding providers must acknowledge a patient's "right of self-determination" despite the provider's own personal objections). The conscience exception applies to both individual and institutional providers, though institutions must give notice. *See, e.g.*, *In re* Jobes, 529 A.2d 434, 450 (N.J. 1987) (stating that nursing home should have given patient's family notice of their policy regarding artificial feeding); PRESIDENT'S COMMISSION FOR THE STUDY OF ETHICAL AND LEGAL ISSUES IN TREATMENT DECISIONS, DECIDING TO FORGO LIFE-SUSTAINING TREATMENT: A REPORT ON THE ETHICAL, MEDICAL, AND LEGAL ISSUES IN TREATMENT DECISIONS 44 (1983). However, the conscience exception is thought to have limited applicability in the futility context because the provider's values are not the central concern. *See* Gampel, *supra* note 59, at 101 ("[T]he values at stake in that judgment are unlikely to be as central to an individual HCP, or to the medical

The UHCDA's authorization of unilateral decisions is comprehensive in at least four important respects. First, the UHCDA permits the provider to decline to comply with a treatment request concerning *any* type of treatment.[320] While some state statutes only authorize unilateral decisions with respect to CPR,[321] the UHCDA authorizes unilateral decisions with respect to CPR, mechanical ventilation, artificial nutrition and hydration, or any other type of medical intervention.[322]

Second, the UHCDA is comprehensive in that it authorizes unilateral decisions even when the patient or surrogate has made an explicit and affirmative request for treatment or has demonstrated explicit and vehement opposition.[323] On the contrary, some state statutes authorize unilateral decisions only where the patient's preferences are unknown—where the patient has no available advance directive or surrogate.[324]

Third, the UHCDA leaves the provider with substantial discretion to determine the circumstances under which treatment is inappropriate.[325] The UHCDA permits providers to decline to comply with requests for treatment that would be medically ineffective.[326] "Medically ineffective" treatment is defined as treatment that would not provide any "significant benefit."[327] However, the UHCDA allows the health care provider broad discretion to determine whether the benefit achievable by a treatment is "significant."[328]

---

profession, as the values that tell against acts such as assisted suicide or abortion."). Since many futility cases are driven by providers' desire to avoid patient suffering, the conscience exception may soon play a greater role. *Cf.* Mark R. Wicclair, *Conscientious Objection in Medicine*, 14 BIOETHICS 205, 216–17 (2000) ("The condition is that an appeal to conscience has significant moral weight only if the core ethical values on which it is based correspond to one or more core values in medicine."). This is especially true because of the increasing breadth and use of conscience clauses in medicine. *See, e.g.*, Maxine M. Harrington, *The Ever-Expanding Health Care Conscience Clause: The Quest for Immunity in the Struggle Between Professional Duties and Moral Beliefs*, 34 FLA. ST. U. L. REV. 779 (2007).

320. *See* UNIFORM HEALTH-CARE DECISIONS ACT §§ 7(e)-(f), 13(d).

321. *See infra* note 374 and accompanying text.

322. *See* UNIFORM HEALTH-CARE DECISIONS ACT § 1(6)(i)-(iii).

323. *See id.* §§ 7(e)-(f), 13(d).

324. *See infra* notes 376–77.

325. *See* UNIFORM HEALTH-CARE DECISIONS ACT §§ 7(f), 13(d).

326. *Id.* § 7(f).

327. *Id.* §7(f) cmt. ("'Medically ineffective health care,' as used in this section, means treatment which would not offer the patient any significant benefit."). As adopted, one UHCDA state defines "medically ineffective treatment" more tightly, as medical procedures which, "to a reasonable degree of medical certainty, . . . will not: (1) Prevent or reduce the deterioration of the health of an individual; or (2) Prevent the impending death of an individual." DEL. CODE ANN. tit. 16, § 2501(m) (2003).

328. *See* Ferguson, *supra* note 16, at 1220–21 ("The UHCDA provides a mere framework . . . giv[ing] only broad platitudes . . . [with] sections [that] seemingly create an open-ended excuse for a physician to withdraw treatment . . . .").

Fourth, as adopted in several states, the UHCDA explicitly confers immunity on providers who exercise the unilateral decision provisions in good faith.[329] California, for example, provides that "[a] health care provider . . . acting in good faith and in accordance with generally accepted health care standards applicable to the health care provider . . . is not subject to civil or criminal liability or to discipline for unprofessional conduct for any actions in compliance with this division."[330]

4. Operation of the UHCDA

Most end-of-life decision making laws are designed to work extra-judicially.[331] The UHCDA is no exception.[332] Providers need not go to court to make a unilateral decision.[333] They need only comply with the following process outlined in the UHCDA.[334]

If the provider is going to decline to comply with a health care decision under the UHCDA, the provider must first inform the patient or surrogate.[335]

---

329. *See* UNIFORM HEALTH-CARE DECISIONS ACT § 9. While UHCDA itself confers immunity for several categories of conduct, it does not confer immunity for complying with the unilateral decision provisions. *See id.*

330. CAL. PROB. CODE § 4740 (West Supp. 2007); *accord* ALA. CODE § 22-8A-8(a) (LexisNexis 2006) ("shall not be liable for such refusal"); DEL. CODE ANN. tit. 16, § 2510(a)(5) (2003); ME. REV. STAT. ANN. tit. 18-A, § 5-809(a)(2) (1995); N.M. STAT. § 24-7A-9(A)(4) (2000); WYO. STAT. ANN. § 35-22-410(a)(v) (2007). Providers in other states may qualify for immunity under related statutes. *See, e.g.*, HAW. REV. STAT. § 663-1.7(b) (Supp. 2005) ("There shall be no civil liability for any member of an . . . ethics committee, or . . . for any acts done in furtherance of the purpose for which the . . . ethics committee . . . was established . . . .").

331. *See generally* CAL. PROB. CODE § 4650(c) (West Supp. 2007) ("In the absence of a controversy, a court is normally not the proper forum in which to make health care decisions, including decisions regarding life-sustaining treatment."); *In re* Rosebush, 491 N.W.2d 633, 637 (Mich. Ct. App. 1992) ("[T]he decision-making process should generally occur in the clinical setting without resort to the courts . . . ."); *In re* Quinlan, 355 A.2d 647, 669 (N.J. 1976) (suggesting that applying to a court for authority to stop LSMT is generally inappropriate, being both cumbersome and an encroachment on the medical profession); Jesse A. Goldner et al., *Responses to Medical Futility Claims*, *in* HEALTH LAW HANDBOOK 404 (Alice Gosfield ed., 1997) ("The final key background theme is that the courts express a clear preference for limiting judicial involvement in these questions."). *But see* Maureen Kwiecinski, *To Be or Not to Be, Should Doctors Decide? Ethical and Legal Aspects of Medical Futility Policies*, 7 MARQ. ELDER'S ADVISOR 313, 353-55 (2006) (suggesting that judicial review should be required).

332. *See* UNIFORM HEALTH-CARE DECISIONS ACT, Prefatory Note ("[T]he Act is in general to be effectuated without litigation . . . ."); *id.* § 14 cmt. ("[T]he provisions of the Act are in general to be effectuated without litigation . . . .").

333. *See id.* at Prefatory Note, § 14 cmt.

334. While neither the UHCDA itself nor the statutes that are modeled on it make any reference to ethics committees, institutional policies almost invariably provide a role for an institutional committee. Most providers supplement the process in their state's statute with that outlined by the AMA.

335. UNIFORM HEALTH-CARE DECISIONS ACT § 7(g)(1); *see also id.* § 7(a) ("Before

This is a sensible requirement, since mutual agreement is reached in most cases.[336] Furthermore, notice gives the surrogate an opportunity to either seek review of the decision or transfer the patient to another physician or institution or both.[337] Informing the surrogate addresses the notorious lack of transparency associated with unilateral DNR orders in the 1980s.[338]

After the provider informs the patient or surrogate of their refusal to comply with the treatment request, the provider must then try to transfer the patient to another provider who is willing to comply with the treatment request.[339] The UHCDA states:

> [U]nless the patient or person then authorized to make health-care decisions for the patient refuses assistance, [the provider shall] immediately make all reasonable efforts to assist in the transfer of the patient to another health-care provider or institution that is willing to comply with the instruction or decision.[340]

Thus, prior to transfer, the provider must comply with the treatment request.[341]

---

implementing a health-care decision made for a patient, a supervising health-care provider, if possible, shall promptly communicate to the patient the decision made and the identity of the person making the decision.").

336. *See supra* note 91 and accompanying text.

337. *See* UNIFORM HEALTH-CARE DECISIONS ACT § 7(g).

338. Where providers were unable to write a unilateral DNR order and CPR was considered inappropriate, providers were known to affix color dots to the patient's wristband or write "N.T.B.R." (Not to Be Resuscitated) in pencil on the chart to be erased after the patient died. *See* Hoffman, *supra* note 46, at 6; Kapp, *supra* note 122, at 173. Some providers did a "Hollywood Code" or "Show Code" in which they performed a half-hearted or mock resuscitation. George P. Smith, II, *Euphemistic Codes and Tell-Tale Hearts: Humane Assistance in End-of-Life Cases*, 10 HEALTH MATRIX: J. L.-MED. 175, 184 (2000); Rosenthal, *supra* note 66, at A1. Still other providers performed "Slow Codes" in which they moved very slowly. *See* Smith, *supra*, at 180; Editorial, *Slow Codes, Show Codes and Death*, N.Y. TIMES, Aug. 22, 1987, at A26; *cf. In re* Quinlan, 355 A.2d 647, 657 (N.J. 1976) (discussing the medical practice of "judicious neglect").

339. UNIFORM HEALTH-CARE DECISIONS ACT § 7(g)(2)–(3); *see, e.g.*, CAL. PROB. CODE § 4736(b) (West Supp. 2007). It is unclear whether the care in this interim period can be billed once a formal decision has been made that the care is inappropriate. *Cf.* 42 U.S.C. § 1320c-5(a) (2007). The United States Code uses the following language:

> It shall be the obligation of any health care practitioner and any other person (including a hospital or other health care facility, organization, or agency) who provides health care services for which payment may be made (in whole or in part) [by Medicare or Medicaid] under this chapter, to assure, to the extent of his authority that services or items ordered or provided . . . will be provided economically and only when, and to the extent, medically necessary."

Id.

340. UNIFORM HEALTH-CARE DECISIONS ACT § 7(g)(2).

341. *Id.* § 7(g)(2).

Interestingly, however, the UHCDA does not specifically address what happens if transfer is not possible.[342] This is significant because those patients, for whom providers deem LSMT inappropriate, typically cannot be transferred.[343] There is almost never a facility available and willing to take such patients.[344] Even in cases where a facility is available, these patients are often not sufficiently stable to be transferred.[345]

The UHCDA requires only that the provider make "all reasonable efforts" to transfer the patient.[346] If the provider is unable to transfer the patient, then

---

342. *See id.* § 7.

343. *See, e.g.*, *In re* Baby K, 16 F.3d 590, 593 (4th Cir. 1994) (noting no hospitals with a pediatric intensive care unit (PICU) were willing to accept Stephanie Keene from Fairfax Hospital, although she was temporarily transferred to a nursing home); Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1073 (La. Ct. App. 1998) (noting physician sought unsuccessfully to transfer patient); Lee, *supra* note 31, at 487 ("[T]ransfer of care is difficult in a medical futility case . . . ."); Miles, *supra* note 74, at 513 (reporting family of Helga Wanglie unsuccessfully tried to transfer her); Keith Shiner, *Medical Futility: A Futile Concept?*, 53 WASH. & LEE L. REV. 803, 845–46 (1996) (stating that the "transfer option, by itself, is an incomplete solution to the problem of medical futility"); Ackerman, *supra* note 43, at B1 ("Memorial Hermann officials said that other pediatric hospitals they consulted concurred with their treatment plan and decision to discontinue care."); Murphy, *supra* note 43, at 37 (reporting that while the family of Joseph Ndiyob eventually found a Los Angeles hospital willing to accept him, the hospital recanted when it learned he lacked health insurance); Baylor Response, *supra* note 144 ("Ultimately, twelve different health care facilities refused to accept the patient in transfer."); News Release, Memorial Hermann, *Statement to the Media Regarding Kyna Dismuke-Howard*, (May 3, 2005), http://www.memorialhermann.org/newsroom/050305a.htm ("[O]ur physicians contacted premier children's hospitals across the country . . . [but] each one reviewed the facts and refused to accept her transfer."). Texas provides a registry of providers willing to accept patients possibly subject to unilateral decisions. TEX. HEALTH & SAFETY CODE ANN. § 166.053 (Vernon 2006). But the registry currently includes only one provider. Texas Department of State Health Services, Registry of Health Care Providers and Referral Groups, http://www.dshs.state.tx.us/THCIC/Registry.shtm (last updated Sept. 25, 2007).

344. *See* sources cited *supra* note 343. Once in a while, providers are able to transfer patients who request inappropriate treatment. *See, e.g.*, Alexander M. Capron, *Baby Ryan and Virtual Futility*, 25 HASTINGS CTR. REP., Mar.-Apr. 1995, at 20 (reporting that the parents of Ryan Nguyen found a facility willing to provide the requested treatment); Paris, *supra* note 185, at 1013 (reporting parents transferred Baby L's care to a consultant pediatric neurologist); Todd Ackerman, *Hospital to End Life Support: Houston Woman Faces Second Fight in 2 Months Over Husband's Care*, HOUSTON CHRON., Apr. 28, 2005, at B5 (noting report by St. Luke's Hospital in Houston that "more than 30 facilities had rejected Nikolouzos before Avalon Place surprised them and agreed to take [him]"); Beck, *supra* note 185, at A13 (stating the guardian ad litem for Baby L "found a pediatric neurologist from another hospital who was willing to do everything the mother wanted").

345. *See, e.g.*, Brief of Respondent at 3, 21, Duarte v. Chino Comty. Hosp., No. E020473 (Cal. Ct. App. Aug. 19, 1998) (arguing that while a transfer would have resolved a conflict where the provider refused to withdraw treatment at the surrogate's request, the patient "was never stable enough to transfer to the proposed facility").

346. UNIFORM HEALTH-CARE DECISIONS ACT § 7(g)(3).

the provider may decline to comply with the treatment request.[347] California, for example, rejected an ultimatum approach which requires the provider to transfer or comply.[348] Tennessee similarly clarifies that if a transfer cannot be made, then the provider shall not be compelled to comply.[349]

If the patient is transferred, then she will receive the requested treatment.[350] If the patient is not transferred, the inability to transfer should serve as confirming evidence that the requested treatment was outside the standard of care and that the provider's refusal to comply with the request was appropriate.[351] To the extent there is variability among providers' judgments of medical appropriateness, transfer thereby serves as an important safety valve function.[352]

---

347. *See id.* § 7(e)-(g).

348. CAL. PROB. CODE § 4736(b)-(c) (West Supp. 2007).

349. TENN. CODE ANN. § 68-11-1808(d) (2006).

350. Schwartz, *supra* note 105, at 162 (stating that "[i]f a patient who desires a particular course of treatment can find a healthcare provider—*any* healthcare provider—who believes that the proposed course of treatment is within the realm of reasonable medical alternatives, that patient will have access to that course of treatment").

351. *See* Anne L. Flamm & Martin L. Smith, Letter to the Editor, *Advance Directives, Due Process, and Medical Futility*, 140 ANNALS INTERNAL MED. 402, 404 (2004) ("The absence of a facility willing to accept transfer may indicate that a community consensus exists on the futility of particular medical interventions for a patient."). On the other hand, the inability to transfer may show nothing about the consensus over medical inappropriateness. First, many facilities do not make a diligent effort to locate potential transferee providers. Second, many providers refuse transfer for purely economic and risk management reasons.

352. AMA Council, *supra* note 53, at 940. The report describes the "fair process approach" as "insist[ing] on full and fair deference to the patient's wishes, placing limits on this patient-centered approach only when the harm to the patient is so unseemly that, even after reasonable attempts to find another institution, a willing provider of the service was not found." *Id.*; *see also* Halevy & McGuire, *supra* note 283, at 38 ("[T]he fact that the registry [of willing transfer providers] is so sparse supports the underlying ethical principle . . . of a professional consensus . . . ."); Lee, *supra* note 31, at 486 ("'Transfer of care' is used as a legal device to ensure the physician's professional rights are balanced against those of the patient-surrogates."); James J. Murphy, Comment, *Beyond Autonomy: Judicial Restraint and the Legal Limits Necessary to Uphold the Hippocratic Tradition and Preserve the Ethical Integrity of the Medical Profession*, 9 J. CONTEMP. HEALTH L. & POL'Y 451, 483-84 (1993) (stating that where no physician will agree to a transfer, this demonstrates consensus); Schwartz, *supra* note 105, at 163 ("When there is universal agreement among healthcare providers that the patient's request seeks something beyond the limits of medicine, that should constitute very strong evidence that the request is inappropriate."). This assumes that the patient's request for a particular course of treatment is based on medical reasons. Schwartz, *supra* note 105, at 162-63.

B. Other Comprehensive Unilateral Decision Statutes

While the UHCDA may be the most common unilateral decision statute, it is not the only one. Other states have adopted comprehensive unilateral decision statutes similar to those in the ten UHCDA jurisdictions.[353]

Like the UHCDA, these statutes are comprehensive in that they authorize providers to make unilateral decisions concerning any type of requested treatment, including situations where the surrogate has made an affirmative request for treatment.[354] Similar to the UHCDA, many of these statutes not only authorize unilateral decisions but also offer immunity for the providers who make those decisions.[355]

The key difference among the non-UHCDA comprehensive unilateral decision statutes concerns the definition of "medically inappropriate." Some statutes provide no definition or standard, leaving providers with maximum

---

353. *See generally* sources cited *infra* note 355 (citing state statutes containing unilateral decision provisions similar to UHCDA). Many of these states' laws were based on earlier NCCUSL uniform acts. *See generally* Thomas J. Marzen, *The "Uniform Rights of the Terminally Ill Act": A Critical Analysis,* 1 ISSUES L. & MED. 441, 474 (1986) (observing that the Act "gives the physician almost unfettered discretion to decide what will be done"); Leslie B. Oliver, *The Right to Die in North Dakota: The North Dakota Living Will Act,* 66 N.D. L. REV. 495, 525 (1990) ("Allowing physicians discretion to enforce the terms of a declaration may require them to become the ultimate authority as to whether life-prolonging treatment will be provided, withheld or withdrawn.").

354. *See, e.g.*, GA. CODE ANN. § 31-36-7(2) (2006) (stating that a provider may refuse to comply with a surrogate's request, but must aide in seeking transfer for the patient to another provider who will comply with the treatment request).

355. *See, e.g.*, ARK. CODE ANN. § 20-17-208(b) (2005) ("A physician or other health care provider, whose actions under this subchapter are in accord with reasonable medical standards, is not subject to criminal or civil liability or discipline for unprofessional conduct with respect to those actions."); GA. CODE ANN. § 31-32-8(b) (2006) ("No person shall be civilly liable for failing or refusing in good faith to effectuate the living will of the declarant patient."); GA. CODE ANN. § 31-36-8(2) (2006) ("No such provider or person shall be subject to any type of civil or criminal liability or discipline for unprofessional conduct . . . ."); GA. CODE ANN. § 31-36-8(3) (2006); IDAHO CODE ANN. § 39-4513(2) (Supp. 2007); 755 ILL. COMP. STAT. § 45/4-8(b), (c) (West 1993); IOWA CODE ANN. § 144A.9(2) (West 2002); KY. REV. STAT. § 311.633(3)-(4) (no penalties by anyone) (LexisNexis 2007); MD. CODE ANN., HEALTH–GEN. § 5-609(a) (LexisNexis 2005); MINN. STAT. ANN. § 145C.11 (West 1998) (establishing immunity to providers if the provider acts in good faith or acts according to a surrogate's decision); MONT. CODE ANN. § 50-9-204(2) (2005); NEB. REV. STAT. § 20-410(2) (1997); NEV. REV. STAT. § 449.630(2)-(3) (1991); NEV. REV. STAT. § 449.640(2) (1993); N.H. REV. STAT. ANN. § 137-J:8(II) (2005); N.C. GEN. STAT. § 90-322(d) (1993); N.D. CENT. CODE § 23-06.5-12(2) (2002); OHIO REV. CODE ANN. § 2133.11(A)(4) (LexisNexis 2006); OKLA. STAT. ANN. tit. 63, § 3101.10 (West 1995); TEX. HEALTH & SAFETY CODE ANN. §§ 166.044(a), 166.045(d), 166.166 (Vernon 1999); VT. STAT. ANN. tit. 18, § 9713(c)(3) (2006) (giving protection to hospital employees only from adverse employment decision); VA. CODE ANN. § 54.1-2988 (2005); WASH. REV. CODE §§ 70.122.051, .122.060(3) (2006); WIS. STAT. ANN. § 154.07(1)(a)(3) (West 1998); WIS. STAT. ANN. § 155.50(1)(b) (West 2003).

discretion to determine the circumstances under which they will refuse to comply with treatment requests.[356] Virginia, for example, provides that a physician is not required to "render medical treatment to a patient that the physician determines to be medically or ethically inappropriate."[357] Other states' statutes provide a more precise formulation, authorizing providers to decline to comply with treatment requests that would require treatment outside their professional medical judgment.[358] States articulate this standard in different ways, but all the formulations are analogous to the UHCDA's standard of "generally accepted health care standards."[359] The most common formulation of medical appropriateness is one based on "reasonable medical practice,"[360] "reasonable medical standards,"[361] "responsible medical practice,"[362] "medical judgment,"[363] or "usual and customary standards of medical practice."[364] Other statutes refer to "professional reasons,"[365]

---

356. *See, e.g.*, GA. CODE ANN. §§ 31-32-8(b), 31-36-7(2) (2006) (mentioning that physicians may refuse to comply with a living will, but not addressing when they may or may not refuse treatment); GA. CODE ANN. § 31-36-8(2), (3) (2006) (requiring that physician's refusal to comply with treatment request must be "substantially in accord with reasonable medical standards"); 755 ILL. COMP. STAT. §§ 35/3(d), 45/4-7(b) (West 1993); IND. CODE ANN. § 16-36-4-13(e) (LexisNexis 1993); IND. CODE ANN. § 30-5-7-4(b) (LexisNexis 2000); IOWA CODE ANN. § 144A.8(1) (West 2002); MINN. STAT. ANN. § 145B.06(1) (West 1991); MINN. STAT. ANN. § 145C.15(b) (West 1998); MO. ANN. STAT. § 459.030(1) (West 1985); MONT. CODE ANN. § 50-9-203 (2005); NEV. REV. STAT. §§ 449.628, 449.640 (1997); OHIO REV. CODE ANN. §§ 1337.16(B), 2133.02(D)(1) (LexisNexis 2006); OKLA. STAT. ANN. tit. 63, § 3101.9 (West 1998); TEX. HEALTH & SAFETY CODE ANN. § 166.046 (Vernon 1999); WIS. STAT. ANN. §§ 154.07(1)(a) (West 1990); WIS. STAT. ANN. § 155.50(1)(b) (West 2003).

357. VA. CODE ANN. § 54.1-2990(A) (2005).

358. Like the UHCDA, the statutes in most states allow providers to refuse to comply with surrogate treatment requests for *moral* reasons. *See, e.g.*, CAL. PROB. CODE § 4734 (West 2007); MD. CODE ANN., HEALTH–GEN. § 5-611(a) (West 2005); VA. CODE ANN. § 54.1-2990 (2005). While these provisions have rarely been used in the context of futility disputes, they are applicable and may soon be invoked more frequently. *See supra* note 319.

359. *See infra* notes 360-69.

360. *See, e.g.*, MINN. STAT. ANN. § 145B.13 (West 1991) ("reasonable medical practice").

361. *See, e.g.*, GA. CODE ANN. § 31-36-8(3) (2006) ("substantially in accord with reasonable medical standards at the time of reference").

362. *See, e.g.*, N.H. REV. STAT. ANN. §§ 137-J:7(I) (2006) (complying "within the bounds of responsible medical practice").

363. *See, e.g.*, LA. REV. STAT. ANN. § 40:1299.58.1(B)(3) (1985) ("It is the intent of the legislature that nothing in this Part shall be construed . . . to require the application of medically inappropriate treatment or life-sustaining procedures to any patient or to interfere with medical judgment with respect to the application of medical treatment or life-sustaining procedures.").

364. *See, e.g.*, CONN. GEN. STAT. ANN. § 19a-571(a) (West 2002) (shielding providers from liability where "the decision to withhold or remove such life support system is based on the best medical judgment of the attending physician in accordance with the usual and customary standards of medical practice"); MO. ANN. STAT. § 459.040 (West 1985).

365. *See, e.g.*, IDAHO CODE ANN. § 39-4513(2) (Supp. 2007) (allowing provider to withhold LSMT if unwilling to provide treatment for "professional reasons").

"professional grounds,"[366] or "professional standards"[367] as means for determining medical inappropriateness. Like the UHCDA, the other comprehensive unilateral decision statutes operate extra-judicially.[368] Additionally, these state statutes require the unwilling provider to attempt to transfer the patient before taking unilateral action.[369]

C. Narrow Unilateral Decision Statutes

While most states with unilateral decision statutes have adopted comprehensive provisions similar to the UHCDA, a few have taken a more "narrow" approach. New York, for example, enacted a narrow unilateral decision statute in 1987.[370] Other states soon enacted statutes similar to New York's, permitting unilateral decisions only in narrowly defined circumstances.[371]

As compared to the UHCDA and other comprehensive unilateral decision statutes, these narrow statutes offer a stricter range of circumstances under which providers can unilaterally stop LSMT.[372] In particular, the statutes are tightly delineated with respect to the following: (1) the type of treatment, (2) the presence of a surrogate request for treatment, and (3) the expected effect of the treatment.[373] First, certain types of medical interventions have been the focus of special attention. Consequently, some statutes limit types of treatment by authorizing unilateral decisions to withhold only CPR,[374] while others

---

366. *See, e.g.*, KY. REV. STAT. ANN. § 311.633(3) (LexisNexis 2007) (allowing providers to refuse treatment on "professional grounds").

367. *See, e.g.*, N.J. STAT. ANN. §§ 26:2H-62(d) (1992) ("Nothing in this act shall be construed to require . . . care in a manner contrary to law or accepted professional standards.").

368. *See, e.g.*, N.J. STAT. ANN. §§ 26:2H-66 (1992) (implementing a dispute resolution process to resolve disagreements between patients, patient's surrogates, and doctors); *cf.* sources cited *supra* note 355 (referencing statutes which give providers immunity from civil and criminal liability for treatment or refusal of treatment, implying that the judicial process may not lead to a satisfactory result for the patient or surrogate).

369. *See, e.g.*, 20 PA. CONS. STAT. ANN. § 5424(b) (West Supp. 2007); MD. CODE ANN., HEALTH-GEN. § 5-613(a)(1)(iii) (LexisNexis 2005); VA. CODE ANN. § 54.1-2987 (2005) ("An attending physician who refuses to comply . . . shall make a reasonable effort to transfer the patient . . . .").

370. N.Y. PUB. HEALTH LAW §§ 2961(12), 2966(1) (McKinney 1988).

371. *See, e.g.*, OR. REV. STAT. § 127.635(1) (2005); S.D. CODIFIED LAWS § 59-7-2.7(1) (2004); UTAH CODE ANN. § 75-2-1107(1) (1993); VT. STAT. ANN. tit. 18 § 9708(a) (Supp. 2006); W. VA. CODE ANN. § 16-30C-6(e) (LexisNexis 2006); *see also* VHA-NEC REPORT, *supra* note 109, at 6 ("VA physicians are not permitted to write a DNR order over the objection of the patient or surrogate, but they are permitted to withhold or discontinue CPR based on bedside clinical judgment at the time of cardiopulmonary arrest.").

372. *See* sources cited *supra* note 371.

373. *See* sources cited *supra* note 371.

374. *See, e.g.*, N.Y. PUB. HEALTH LAW § 2966(1) (McKinney 1988); VT. STAT. ANN. tit. 18, § 9708(a) (Supp. 2006); W. VA. CODE ANN. § 16-30C-6(e) (LexisNexis 2006).

authorize unilateral decisions to withhold only artificial nutrition and hydration.[375] Second, the narrow unilateral decision statutes limit not only the type of treatment, but they also limit unilateral decisions to situations where neither the patient nor the patient's surrogate has made a contrary decision.[376] The health care provider can only unilaterally stop LSMT when no other decision maker is available.[377] Third, the narrow unilateral decision statutes authorize a provider to make unilateral decisions only in narrow, verifiable circumstances of medical inappropriateness.[378] Rather than giving providers discretion to determine medical inappropriateness, these narrow statutes authorize unilateral decisions only in cases of brain death, physiological futility, or permanent unconsciousness.[379]

---

375. *See, e.g.*, OR. REV. STAT. §§ 127.580(1), 127.635(1) (2005); S.D. CODIFIED LAWS § 59-7-2.7 (2004).

376. *See, e.g.*, N.Y. PUB. HEALTH LAW § 2966(1) (McKinney 1988); OR. REV. STAT. § 127.580 (2005); UTAH CODE ANN. § 75-2-1107(1) (1993); VT. STAT. ANN. tit. 18, § 9708(a) (Supp. 2006); W. VA. CODE ANN. § 16-30C-6(e) (LexisNexis 2006). The New York Department of Health has apparently expanded the exception to permit physicians to write DNR orders even over the objection of a surrogate. *See* Edward F. McArdle, *New York's Do-Not-Resuscitate Law: Groundbreaking Protection of Patient Autonomy or a Physician's Right to Make Medical Futility Determinations?*, 6 DEPAUL J. HEALTH CARE L. 55, 73-74 (2002-2003).

377. *See* sources cited *supra* note 376.

378. *See, e.g.*, OR. REV. STAT. § 127.635(1) (2005). Oregon's unilateral decision statute requires one of four specified conditions:

> Life-sustaining procedures [such as artificial nutrition and hydration] . . . may be withheld or withdrawn . . . if the [patient] has been medically confirmed to be in one of the following conditions: (a) A terminal condition; (b) Permanently unconscious; (c) A condition in which administration of life-sustaining procedures would not benefit the principal's medical condition and would cause permanent and severe pain; or (d) The person has a progressive illness that will be fatal and is in an advanced stage, the person is consistently and permanently unable to communicate by any means, swallow food and water safely, care for the person's self and recognize the person's family and other people, and it is very unlikely that the person's condition will substantially improve.

Id. South Dakota similarly enumerates three circumstances:

> [A]rtificial nutrition or hydration may be withheld or withdrawn if: (1) Artificial nutrition or hydration is not needed for comfort care or the relief of pain and the attending physician reasonably believes that the principal's death will occur within approximately one week; or (2) Artificial nutrition or hydration cannot be physically assimilated by the principal; or (3) The burden of providing artificial nutrition or hydration outweighs its benefit, provided that the determination of burden refers to the provision of artificial nutrition or hydration itself and not to the quality of the continued life of the principal . . . .

S.D. CODIFIED LAWS § 59-7-2.7 (2004).

379. *See* sources cited *supra* note 378.

V. EFFECTS OF THE UNILATERAL DECISION STATUTES

In the early 1990s, health care providers were unwilling to make unilateral decisions to stop LSMT without legal protection.[380] Consequently, over the past eighteen years, state legislatures have promulgated statutes that purport to provide this protection.[381] Now, it is time to assess the effects of these statutes.

While little empirical data exists, there is sufficient evidence to detect four broad trends and identify focused issues for empirical research. The first two trends are reasonably negative, at least from the perspective of statutory effectiveness. First, even in states with comprehensive unilateral decision statutes, many hospitals still do not have futility policies.[382] Second, those few hospitals with futility policies rarely implement them to make a unilateral decision in cases of intractable conflict.[383]

Two additional trends have more positive attributes. First, the unilateral decision statute in one state, Texas, does appear to work.[384] Texas hospitals both have *and* implement futility policies.[385] Second, unilateral decision statutes appear to facilitate the informal resolution of futility disputes, reducing, although not eliminating, the need to resort to unilateral decision making.[386]

A. Hospitals Do Not Have Futility Policies.

Unfortunately, there is a "disturbing lack of information" on the prevalence of hospital futility policies.[387] The two most populated states in the country failed to implement any reporting mechanism as part of their unilateral decision statutes.[388] Consequently, as one distinguished health law scholar concluded, "No data exist on futility policies adopted by [institutions] in California [or Texas], much less across the nation."[389]

---

380. *See supra* notes 276-84.

381. *See supra* note 355.

382. *See infra* notes 387-97.

383. *See infra* notes 398-403.

384. *See infra* notes 404-13.

385. *See infra* notes 404-08.

386. *See infra* notes 409-13.

387. Kwiecinski, *supra* note 331, at 329.

388. In 2002, California, a UHCDA state, considered legislation that would "study the extent to which health care providers and institutions are denying patients life-sustaining health care that they desire." *Hearing on S.B. 1344 Before the S. Assembly Comm. on Appropriations* (2002). Unfortunately, that legislation was never enacted. *S.B. 1344 Status Rep.* (2002), http://www.leginfo.ca.gov/pub/01-02/bill/sen/sb_1301-1350/sb_1344_bill_20021130_status.html. Similarly, Texas failed to monitor the use of its unilateral decision statute. *See* Ramshaw, *supra* note 84. A bill introduced in March 2007 proposed to change this, H.B. 3474, 80th Leg. (Tex. 2007), but that bill died with the close of Texas's 80th legislative session. Texas Legislature Online, http://www.capitol.state.tx.us (search by bill number).

389. Schneiderman & Capron, *supra* note 243, at 529; *see also* Kwiecinski, *supra* note 331, at 329 ("At the writing of this essay, no reports surveying the circumstances in which

In fact, a distinguished group of scholars conducted an empirical research study in 1996 on this very issue.[390] They surveyed 1,990 large hospitals in the United States and received 537 responses.[391] Of these, only 29 (about 5%) were "clearly denominated as medical futility policies and . . . reached beyond DNR orders, more traditional life-sustaining treatment decisionmaking, and the determination of death."[392] Moreover, most of these 29 policies "envisioned a primarily consultative, consensus-building approach."[393] Almost none of the hospitals resolved what would happen if neither consensus nor transfer were possible in a case.[394] Additionally, there was no specification or authorization of a mechanism for the unilateral termination of LSMT.[395]

These statistics have not improved over the past decade. Recent evidence indicates that while unilateral decision statutes authorize health care providers to refuse compliance with inappropriate treatment requests, providers in these jurisdictions *reluctantly* continue to comply with such requests.[396] Although a number of health care institutions would like to have futility policies, only a few have adopted such policies.[397]

---

institutional futility policies have been invoked have been published.").

390. Sandra H. Johnson et al., *Legal and Institutional Policy Responses to Medical Futility*, 30 J. HEALTH L. 21, 27 (1997).

391. *Id.*

392. *Id.* Within the study, 137 hospitals responded that they had futility policies. Of those, 115 hospitals submitted their policy to the research team, who determined that most of the policies just pertained to traditional LSMT decision making with consent or determining brain death. *Id.*; Goldner, *supra* note 331, at 412.

393. Johnson, *supra* note 390, at 32.

394. *Id.* ("Because these transfer . . . provisions provided for permissible or optional courses of action, many do not resolve what will happen if transfer is not available, is burdensome, or is not desired by the patient/surrogate.").

395. *Id.* ("It was quite frequently the case that a policy . . . failed to specify an ultimate decisionmaker or decisionmaking body if conflict were to persist after all the processes were followed.").

396. *See* Bowman, *supra* note 89, at 1527 ("The reluctance of providers to act unilaterally comes in part . . . from a lack of medical agreement on a workable definition for futility and a lack of legal support for overriding patient choice.").

397. *See id.* at 1528 ("'A lot of people want to have policies, but a lot of people don't [have them].'" (quoting Shirley J. Paine)); Moldow, *supra* note 12, at 39 ("Fear of legal action has previously discouraged many institutions from adopting policies in the area of medical futility . . . ."); Nasraway, *supra* note 89, at 216 ("[I]t is much more common for hospital lawyers to argue in favor of doing the easy thing, i.e., to acquiesce to unreasonable demands . . . ."); Email from Ronald Cranford, Faculty Associate, Univ. of Minn.'s Center for Bioethics, to Thaddeus Pope, Assistant Professor of Law, Univ. of Memphis Cecil C. Humphreys School of Law (July 11, 2004, 07:41 PM) ("Many hospital lawyers, much more concerned about legal liability and adverse publicity for their institutions, have been extremely tentative, if not outright hostile, to ethics committees formulating and implementing futility policies, even though many of us in the field of clinical ethics feel these guidelines are badly needed."); *cf.* Anderson-Shaw, *supra* note 159, at 299 ("Absent state or federal statutes specifically guiding futile care activity, many institutions work under a much more informal approach to futile care.").

B. Hospitals Do Not Enforce Their Futility Policies

While a hospital without a futility policy is unlikely to make a unilateral decision to stop LSMT, the existence of a futility policy hardly means it will be fully utilized. It appears that many institutions that have futility policies either are not implementing them at all or are implementing them only in a very narrow and infrequent manner.[398]

For example, a health care provider in an institution with a futility policy may invoke that policy in an attempt to resolve a dispute.[399] However, if the dispute is intractable, the provider may be reluctant to invoke the unilateral decision provisions of the policy.[400] Instead, the provider will ultimately accede to the surrogate's treatment request.[401] In sum, while futility policies facilitate the informal resolution of disputes, providers defer when the dispute proves intractable.

The unilateral decision statutes in most states seem to have had limited effect. Commentators noted that before the passage of state statutes authorizing unilateral action, hospitals typically deferred to family wishes because they feared being sued.[402] Now, even *with* such laws, hospitals still accede to family wishes for fear of being sued.[403] The statutes have failed to change the behavior of providers.

C. Hospitals in Texas Enforce Their Futility Policies

There is an exception to this general failure in unilateral decision statutes: Texas's statute appears to have had a significant impact since its adoption in

---

398. *See, e.g.*, Bowman, *supra* note 89, at 1527 ("While physicians sometimes disagree with patients or their surrogates over end-of-life care, however, they rarely end care in violation of patient wishes. . . . 'If you're still at an impasse, the hospital continues to provide maximum support.'"); Burns, *supra* note 217, at 3 ("[D]espite an increasing number of ethics consults on questions of futility we do not invoke our own futility policy."); Mary Pat Flaherty, *Right to Die Decision Has Little Impact Here*, PITT. POST-GAZETTE, June 27, 1990, at A1 (reviewing policies at Pittsburgh-area hospitals and observing that "[c]are usually continues—full bore—when an incapacitated patient's family or his designated decision-maker cannot agree with recommendations made by doctors that further care would be futile"); Fletcher, *supra* note 10, at S:230 (noting a "moratorium" on the use of UVA's policy after the *Baby K* decision); Wlazelek, *supra* note 46 (reporting reluctance at Lehigh Valley Hospital-Muuhlenberg in Bethlehem to utilize its unilateral decision policy). Reporter Ann Wlazelek remarked that the option to refuse treatment "takes courage on the part of the physician because he or she will most likely be sued. No doctor at LVH has refused to treat a patient but some patients have been transferred to other facilities." *Id.*

399. *See supra* Part I.D.

400. *See supra* note 398.

401. *See supra* note 398.

402. *See supra* notes 276-81.

403. *See infra* notes 415-18.

1999.[404] In one study at Baylor University Medical Center in Dallas, researchers found that the statutory authorization gave physicians "more comfort," thereby increasing ethical consultations regarding futility disputes by 67%.[405] Not only did physicians and hospitals across Texas begin the dispute resolution process but also, in approximately two percent of cases that were proven intractable, the providers gave notice that they were going to unilaterally stop LSMT.[406]

A broader study of sixteen Texas hospitals over a five-year period found that, on average, each hospital made the decision to unilaterally stop treatment at least one time each year.[407] Indeed, Texas hospitals unilaterally stopped or decided to stop LSMT, even in the face of significant controversy and mass media coverage urging otherwise.[408] In short, the Texas statute has truly changed provider conduct.

D. Unilateral Decision Statutes Facilitate the Informal Resolution of Futility Disputes

Even in cases where the unilateral decision statutes do not facilitate unilateral decisions, the statute may still help the informal resolution of futility disputes because most disputes are not intractable.[409] These statutes help ensure that *earlier* steps in the dispute resolution process work better.[410] They facilitate informal resolution by setting "temporal and conceptual boundaries."[411] For example, surrogates might say, "If you are asking us to agree with the recommendation to remove life support from our loved one, we cannot. However, . . . if the law says it is OK to stop life support, then that is what should happen."[412] The existence of a hospital policy and state law helps

---

404. Fine & Mayo, *supra* note 84, at 744.

405. *Id.* at 744-45.

406. Ramshaw, *supra* note 84.

407. *Id.*

408. *See, e.g.*, Todd Ackerman, *Transfer Resolves Latest Futile-Care Case: Nursing Home in Lubbock to Take Memorial Hermann Patient*, HOUSTON CHRON., July 31, 2006, at B1; Ackerman, *supra* note 43, at B1; Todd Ackerman, *Relocation of Heart Patient on Life Support Called Off: The Controversy is Not Put to Rest as Midwest Facility Says Her Condition is Too Complicated*, HOUSTON CHRON., Apr. 29, 2006, at B1 [hereinafter Ackerman, *Relocation of Heart Patient*]; Ackerman, *supra* note 344, at B5; Belluck, *supra* note 38, at A1; Robert H. Frank, *Weighing the True Costs and Benefits in a Matter of Life and Death*, N.Y. TIMES, Jan. 19, 2006, at C3; Murphy, *supra* note 45, at A29; Emily Ramshaw, *Judge Gives Family Time to Move Woman*, DALLAS MORNING NEWS, Feb. 16, 2007, B2; Ramshaw, *supra* note 45, at B1; Mary Ann Roser, *Where Doctors See Futility, Family Sees Hope*, AUSTIN AM.-STATESMAN, Apr. 28, 2006, at A1.

409. *See supra* notes 91, 331-36 and accompanying text.

410. *See* Fine & Mayo, *supra* note 84, at 744 (noting statute provides for consultations to address disputes between providers and patients or surrogates concerning treatment options).

411. Fine, *supra* note 92, at 70-71.

412. Fine & Mayo, *supra* note 84, at 745 (internal quotations omitted).

families accept the fact that death cannot be postponed forever and that, eventually, LSMT is inappropriate.[413]

VI. CAUSES OF UNILATERAL DECISION STATUTE DISUSE

Providers in unilateral decision statute jurisdictions, other than Texas, generally do not make unilateral decisions to stop inappropriate treatment. Why is this? Why do providers continue to accede to surrogate requests for treatment that they consider medically inappropriate? Why do the unilateral decision statutes remain unused?[414] Many have suggested that the primary reason unilateral decision statutes are not working is because of legal uncertainty and the fear of litigation.[415] Surely, other factors, such as the fear of

---

413. *See, e.g.*, Fine, *supra* note 48, at 80 ("[F]amilies come to understand that there is a finite limit . . . [and] that they are not in total control of the situation."); Fine, *supra* note 92, at 70–71; Fine, *supra* note 100, at 1221 ("[T]he family was relieved because they had 'put up the good fight' . . . but now the decision was out of their hands."); Fine & Mayo, *supra* note 84, at 746 ("[T]he greatest significance of the law is how it changes the nature of conversations . . . about futile-treatment situations by providing conceptual and temporal boundaries."). *But see* Burns, *supra* note 217, at 3 (suggesting that a formal futility policy leads to "confrontation" and "polarization").

414. Unfortunately, non-anecdotal, statistical evidence of the prevalence and use of hospital futility policies is unavailable. It is imperative that academics and policymakers engage in more empirical research to uncover the reasons why providers accede to inappropriate requests. This research and analysis will aid in identifying the problems with sufficient precision and in developing appropriately tailored solutions.

415. *See, e.g.*, WHEN CHILDREN DIE, *supra* note 284, at 322 ("[I]t is increasingly clear that before a physician may terminate life support on any patient [where the family objects] . . . she or he should assume that it is necessary to ask a court for an order."); Brett, *supra* note 70, at 283–84 ("[T]he threat of litigation is an important reason, perhaps the major reason, that physicians are reluctant to withhold or withdraw 'futile' life-sustaining treatment unilaterally against the wishes of family members."); Fletcher, *supra* note 10, at S:230 (noting that health care organizations must "wait for clarification of the law . . . on which futile treatments can be withheld or withdrawn" and in the meantime must "treat until the dispute is resolved"); Hall, *supra* note 246, at 119 ("[T]o the extent that a crisis is in fact widely perceived, it has the quality of a self-fulfilling prophecy . . . ."); Kapp, *supra* note 122, at 242 (recommending that in the absence of "unambiguous legal guidance," providers should accede to surrogate requests); Marshall B. Kapp, *Legal Anxieties and End-of-Life Care in Nursing Homes*, 19 ISSUES L. & MED. 111, 119 (2003) (discussing how a "broad fear of regulatory sanctions for providing too little aggressive LSMT" and the likelihood of civil malpractice actions means that "demand for aggressive LSMT virtually always controls the situation regardless of how inappropriate that demand may be"); Lantos, *supra* note 12, at 587 (explaining that many doctors are unwilling to "take the risk that punishment, rather than forgiveness, may come their way"); Valerie A. Palda et al., *"Futile" Care: Do We Provide It? Why? A Semistructured Canada-Wide Survey of Intensive Care Unit Doctors and Nurses*, 20 J. CRITICAL CARE 207 (2005) (finding that 75% of physicians provided futile care because of legal pressures). Notably, whatever the actual risks, they may be overestimated by providers. *See* McArdle, *supra* note 376, at 71 ("Numerous articles have warned physicians of the serious legal risk in unilaterally writing a DNR order

adverse publicity, also intimidate providers from making unilateral decisions.[416] However, legal factors appear to be the most material cause and, therefore, will be the focus of this Article.

In 1999, when the American Medical Association encouraged hospitals to adopt futility guidelines, it noted that "the legal ramifications of this course of action are uncertain."[417] Now, even with statutory authorization, there is *still* significant legal uncertainty.[418]

There are three potential sources of this uncertainty. First, the unilateral decision statutes are vague, leaving providers and hospital counsel unsure of what standards are required to obtain safe harbor status.[419] Second, there is uncertainty concerning whether and when these state statutes are preempted by conflicting federal law.[420] Third, there is uncertainty concerning the constitutionality of the statutes.[421]

It is impossible to conclude that these sources of uncertainty affect a providers' willingness to use unilateral decision statutes. To definitively answer this question, empirical research must be employed to assesses the motivation for provider behavior. But although no such evidence currently exists, one state presents a case study: Texas. The Texas statute effectively facilitates unilateral decisions, yet it is equally subject to federal preemption and constitutional requirements.[422] Therefore, it seems that the only *material* uncertainty must concern that of the non-Texas unilateral decision statutes themselves.

. . . .").

416. *See, e.g.*, WRONG MEDICINE, *supra* note 37, at 134 ("[I]f the decision to withdraw life-sustaining treatment became known to any of the patient's friends or to the public, the hospital might have to face embarrassing publicity (or, as they put it, 'bad headlines')."); Fletcher, *supra* note 10, at S:226 (observing hospitals can "'engender ill will in their communities'" (quoting Alan Meisel)); Rivin, *supra* note 30, at 391 ("The 'pay or leave' demand is probably too coercive for the hospital or the physician's malpractice carrier or public relations advisor to accept."); Schneiderman & Capron, *supra* note 243, at 525-26 ("[I]n a survey of representatives of all 43 children's hospitals in the country . . . almost all acknowledged [that] their own hospital would probably yield to demands for life-sustaining treatment . . . because of fears of lawsuits and bad headlines."); Ackerman, *Relocation of Heart Patient*, *supra* note 408, at B1 ("St. Luke's was flooded with angry calls about the plan to pull the plug on Clark . . . ."); Andrea Clarke's Struggle for Life, Posting to ProLifeBlogs.com, http://www.prolifeblogs.com/articles/archives/2006/04/andre_clarkes_s.php (Apr. 25, 2006, 01:04 AM) (describing unilateral decision making as a "flagrant act of (passive) euthanasia").

417. AMA Council, *supra* note 53, at 940.

418. *See infra* Part VI.A-C.

419. *See infra* notes 423-68.

420. *See infra* notes 440-51.

421. *See infra* notes 452-64.

422. *See infra* Part VII.B.

A. Uncertainty from Statutory Vagueness

Lawyers, bioethicists, health care providers, and policymakers have had enormous difficulty defining "medically inappropriate."[423] Years of debate have failed to produce any consensus.[424] As a result, policymakers designed an approach with vague standards, thereby giving substantial discretion to the health care providers and institutions.[425] Rather than establishing a clear framework for determining medical inappropriateness, the statutes leave that determination to the judgment and discretion of the individual health care

---

423. *See generally* Anderson-Shaw, *supra* note 159, at 303 (noting that all state statutes use similar terms like "medically inappropriate" or "medically ineffective" to define futility, yet the definitions of these terms are left to the discretion of the providers); Tomlinson & Czlonka, *supra* note 59, at 33 (arguing "against any attempt to base a futility policy on some concrete definition of *futility*"); David G. Warren, *The Legislative Role in Defining Medical Futility*, 56 N.C. MED. J. 453, 454 (1995) ("[T]here may be another wave of proposals in state legislatures to address the question of . . . medical futility. Drafting difficulties are obvious . . . .").

424. *See* Moseley, *supra* note 4, at 211 ("[D]espite years of debate in scholarly journals, professional meetings, and popular media, consensus on a precise definition eludes us still."); *see also* Burt, *supra* note 66, at 249-50 ("[W]ithin the medical community no consensus has emerged to give practical content to the futility concept . . . ."); Judith F. Daar, *A Clash at the Bedside: Patient Autonomy v. a Physician's Professional Conscience*, 44 HASTINGS L.J. 1241, 1246 (1993) (viewing this struggle as a "clash at the bedside"); Goldner, *supra* note 331, at 416 (empirical research study "suggests an absence of consensus"); Lee, *supra* note 31, at 482; Mark Strasser, *The Futility of Futility? On Life, Death, and Reasoned Public Policy*, 57 MD. L. REV. 505, 514 (1998) (describing current formulations of the term as either under-inclusive, over-inclusive, or both); Richard L. Wiener et al., *A Preliminary Analysis of Medical Futility Decisionmaking: Law and Professional Attitudes*, 16 BEHAV. SCI. L. 497, 499 (1998); Zientek, *supra* note 82, at 251 ("Because of the difficulty in defining futility . . . the [Texas] statute is vague on a number of central issues."). *But see* Levine, *supra* note 10, at 73 (suggesting that there is a general consensus among health care providers that some types of treatment are medically inappropriate).

425. *See, e.g.*, THE RIGHT TO DIE, *supra* note 17, § 13.02 at 13-6 to 13-7; AMA Council, *supra* note 53, at 939 (rejecting an absolute definition in favor of a process-based approach); Ferguson, *supra* note 16, at 1220 ("[T]he statute provides no clear standard regarding the propriety of such decisions."); *id.* (arguing that the UHCDA does not "provide a clear definition of futility and fails to supply adequate ethical context or constraints to guide difficult decisions"); Keith Shiner, *Medical Futility: A Futile Concept?* 53 WASH. & LEE L. REV. 803, 810 (1996) (stating that the legislative bodies failed to deal with the problem of medical futility, instead creating undefined statutes); *cf.* Johnson, *supra* note 390, at 36 ("Developing clarity in the boundaries of futility is fundamental.").

provider.[426] In this sense, the statutes can be described as "purely enabling legislation."[427]

It is not unusual for policymakers to delegate responsibility when they cannot agree on rules or guidelines.[428] Moreover, this deference is typical with respect to the medical profession.[429] The discretion afforded by the unilateral decision statutes, however, is purchased at the expense of significant uncertainty.[430] Because of the statutory vagueness, providers have difficulty ensuring that they are satisfying the required standards.[431]

---

426. The legislature's failure to create a clear framework to determine medial inappropriateness is hardly surprising. The inappropriate treatment question "address[es] issues concerning the meaning that we attach to life, particularly diminished life; self-determination; the nature of the physician-patient relationship; and the just allocation of scarce health care resources." Shiner, *supra* note 425, at 808-09.

427. MASON & LAURIE, *supra* note 77, at 596; *see* Ferguson, *supra* note 16, at 1220 ("The UHCDA provides a mere framework . . . [and] gives only broad platitudes . . . ."); *id.* at 1221 ("These sections seemingly create an open-ended excuse for a physician to withdraw treatment. . . ."); *see also* DWORKIN, *supra* note 124, at 144-45 (arguing that given factual variability of the issues and the lack of public consensus, matters should be left to the medical profession with minimal legal oversight); Elizabeth Day, *Do Not Resuscitate—and Don't Bother Consulting the Family,* SUNDAY TELEGRAPH [UK], Mar. 14, 2004, at 22, *available at* 2004 WL 4176646 ("There is the possibility of legislation, but in a field as controversy-strewn as medical ethics, a blanket law remains an imperfectly blunt tool.").

428. *Cf.* Carl E. Schneider, *Discretion, Rules, and Law: Child Custody and the UMDA's Best-Interest Standard,* 89 MICH. L. REV. 2215, 2244-45 (1991) (discussing the "'rule-building discretion'" which arises from a "direct and deliberate grant of discretionary authority").

429. *Cf.* Carl E. Schneider, *Void for Vagueness,* 37 HASTINGS CTR. REP., Jan.-Feb. 2007, at 10 ("In short, lawmakers have essentially established rules intended to hold medicine to its own standards and then mostly left the system to work unmolested.").

430. *See, e.g.*, HALL ET AL., *supra* note 78, at 451 ("On balance, it is difficult to offer much assurance about the existing legal climate regarding futility policies."); Ferguson, *supra* note 16, at 1243 (noting that the statute fails to provide a "usable, clear standard that protects the physician"); Flamm, *supra* note 10, at 4 ("The promise of immunity, of course, is not guaranteed; patients can challenge a provider's adherence to [the statute] or more generally dispute the reasonableness of actions taken."); Kwiecinski, *supra* note 331, at 349-50 ("When treatment can be or should be described as 'inappropriate' is not defined by the statute. . . . This lack of boundaries and oversight allows the providers far too much discretion."); Meisel & Jennings, *supra* note 11, at 75 ("[T]he law is unclear on what should be done."); Rowland, *supra* note 214, at 297 ("[T]hese statutes provide little guidance in regards to the limiting of the obligation for physicians to provide ongoing care they believe futile."); Schneiderman & Capron, *supra* note 243, at 528 ("For if limits to physicians' obligations are not defined, end-of-life outcomes are likely to be determined less by medical circumstances and justifiable standards and more by individual healthcare providers' tolerance for risk, patients' and families' varying degrees of knowledge and rhetorical skills, and economic considerations."); Tovino & Winslade, *supra* note 41, at 29 (observing that in futility cases "no widely accepted ethical and legal framework exists to govern decision-making"); *cf. In re* Bowman, 617 P.2d 731, 738 (Wash. 1980) (noting, with respect to brain death, that "[a]doption of [a legislative] standard will alleviate concern among medical practitioners that legal liability might be imposed when life support systems are withdrawn . . . ."). *But cf.* Goldner, *supra* note 331, at 409 ("[C]ourts

Indeed, the drafters of the UHCDA recognized this very shortcoming, observing that the statute really "provides no immunity at all . . . [because] virtually every question of reasonable care is a jury question."[432] The lack of immunity was "one of the reasons why [providers] want[ed] to get something in the black letter that talks about acceptable health-care standards."[433]

Some have suggested that the unilateral decisions statutes could have been effective, despite their vagueness, if "the medical profession . . . articulate[d] and thereafter follow[ed] uniform practice standards regarding futile care."[434] For example, recognizing the dynamic advancement in technology, the drafters of the Uniform Determination of Death Act (UDDA) did not specify any exact diagnoses in the statute itself.[435] Providers did, however, develop clinical criteria necessary to implement the UDDA.[436] In contrast, with respect to medical inappropriateness under the UHCDA, providers have neither articulated nor adhered to any clear universal standards of practice.[437]

---

are hesitant to penalize physicians who reasonably rely on what they perceive to be professional standards . . . .").

431. *Cf.* Blumstein, *supra* note 172, at 1049 (noting that flexibility is not a desirable objective for a safe harbor); *Final Rule: Medicare and State Health Care Programs: Fraud and Abuse; Safe Harbor for Federally Qualified Health Centers Arrangements Under the Anti-Kickback Statute*, 72 Fed. Reg. 56,632, 56,639 (Oct. 4, 2007). A trade association commented that requiring health care centers to implement and document "reasonable, consistent, and uniform standards" provides "insufficient guidance" as well as "a chilling effect on parties' participation in safe harbored arrangements, as parties would be unsure whether their standards would satisfy the requirements of the safe harbor." *Id.* On the other hand, at least one statute defines the provider's discretion subjectively rather than objectively. *See, e.g.*, N.M. STAT. § 24-7A-7(F) (2006) ("'Medically ineffective health care' means treatment that would not offer the patient any significant benefit, *as determined by a physician*.") (emphasis added).

432. Nat'l Conference of Comm'rs on Uniform State Laws, Proceedings in Comm. of the Whole, Uniform Health-Care Decisions Act, July 30, 1993, at 141-42 (statement of Comm'r Windsor Dean Calkins). Louisiana, for example, had a unilateral decision statute in 1998 exempting providers from care that was "medically inappropriate" and "contrary to medical judgment." Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1076 (La. Ct. App. 1998). Because these terms were not defined, however, an appellate court remanded a malpractice case for further litigation to determine the standard of care. *Id.*

433. Nat'l Conference of Comm'rs on Uniform State Laws, Proceedings in Comm. of the Whole, Uniform Health-Care Decisions Act, July 30, 1993, at 144 (statement of Comm'r Michael Franck).

434. Isackson, *supra* note 290, at 11; *see also* Kapp, *supra* note 122, at 172 (noting the need for "broad consensus within the medical community" and "societal agreement").

435. *See* Bernat, *supra* note 119, at 39 (stating that the "distinction between the brain's clinical functions and brain activities, recordable electronically or through other laboratory means," was not found within the UDDA).

436. *See* Bernat, *supra* note 119, at 40.

437. *See supra* notes 214-17. There are a few narrow exceptions. For example, providing only comfort care for anencephalic infants is a well-settled standard of care. Not even the opposing experts in *Baby K* contracted this. Brief of Appellants at 15-16, *In re* Baby K, No. 93-1899 (4th Cir. 1993), 1993 WL 13123742.

Consequently, the practice of deferring to surrogate demands has become the standard of care.[438]

B. Uncertainty from Fear of Preemption[439]

Even if providers could be reasonably certain of compliance with state unilateral decision statutes, this clarity would provide no legal comfort to providers if unilaterally stopping LSMT violated federal law. Preemption outside the futility context remains an obstacle to state efforts to develop more rational allocation systems.[440] Preemption may similarly stand as an obstacle to the effectuation of state unilateral decision laws.

Notably, the Fourth Circuit has held that Virginia's unilateral decision statute was preempted by the Emergency Medical Treatment and Active Labor Act (EMTALA).[441] Some commentators have since suggested that the preemptive scope of EMTALA is "limited" and that the duty imposed by EMTALA "cannot be invoked to require treatment in the vast majority of futility cases."[442] After all, EMTALA does not apply to inpatients.[443] Once the

---

438. *Cf.* Peter Albertson, *A 72-Year-Old Man With Localized Prostate Cancer*, 274 JAMA 69, 73 (1995) ("[T]here's an interesting catch-22—the medicolegal standard of care becomes what physicians do. If... physicians all [provide inappropriate treatment]... for fear of being sued if they don't, then eventually if enough of them do it, they'll create the truth of their fear."); Clark C. Havighurst, *Practice Guidelines as Legal Standards Governing Physician Liability*, 54 L. & CONTEMP. PROBS. 87, 97-98 (1991) ("Customary medical practices have evolved in the United States under systems of paying for medical care that create economic incentives for both physicians and patients to overutilize services, spending more on marginal benefits than they are in any sense worth.")

439. While this Article does not fully develop the preemption analysis under each of these statutes, Part VI.B. examines the scope of potential preemption.

440. *Cf.* Mary A. Crossley, *Medical Futility and Disability Discrimination*, 81 IOWA L. REV. 179, 181 (1995-1996) (discussing the preemption effect of the Americans with Disabilities Act on providers' attempts to ration health care).

441. *In re* Baby K, 16 F.3d 590, 597 (4th Cir. 1994).

442. THE RIGHT TO DIE, *supra* note 17, § 13.06[C] at 13-30.

443. *See* Bryan v. Rectors & Visitors of the Univ. of Va., 95 F.3d 349, 353 (4th Cir. 1996). The court acknowledged the "legal reality" that "[o]nce EMTALA has met that purpose of ensuring that a hospital undertakes stabilizing treatment for a patient, who arrives with an emergency condition, the patient's care becomes the legal responsibility of the hospital and the treating physicians." *Id.* In the court's analysis, "the legal adequacy of that care is then governed not by EMTALA but by the state malpractice law that everyone agrees EMTALA was not intended to preempt." *Id.* The court also distinguishing *Baby K* in part because that case did not focus on the temporal duration of obligation. *Id.*; *see also In re* AMB, 640 N.W.2d 262, 289 (Mich. Ct. App. 2001) (holding that there was no EMTALA violation where patient had been admitted to hospital for more than a week before withdrawal of LSMT); Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1075 n.2 (La. Ct. App. 1998) ("Agreeing with *Bryan*, we find that EMTALA provisions are not applicable to the present case.").

hospital has screened and stabilized a patient, it no longer has any obligation under EMTALA to provide medical services.[444]

While EMTALA's preemptive scope is circumscribed, the restrictions that it continues to impose on the scope and applicability of state unilateral decision statutes remain noteworthy. In particular, while the requisite treating period under EMTALA is limited, it is significant under the circumstances in which medically inappropriate care is often requested.[445] Specifically, EMTALA does not apply to inpatients; however, the subjects of many futility disputes were not inpatients. For example, in *Baby K,* by the time Fairfax Hospital sought declaratory relief, Baby K had already been transferred to a nursing home.[446] Yet over the next four months, she returned to the hospital three times due to breathing difficulties.[447] This status is even more common among adult patients, who are transferred from nursing homes to hospitals upon the occurrence of an acute event.[448] Furthermore, in similar circumstances, unilateral decisions to stop LSMT may be preempted by competing obligations under other federal statutes, including the following: (1) the Americans with Disabilities Act,[449] (2) section 504 of the Rehabilitation Act of 1973,[450] and (3) the Child Abuse Prevention and Treatment Act.[451]

C. Uncertainty from Fear of Unconstitutionality

Generally, there is little guidance regarding which state futility statutes violate the U.S. Constitution, because futility disputes are rarely litigated and

---

444. 42 C.F.R. § 489.24(d)(2) (2007).

445. *See* Fletcher, *supra* note 10, at S:230 (expressing concerned about patient who spent "seventeen days in intensive care").

446. *Baby K*, 16 F.3d at 593; *In re* Baby K, 832 F. Supp. 1022, 1024-27 (E.D. Va. 1993).

447. *Baby K*, 16 F.3d at 593; *Baby K*, 832 F. Supp. at 1024-25. It is unclear whether Baby K was discharged from the nursing home and presented at the emergency department of the hospital or was transferred from the nursing home to the hospital. Regardless, EMTALA would be triggered in either case. *Baby K*, 16 F.3d at 594-95 n.6 (citing 42 U.S.C. § 1395dd(g)).

448. *See, e.g.*, Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1073 (La. Ct. App. 1998) ("Having suffered cardiorespiratory arrest, Sonya Causey was transferred to St. Francis Medical Center (SFMC) from a nursing home."); *Barriers*, *supra* note 22, at 16 (reporting "an increasing number of terminally ill nursing home patients coming to the emergency department . . . when they experience life threatening symptoms"). Spiro Nikolouzos was transferred from St. Luke's Hospital to Avalon Place, a nursing home, but then back to Southeast Baptist Hospital after he developed pneumonia; Southeast Baptist sought to unilaterally terminate care. Ackerman, *supra* note 344, at B5.

449. *See Baby K*, 832 F. Supp. at 1027-28; Bopp & Coleson, *supra* note 45, at 842-44; Crossley, *supra* note 440, at 202-05.

450. *See Baby K*, 832 F. Supp. at 1026-27; Bopp & Coleson, *supra* note 45, at 842.

451. *See* Montalvo v. Borkovec, 647 N.W.2d 413, 419 (Wis. Ct. App. 2002); Sadath A. Sayeed, *Baby Doe Redux? The Department of Health and Human Services and the Born-Alive Infants Protection Act of 2002: A Cautionary Note on Normative Neonatal Practice*, 116 PEDIATRICS e576, e580-81 (2005).

courts tend to avoid deciding constitutional questions. Nevertheless, limited accounts of judicial treatment and academic legal commentary suggest that there is a reasonable risk of unconstitutionality for some unilateral decision statutes.

Where surrogate insistence on treatment is based on "religious convictions," the unilateral termination of LSMT may implicate the patient's First Amendment rights.[452] Where the patient is a prisoner, unilateral termination of LSMT could implicate the Eighth Amendment prohibition against cruel and unusual punishment.[453] Further, some have argued that unilateral termination is inconsistent with equal protection,[454] the right to life,[455] and the freedom of expression.[456]

Some litigants and commentators have even argued that the unilateral termination of LSMT would effectively constitute a usurpation of the patient's fundamental right to refuse LSMT.[457] However, to the extent that *Cruzan* established such a constitutional right, it is probably only a negative right to be free from unwanted treatment, not an affirmative right to LSMT.[458] Nevertheless, more than one court has held that the Fourteenth Amendment Due Process Clause prohibits unilaterally stopping LSMT.[459]

In any case, there is state action and a constitutionally protected interest in life is at stake.[460] Therefore, the procedures attendant to the deprivation of this

---

452. *Baby K*, 832 F. Supp. at 1029; *see also* Darren P. Mareiniss, *A Comparison of* Cruzan *and* Schiavo*: The Burden of Proof, Due Process, and Autonomy in the Persistently Vegetative Patient*, 26 J. LEGAL MED. 233, 252-53 (2005) (presenting the argument that the very concept of futility might offend one's belief in faith healing and the absolute sanctity of life).

453. George P. Smith, II, *Futility and the Principle of Medical Futility: Safeguarding Autonomy and the Prohibition Against Cruel and Unusual Punishment*, 12 J. CONTEMP. HEALTH L. & POL'Y 1 (1995).

454. Bopp & Coleson, *supra* note 45, at 837-39.

455. *Id.* at 839-40.

456. *Id.* at 841-42.

457. *See, e.g.*, Rideout v. Hershey Med. Ctr., 30 Pa. D. & C.4th 57, 62 (Dauphin County Ct. C.P. Dec. 29, 1995) (No. 872S1995), 1995 WL 924561 (reasoning that hospital had violated the constitution and usurped the patient's interest in her own life when the hospital unilaterally disconnected her life support).

458. *See* Johnson v. Thompson, 971 F.2d 1487, 1495 (10th Cir. 1992) (citing DeShaney v. Winnebago County Dep't Soc. Servs., 489 U.S. 189, 195 (1989)); Abigail Alliance for Better Access to Developmental Drugs v. von Eschenbach, 495 F.3d 695 (D.C. Cir. 2007), *cert. denied* No. 07-444, 2008 WL 114305 (Jan. 14, 2008); Mareiniss, *supra* note 452, at 251, 258.

459. *See In re* Baby K, 832 F. Supp. 1021, 1030 (E.D. Va. 1993) ("A parent has a constitutionally protected right to 'bring up children' grounded in the Fourteenth Amendment's due process clause. . . . [and] '[w]hen parents do not agree on the issue of termination of life support . . . this Court must yield to the presumption in favor of life.'"); *Rideout*, 30 Pa. D. & C.4th at 83-84 (allowing parents to assert right to life claim on behalf of child because "their privacy-based rights were violated under both state and federal constitutions").

460. Thaddeus Mason Pope, *Hospital Ethics Committees as a Forum of Last Resort under The Texas Advance Directives Act: A Violation of Procedural Due Process* (unpublished

interest must provide sufficient protection from error or abuse.[461] At a minimum, the surrogate must be afforded proper notice, an opportunity for a meaningful hearing, and access to an impartial tribunal.[462] Otherwise, unilateral termination could violate procedural due process.[463]

In sum, unilateral decision making may be constrained by constitutional and federal statutory constraints. Determining the parameters of those constraints merits further legal analysis. Yet, regardless of the nature of those constraints, they are not deterring Texas providers from making unilateral decisions to stop LSMT.[464] Therefore, it seems that somewhere deep in the heart of Texas lies the answer to making other state unilateral decision statutes more effective.

VII. SOLUTIONS: MAKING THE SAFE HARBOR NAVIGABLE

Because the unilateral decision statutes are too vague and open-ended, their purported safe harbors are not navigable. Can we make them more navigable? Can we reduce the uncertainty? There are two alternatives: (1) make the statutory standards concrete and precise or (2) abandon substantive standards altogether and use a purely process-based approach, like that used in Texas.[465]

A. Eliminating Uncertainty with Precise Standards

Consensus on precise, substantive, and legislatable measures of medical inappropriateness has proven unachievable.[466] Perhaps this should not be too surprising. Very few areas of medicine have professional standards that are "sufficiently mandatory and concrete" to operate as a safe harbor.[467] Rarely do providers have what is necessary for immunity: "a precise and plain statement of the acceptable medical practice."[468] Instead, professional standards are typically set "*ex post* by selectively drawn expert witness testimony."[469]

---

manuscript) (on file with author).

461. Kwiecinski, *supra* note 331, at 347.

462. *Id.*

463. *See id.* at 345-47.

464. *See supra* note 404-08 and accompanying text.

465. *Cf.* John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965, 999-1000 (1984) (proposing that the uncertainty may be reduced by an enhanced fact-finding process, the promulgation of enforcement guidelines, or the implementation of a bright-line test).

466. *See supra* notes 214-17 and 423-37 and accompanying text.

467. *See* Hall, *supra* note 246, at 121, 127-28, 144-45.

468. *Id.* at 134.

469. Blumstein, *supra* note 172, at 1028; *see also* Causey v. St. Francis Med. Ctr., 719 So. 2d 1072, 1075-76 (La. Ct. App. 1998) (holding that because the statute failed to define "medically inappropriate" and "medical judgment," the case had to be sent to a medical review panel to determine the appropriate standard of care).

If doctors cannot achieve even professional consensus, they are even less likely to achieve the social consensus necessary for legislation.[470] Therefore, it seems that only a pure process-based approach like that adopted in Texas could be effective in inducing the conduct that the futility statutes intended.[471]

B. The Texas Pure Process Approach

In Texas, when a provider refuses to honor a surrogate's request for continued LSMT, the provider must commence a multi-stage review process. LSMT must be provided during this review process.[472] The first stage entails an ethics committee review of the attending physician's determination.[473] The surrogate must be notified of the ethics committee review process at least forty-eight hours before the committee meets.[474] The surrogate is also entitled to attend the meeting and to receive a written explanation of the committee's decision.[475]

If the ethics committee agrees with the treating physician that LSMT is inappropriate, the provider must attempt to transfer the patient to another provider that is willing to comply with the surrogate's treatment request.[476] The provider is obligated to continue providing LSMT for ten days after the surrogate is given the ethics committee's written decision.[477] If the patient has not been transferred or granted an extension, then the provider may unilaterally stop LSMT on the eleventh day.[478]

When the Texas Advance Directives Act (TADA) first went to Governor Bush in 1997, he vetoed the bill because it "eliminate[d] the objective negligence standard for reviewing whether a physician properly discontinued the use of life-sustaining procedures."[479] However, replacing the objective

---

470. Perhaps with the growth of palliative care and greater awareness of resource limitations, our culture will become less death-defying and more reluctant to conclude that more is better.

471. *House of Delegates Action*, *supra* note 53, at 91 (referencing the Wisconsin Medical Society Resolution 1-2004, which "support[s] the passage of state legislation which establishes a legally sanctioned extra-judicial process for resolving disputes regarding futile care, modeled after the Texas Advance Directives Act of 1999").

472. TEX. HEALTH & SAFETY CODE ANN. § 166.046(a) (Vernon 2003).

473. *Id.*

474. *Id.* § 166.046(b)(2).

475. *Id.* § 166.046(b)(4). The surrogate is also entitled to a copy of a registry with the name of providers willing to accept the patient upon transfer. *Id.* § 166.046(b)(3)(B).

476. *Id.* § 166.046(d).

477. *Id.* § 166.046(e). A court may extend this time period only if "there is a reasonable expectation" that a transfer can be made. *Id.* § 166.046(g).

478. *Id.* § 166.046(e) ("The physician and the health care facility are not obligated to provide life-sustaining treatment after the 10th day after the written decision . . . .").

479. Tex. Legis. J. 4926 (June 20, 1997), *vetoing* Tex. S. Bill 414, 76th Leg. (1997); *see also* Interim Report, *supra* note 113, at 33-34 (referencing Governor Bush's veto proclamation of the first TADA).

standard of negligence with measurable procedures was precisely the point, as reflected in the 1999 legislation that Bush did sign:

> A physician, health care professional acting under the direction of a physician, or health care facility is not civilly or criminally liable or subject to review or disciplinary action by the person's appropriate licensing board if the person has *complied with the procedures* outlined in Section 166.046.[480]

Unlike the UHCDA and other unilateral decision statutes which specify vague substantive standards such as "significant benefit," the safe harbor of TADA is defined solely in terms of process.[481] Texas providers who follow TADA's prescribed notice and meeting procedures are therefore immune from disciplinary action and civil and criminal liability.[482] Because the statute's requirements are concrete and measurable, there is little, if any, uncertainty of compliance.

The TADA is far from perfect. Ten days may not be a reasonable or sufficient time for surrogates to locate an alternative facility willing to accept the patient.[483] There may be procedural due process implications by placing the ultimate decision in the hands of an institutional ethics committee, which is comprised of physicians and administrators who look to the hospital for their economic livelihood.[484] However, these mechanics of the TADA process can and are being considerably refined.[485] The TADA demonstrates that a pure process approach works and that such an approach now serves,[486] and should continue to serve, as a model for other states.

CONCLUSION

Unilateral decision statutes provide the legal protection that health care providers have long sought for their hospital futility policies. Yet without more

---

480. TEX. HEALTH & SAFETY CODE ANN. § 166.045(d) (emphasis added).

481. *See* Iliana L. Peters, *Perspectives on the Texas "Medical Futility Statute," as Amended in 2003*, HEALTH LAW. WKLY., Oct. 22, 2004, *available at* http://www.ahla.org/hlw/issues/041022/041022_a_art_01_Peters.cfm ("Importantly, the statute does not attempt to define 'medical futility.' Any attempt to do so might result in a definition that is either too broad or too narrow.").

482. TEX. HEALTH & SAFETY CODE ANN. § 166.045(d); *see also* Truog & Mitchell, *supra* note 98, at 20 ("Clinicians in Texas may therefore be much more confident and bold in applying the policy, knowing that they are protected by the law.").

483. *Hearing on S.B. 439 Before the S. Comm. on Health and Human Servs.*, 80th Leg. (Tex. 2007).

484. *Id*; *see also Hearings on Advance Directives Before the H. Comm. on Public Health*, 80th Leg. (Tex. 2007); Burns & Truog, *supra* note 214, at 1990-91; Pope, *supra* note 460.

485. *See, e.g.*, S.B. 439, 80th Leg. (Tex. 2007) (amendments relating to advance directives and health care and treatment decisions).

486. State medical societies in Wisconsin and North Carolina have formally considered recommending TADA-type statutes to their state legislatures.

precise formulation, this authority is only illusory. The illusion will remain until there is consensus on (1) the proper ends of medicine, (2) the acceptable criteria for rationing, and (3) the legitimate restrictions on patient autonomy. Such consensus is not imminently forthcoming, however. In the meantime, providers and policymakers should look to Texas's pure process approach as a model, just as California, Vermont, and other states look to Oregon for guidance on physician-assisted suicide legislation.

DOES AN INDEPENDENT BOARD IMPROVE NONPROFIT CORPORATE GOVERNANCE?

KATHLEEN M. BOOZANG*

I. INTRODUCTION

Congress and the stock exchanges have responded to the rash of recent corporate scandals with a whirlwind of governance reforms designed to enhance transparency, officer accountability, and director independence. Despite mounting evidence that many of these costly changes have little or no bearing on corporate performance, the nonprofit sector is aggressively pursuing similar governance reforms through recommended "best practices,"[1] proposed revisions to the Model Nonprofit Corporation Act,[2] and proposed new American Law Institute (ALI) Principles of the Law of Nonprofit Organizations.[3] This voluntary revamping of nonprofit governance structures assumed a sense of urgency after Senator Chuck Grassley held Senate Finance Committee hearings on charitable corporations, with the push for stronger governance by institutional credit ratings agencies and on the greatly hyped—but, as yet, largely unfulfilled—expectation that states would enact nonprofit governance regulation mirroring Sarbanes-Oxley.[4] The Internal Revenue

---

* Professor of Law and Associate Dean for Academic Advancement, Seton Hall University School of Law. B.S., Boston College School of Management, 1981; J.D., Washington University School of Law, 1984; LL.M., Yale Law School, 1990. My thanks to the research support of Jason Watson '07 and Jeremy Jacobsen '09, Melanie DiPietro, Timothy Glynn, Thomas L. Greaney, Larry Mitchell, and Charles A. Sullivan, as well as the Seton Hall librarians.

1. *See, e.g.*, Michael W. Peregrine & Bernadette M. Broccolo, *"Independence" and the Nonprofit Board: A General Counsel's Guide*, 39 J. HEALTH L. 497, 498 (2006) (noting the recent "desire to adopt best practices" in the nonprofit sector); McDermott Will & Emery, Best Practices: Nonprofit Corporate Governance 1 (June 2004), http://www.mwe.com/info/news/wp0604a.pdf (suggesting best practices guidelines for the nonprofit sector).

2. MODEL NONPROFIT CORP. ACT THIRD EDITION (Proposed Exposure Draft 2006). This is a project of the American Bar Association Section on Business Law's Committee on Nonprofit Corporations.

3. PRINCIPLES OF THE LAW OF NONPROFIT ORGS. (Discussion Draft 2006) [hereinafter ALI DRAFT PRINCIPLES].

4. PANEL ON THE NONPROFIT SECTOR, INDEPENDENT SECTOR, STRENGTHENING TRANSPARENCY GOVERNANCE ACCOUNTABILITY OF CHARITABLE ORGANIZATIONS: A FINAL REPORT TO CONGRESS AND THE NONPROFIT SECTOR 13 (2005), *available at* http://www.nonprofitpanel.org/final/Panel_Final_Report.pdf; Peregrine & Broccolo, *supra* note 1, at 500–01.

Service (IRS) further heightened tension in 2006 by issuing a draft paper on "Good Governance Practices for 501(c)(3) Organizations."[5]

Reformers argue that board independence from management is crucial to improving oversight of management on behalf of the shareholder-owners.[6] Yet rarely-acknowledged pre- and post-Sarbanes empirical studies suggest that for-profit independent boards have had a mixed, even negative, impact on a number of key areas of performance.[7] These mixed results have led some observers to caution that corporations are better off experimenting with board composition to determine what mix best achieves the particular board's goals.[8] As one good-governance advocate has observed, "'[W]e may not know what we are doing' in the area of governance reform."[9]

With this background and the paucity of governance studies in the nonprofit sector,[10] nonprofits' willingness to jump on the independent-board bandwagon is inexplicable.[11] What is it that independent boards are supposed

---

5. IRS, U.S. Dep't Treasury, Good Governance Practices for 501(c)(3) Organizations 1 (Discussion Draft, n.d.), http://www.irs.gov/pub/irs-tege/good_governance_practices.pdf [hereinafter IRS, Good Governance Practices] (opining that boards following good governance practices are more likely to pursue a proper exempt purpose, act in the public's interest, and avoid pursuit of private interests). The IRS also intends to expand the information required on its form 990, the return filed by tax-exempt charities, effective tax year 2008. Tax-Exempt & Gov't Entities Div., IRS, Background Paper: Redesigned Draft Form 990, at 1 (n.d.), http://www.irs.gov/irs-tege/form_990_Cover_Sheet.pdf. The IRS's goals in revising the form are "enhancing transparency, promoting tax compliance, and minimizing the burden on the filing organization." *Id.* at 2.

6. *See, e.g.*, PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 75, 77; *see also, e.g.*, McDermott Will & Emery, *supra* note 1, at 2 (including board independence among their recommended "best practices" for nonprofit governance). The structural improvements being pursued in the for-profit sector seek to overcome the natural asymmetry of interests between management and investor caused by the separation of ownership and control. *See* Stephen M. Bainbridge, *Why a Board? Group Decisionmaking in Corporate Governance*, 55 VAND. L. REV. 1, 4 (2002) (discussing the efficiency justification for separation of ownership and control).

7. Robert Charles Clark, *Corporate Governance Changes in the Wake of the Sarbanes-Oxley Act: A Morality Tale for Policymakers Too*, 22 GA. ST. U. L. REV. 251, 298–99 (2005).

8. *See, e.g.*, *id.* at 301–02; *see also, e.g.*, PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 77–78 (suggesting experimentation for nonprofit boards).

9. Clark, *supra* note 7, at 302.

10. *See* Jeffrey L. Callen, April Klein & Daniel Tinkelman, *Board Composition, Committees, and Organizational Efficiency: The Case of Nonprofits*, 32 NONPROFIT & VOLUNTARY SECTOR Q. 493, 496 (2003), *available at* http://nvs.sagepub.com/cgi/content/abstract/32/4/493 (noting that "empirical literature dealing with the actual impact of nonprofit boards is far more limited, exploratory, and diffuse" than the proliferating nonprofit governance how-to literature would suggest).

11. *See, e.g.*, McDermott Will & Emery, *supra* note 1, at 2 (counseling nonprofit clients that "[a]t least a majority" of board members should be "'independent,' both in fact and appearance," and defining independence as "the absence of any material direct or indirect relationship with the Corporation").

to be uniquely able to accomplish? How many independent directors are required to ensure board independence? What evidence exists that independent boards are effective at achieving the articulated goals—and are such goals even quantifiable and measurable? There have been no convincing answers to any of these questions, nor has there been any clear articulation of why nonprofit boards, in particular, should be independent.

The law regulating nonprofit governance remains surprisingly undeveloped. For example, scholars, attorneys general, and corporate counsel do not even agree on the answer to a most basic corporate law question: To whom is the nonprofit board accountable?[12] While for-profit directors are ultimately and primarily responsible to shareholders and generally pursue the goal of maximizing stock value, nonprofit boards have multiple constituencies and operate with few guiding principles as to how to prioritize competing claims for their resources.[13] In short, the drive for independent directors in the nonprofit sector seems directionless.

Several factors explain the disconnect between nonprofit governance reform and the very essence of nonprofit entities. Part of the problem is that proposed reforms are untethered from a realistic view of the myriad entities that comprise the nonprofit sector. Although the sector is overwhelmingly composed of nonprofit entities with annual revenues of less than $1 million (many of those with revenues under $25,000 per year),[14] it also includes multibillion dollar commercial entities that operate in extremely competitive markets. Directors are largely volunteers, and while nonprofit boards tend to be dominated by CEOs or directors of business corporations, they also include—especially as they become smaller—passionate community organizers with no

---

12. *See, e.g.*, Thomas L. Greaney & Kathleen M. Boozang, *Mission, Margin, and Trust in the Nonprofit Health Care Enterprise*, 5 YALE J. HEALTH POL'Y L. & ETHICS 1, 14–15 (2005) (noting nonprofit health plans' relationship with the serviced community); Henry B. Hansmann, *The Role of Nonprofit Enterprise*, 89 YALE L.J. 835, 845 (1980) (describing "the purpose of the [corporate] charter [as] primarily to protect the interests of the organization's *patrons* from those who control the corporation"); Lumen N. Mulligan, *What's Good for the Goose Is Not Good for the Gander: Sarbanes-Oxley-Style Nonprofit Reforms*, 105 MICH L. REV. 1981, 2006–07 (2007) (discussing the conclusion of several scholars that, much like for-profit corporations are beholden to stockholders, nonprofits primarily owe duties to stakeholders).

13. *See* Mulligan, *supra* note 12, at 2006 (observing that these multiple constituencies frequently push towards differing goals, which "affects the board's concept of accountability").

14. NAT'L COUNCIL NONPROFIT ASS'NS, THE UNITED STATES NONPROFIT SECTOR 1 (2006), http://www.ncna.org/_uploads/documents/live//us_sector_report_2003.pdf (stating that 66% of all nonprofit public charities had revenues of less than $25,000 in the year 2003). Between 1996 and 2006, small nonprofits with yearly revenues of under $25,000 more than doubled the growth of nonprofits overall, growing by more than 78% nationally, Nat'l Ctr. for Charitable Stat., Urb. Inst., Number of Nonprofit Organizations in the United States, 1996–2006, http://nccsdataweb.urban.org/PubApps/profile1.php?state=US (last visited Nov. 7, 2007), and by over 100% in the state of New York, Nat'l Ctr. for Charitable Stat., Urb. Inst., Number of Nonprofit Organizations in New York, 1996–2006, http://nccsdataweb.urban.org/PubApps/profile1.php?state=NY (last visited Nov. 7, 2007).

exposure to the corporate boardroom and rabbis and ministers whose vocation is service to God rather than financial management. Additionally, definitions of nonprofit director independence, which are modeled on those from the for-profit context, fail to reflect the inextricable link between a nonprofit's purpose for existence and its board's need to be composed of individuals dedicated to, and steeped in, that mission.[15]

Early results of governance reform suggest that "corporate compliance" supersedes preservation and pursuit of mission in many of today's nonprofit boardrooms.[16] Unquestionably, there exist nonprofit directors who act in self-interest, behave illegally (if often out of naïveté), or mishandle the assets entrusted to their stewardship.[17] Nevertheless, a disproportionate focus on legal and financial accountability, with the attendant pressure to appoint directors qualified for performing compliance activities, can divert attention from the more important question: What kind of board will best steward the entity's resources as it pursues its mission and serves its constituencies? Bureaucracy driven by oversight concerns threatens to stifle the volunteerism and democratic experimentation that have been the hallmarks of charities since their inception.[18] Corporate compliance, which requires business sophistication, threatens to exacerbate the exclusion of beneficiaries, congregants, and the ordinary civic-minded activist from nonprofit boards. The current governance reform movement ignores the radical differences in purpose and role between the business and the nonprofit corporation. Indeed, the mission orientation of charitable corporations raises the question of whether independent boards even make sense.

In sum, the most important insight to be gained from the Sarbanes-Oxley governance reform is that the nonprofit's primary goal of mission integrity,

---

15. *See* Greaney & Boozang, *supra* note 12, at 1 (arguing that governance reform has not established "mechanisms to ensure fidelity to the organization's charitable mission").

16. *See generally* Mulligan, *supra* note 12 (arguing that Sarbanes-Oxley-type governance reforms hamper the charitable efforts of nonprofits). One could argue that internal accountability is even more important in the nonprofit because there is so little government oversight, with respect to either mission fidelity or use of charitable assets. *See, e.g.*, PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 13 (noting that the "serious shortage of resources has often made it difficult for government officials to identify and punish most violators" in the nonprofit sector); MARION R. FREMONT-SMITH, GOVERNING NONPROFIT ORGANIZATIONS 443 (2004) (observing that only twelve states' attorney general offices exercise their oversight of charities "in a manner that impacts positively on the behavior of charitable fiduciaries"). Because Congress has severely limited the IRS's oversight of religious organizations, *id.* at 9, accountability of churches and related organizations may be an even more significant concern.

17. Hansmann, *supra* note 12, at 874–75. However, nonprofit wrongdoing is difficult to quantify, largely because of the persistently poor oversight of regulators. FREMONT-SMITH, *supra* note 16, at 13.

18. *See generally* Dana Brakman Reiser, *Dismembering Civil Society: The Social Cost of Internally Undemocratic Nonprofits*, 82 OR. L. REV. 829 (2003) [hereinafter Reiser, *Dismembering Civil Society*] (discussing the detrimental effects of the trend toward undemocratic boards in nonprofits).

with management oversight in a secondary role, might not be best accomplished by independent boards. These goals might be served just as effectively by: encouraging nonprofit boards to seek diversity in the skill sets of their directors and to aim for the representation of their constituencies, especially those not able to make significant financial donations; closing the gaps in current nonprofit statutes that permit weak governance structures; statutorily requiring financial audits of nonprofits over a certain size; recommending the presence of "monitoring directors"; and legally imposing an aggressively expanded conception of transparency.

Part II of this Article presents two "morality tales" of the disasters that can occur in nonprofits overseen by dysfunctional boards that lose sight of their role and the entity's purpose or prioritize self-interest over the nonprofit's mission. Part III explores the concept of board independence, beginning with its application in the corporate realm and then extrapolating to the nonprofit sector. Finally, Part IV offers a critique of current statutory attempts to regulate nonprofit governance and presents suggestions for strengthening nonprofit governance.

II. STORIES OF NONPROFIT DIRECTORS' FAILURES

A. The New York Stock Exchange: A Nonprofit Board that Should Have Known Better

After more than a decade of respected leadership,[19] Richard Grasso resigned as CEO and Chairman of the then-nonprofit New York Stock Exchange (NYSE) in September 2003 amid controversy about the size of his compensation package.[20] For governance specialists, more troubling than the alleged excessiveness of Grasso's pay and benefits was how it happened. Mr. Grasso allegedly manipulated board committee and chair appointments—particularly of the compensation committee—and got away with providing incomplete information to directors, many of whom were ultimately unaware of the substance of the compensation packages they approved.[21] These details

19. Grasso began his tenure with the NYSE in 1968; after holding various executive positions over the years, he became CEO and Chairman of the Board in 1995. DAN K. WEBB, WINSTON & STRAWN LLP, REPORT TO THE NEW YORK STOCK EXCHANGE ON INVESTIGATION RELATING TO THE COMPENSATION OF RICHARD A. GRASSO 7 (2003), http://fl1.findlaw.com/news.findlaw.com/nytimes/docs/grasso/webrptgrassonyse1203.pdf [hereinafter WEBB REPORT].

20. *See id.* at 1. Dan K. Webb of Winston & Strawn was retained by the NYSE to analyze Grasso's compensation package. *Id.* The resultant report is now popularly known as the "Webb Report."

21. *Id.* at 25, 96–99. This is partially attributed to committee turnover, which resulted in newer members' lack of understanding regarding the progression of Grasso's compensation package. *Id.* at 25.

were elicited by New York's then-Attorney General Eliot Spitzer, who sued both the NYSE and individual directors, including Grasso, for breaching their fiduciary duties under the New York Not-for-Profit-Corporation Law (N-PCL).[22]

From its inception in 1970 until March 2006, the NYSE was incorporated as a Type A not-for-profit private membership organization under the N-PCL.[23] The NYSE operated as a board of trade; its members comprised the corporations that paid dues for seats on the Exchange, which permitted them to effect trades.[24] Like any other corporation with close to a billion dollars in annual revenue, the NYSE determined its executives' compensation through a process of benchmarking established and implemented by its compensation committee.[25]

During his tenure, Mr. Grasso executed three separate employment agreements that made him eligible for a variety of benefits programs in addition to his base salary.[26] The benefits included a deferred compensation program, two different kinds of incentive compensation plans, and a supplemental retirement plan.[27] Mr. Grasso's annual salary of $1.4 million[28] was really the most inconsequential part of his compensation package. His incentive awards annually exceeded $13 million, and his 2003 employment agreement, which he signed weeks before his departure, included a $139.5 million immediate lump sum deferred compensation payment as well as a subsequent $48 million payout.[29] Had the board added everything up, it would have discovered that Mr. Grasso's annual compensation package fell just short of the NYSE's net income in some years.[30] The Webb Report concluded that "[i]n total, Grasso

---

22. *See* People *ex rel.* Spitzer v. Grasso (*Grasso II*), 836 N.Y.S.2d 40, 41 (App. Div. 2007).

23. *Id.* at 54 (Mazzarelli, J.P., dissenting). In 2006, the NYSE reorganized into two separate entities: "a New York not-for-profit regulatory entity and . . . a Delaware for-profit public corporation." *Id.* at 55.

24. *Id.* at 54.

25. WEBB REPORT, *supra* note 19, at 22.

26. People *ex rel.* Spitzer v. Grasso (*Grasso I*), 816 N.Y.S.2d 863, 866 (Sup. Ct. 2006), *rev'd, Grasso II*, 836 N.Y.S.2d 40.

27. *Id.* at 867.

28. *Id.* at 866.

29. *Id.* at 867. In contrast, at the time Grasso's predecessor, William Donaldson, left the NYSE, Donaldson was receiving $1.65 million in annual compensation and had a pension worth approximately $3.6 million. WEBB REPORT, *supra* note 19, at 7.

30. *Grasso I*, 816 N.Y.S.2d at 867 (citing Spitzer's complaint). In support of their motion to dismiss, the defendants argued that because the members of the NYSE were sophisticated, the state attorney general did not need to pursue claims on their behalf. *Id.* at 870. The trial court disagreed, observing that, because of its nonprofit status, the stock exchange lacked shareholders with any financial incentive to prosecute board irregularities: "'Given the absence of shareholders, profits and other market devices to ensure the efficacy of contracts and regularity of operations, the statute contemplates significant public oversight of the finances and major transactions of such entities.'" *Id.* (quoting 64th Assocs., L.L.C. v. Manhattan Eye, Ear &

received approximately $144.5 million to $156.7 million in excessive compensation and benefits."[31] Apparently, the "excessive" amounts of Grasso's benefits partly resulted from his repeatedly being allowed to withdraw from his retirement plans even while he remained employed by the NYSE.[32]

In his complaint, Spitzer alleged that the NYSE board breached its fudiciary duties by allowing Grasso's compensation to balloon to such levels.[33] Furthermore, Spitzer charged that the ultimate blame rested with Grasso and that Grasso's management of the committee appointment process constituted a breach of his fiduciary obligations.[34] In addition to his strong influence over the nominating committee and the director selection process—Grasso believed that only CEOs should be appointed and provided a list of names to the committee each year[35]—Grasso also controlled the compensation committee chair and membership appointment process and appointed several of his friends to those positions.[36] The compensation committee frequently received incomplete or incorrect information regarding the key components of Grasso's compensation package.[37] In addition, Grasso set the criteria for annual "Chairman's Award" rewards to senior executives, which constituted 35% of the formula employed in assessing his own annual performance.[38]

Throat Hosp., 813 N.E.2d 887, 889 (N.Y. 2004)). The court further observed that the attorney general had an interest in protecting investors, many of whom were New York residents. *Id.* at 871, 873.

31. WEBB REPORT, *supra* note 19, at 2.

32. *Id.* at 3.

33. *See Grasso I*, 816 N.Y.S.2d at 866, 869. The appellate court rejected four common law causes of action based upon violations of the statute because the legislature had not authorized the attorney general to pursue them. *Grasso II*, 836 N.Y.S.2d 40, 41-42 (App. Div. 2007).

34. *See Grasso I*, 816 N.Y.S.2d at 866, 869.

35. People *ex rel.* Spitzer v. Grasso (*Grasso III*), No. 401620/04, 2006 WL 3016952, at *5 (N.Y. Sup. Ct. Oct. 18, 2006); WEBB REPORT, *supra* note 19, at 96–97 (reporting that one former director referred to the nominating committee as Grasso's "team").

36. WEBB REPORT, *supra* note 19, at 98; *see also id.* at 98–99 (stating that Grasso had "unfettered authority" over appointments to the compensation committee and seemed to favor appointment of directors with poor meeting attendance records). Spitzer also alleged that Grasso used his regulatory authority to punish and reward the firms whose directors did or did not vote his way at board and committee meetings. *Grasso I*, 816 N.Y.S.2d at 868.

37. WEBB REPORT, *supra* note 19, at 5–6. Individual committee members were shown the primary sources upon which the compensation calculations were based in pre-committee half-hour meetings with the head of Human Resources. *Id.* at 43. However, they were allowed to retain for review only worksheets, which were not always provided in advance of the meeting and were not always identical to the worksheets actually used to assess the recommendations being made to the committee. *Id.* at 43–44.

38. *See id.* at 5, 28. The Webb Report discusses the standards set for the Chairman's Award between 1995 and 2002. *Id.* at 28–30. The Chairman's rating exceeded its target in each year that Grasso was CEO, which caused some committee members to question whether the targets were set too low. *Id.* at 29.

Yet the blame cannot rest entirely with Grasso. The compensation committee obtained insufficient support from compensation consultants.[39] It also relied in "unconventional ways" upon incomplete data from the wrong compensation comparator group—for-profit entities—in benchmarking Grasso's salary.[40] By 2003, at least some directors did not even understand Grasso's compensation package before voting for it.[41]

The NYSE board was independent in the sense that the term is used in Sarbanes-Oxley; the directors were not NYSE employees, and Grasso chaired neither the nominating nor the compensation committee.[42] Grasso likely oversaw some of the Stock Exchange's best years; but for the controversy over his compensation, he might still hold his position as CEO. Thus, while one clearly could argue that plenty was wrong with the governance of the NYSE, most indicia of good governance were technically present. Nevertheless, a closer look reveals that the board's independence was a mere specter; many of the directors were Grasso's friends, while others were successfully negotiating the politics and politeness that pervade almost every significant enterprise.[43]

B. University of Medicine and Dentistry of New Jersey: A Morality Tale on the Stewardship of Self-interest

The breadth and depth of problems at the University of Medicine and Dentistry of New Jersey (UMDNJ),[44] the nation's largest public hospital system,[45] will probably never be completely unraveled.[46] Despite the stories of

---

39. *Id.* at 118–19.

40. *Id.* at 4. While the committee discussed including other exchanges among the benchmarks, it never did. *Id.* at 31. The committee also decided not to use investment banker salaries and at no time did it seriously consider using not-for-profit CEO salaries as benchmarks. *Id.* Although the group disagreed about using financial services salaries, the ultimate comparator group was heavily weighted in favor of this industry; the factors used in selecting the peer group did not include firm size and financials, which are typical in setting executive compensation. *Id.* at 32–33. Furthermore, the comparator group was not adjusted from year to year. *See id.* at 35.

41. *Id.* at 5. The committee monitored neither the growth in his pension plans, nor the impact of awarding huge bonuses on the accumulation of his retirement benefits. *Id.* at 3.

42. See *id.* at 96.

43. *See* WEBB REPORT, *supra* note 19, at 98–99.

44. See Lisa Brennan, *Exit Plan for UMDNJ Monitor*, 185 N.J. L.J. 737, 740 (2006) [hereinafter Brennan, *Exit Plan*] (reporting that a federal monitor "found $243 million in waste and fraud" in his investigation of UMDNJ).

45. UMDNJ has an annual budget of $1.6 billion, more than 5500 students, and over 14,000 employees on five campuses that occupy 185 acres. *Id.*; Sue Reisinger, *Casualty of Deferred Prosecution: Did UMDNJ's General Counsel Take the Fall for the School's Administrators? Or Was She Part of the Problem?*, 186 N.J. L.J. 397, 397 (2006). UMDNJ operates three medical schools, New Jersey's only dental school, graduate schools of biomedical sciences and public health, a school for the health-related professions, and a nursing school; its facilities treat more than two million patients each year. UMDNJ Fast Facts (Aug. 2007),

directors' conflicts and absences that regularly dribbled out of UMDNJ, it still came as a surprise when the United States Attorney for the District of New Jersey dramatically appeared at a meeting of the UMDNJ board of directors, reporters in tow, and presented it with a draft of a criminal complaint.[47] The board was given a choice between criminal prosecution or submission to the oversight of a federal monitor and the conditions of a Deferred Prosecution Agreement (DPA).[48] This dramatic opening scene was merely the prelude to months of investigation, firings, whistleblower suits, and grand jury hearings.[49]

---

http://www.umdnj.edu/about/fastfacts.pdf; *see* ADA.org, DDS/DMD Programs – U.S., http://www.ada.org/prof/ed/programs/search_ddsdmd_us.asp (last visited Nov. 12, 2007) (identifying UMDNJ's New Jersey Dental School as the only accredited dental education program in the state).

46. While the problems at UMDNJ are uniquely egregious, board oversight has unquestionably been an issue at other schools. *See generally* Julianne Basinger, *Boards Crack Down on Members' Insider Deals: Recent Scandals Trigger New Scrutiny of Trustees*, CHRON. HIGHER EDUC. (Wash., D.C.), Feb. 6, 2004, at A1, *available at* http://chronicle.com/weekly/v50/i22/22a00101.htm (noting Auburn's placement on accreditation probation because of trustee behavior; questionable business deals between Boston University and its trustees; and criticism of The University of Georgia Foundation for doing more than $30 million of business with companies linked to many of its trustees); Richard P. Chait, *When Trustees Blunder*, CHRON. HIGHER EDUC. (Wash., D.C.), Feb. 17, 2006, at B6, *available at* http://chronicle.com/weekly/v52/i24/24b00601.htm (discussing "dysfunctional governance" at Adelphi University, American University, Boston University, and Cornell University, but arguing that the "more pervasive and corrosive" issue in higher education is "mediocre governance"); Paul Fain, *Thanks, Enron: Auditors Gain Clout on Campuses*, CHRON. HIGHER EDUC. (Wash., D.C.), June 10, 2005, at A1, *available at* http://chronicle.com/weekly/v51/i40/40a00101.htm (discussing the 1991 federal grant scandal at Stanford that led to Congressional revamping of distribution of federal research funds); Sara Hebel, Paul Fain & Goldie Blumenstyk, *Relations Between Presidents and Boards Top Agenda at Leadership Forum*, CHRON. HIGHER EDUC. (Wash., D.C.), June 23, 2006, at A31, *available at* http://chronicle.com/weekly/v52/i42/42a03101.htm (quoting a forum participant's statement that "'[g]overnance is in a state of upheaval' [because] the misbehavior and greed of a few college presidents have led to an increased focus on oversight in higher education"). *But see* Alexander E. Dreier, *Sarbanes-Oxley and College Accountability*, CHRON. HIGHER EDUC. (Wash., D.C.), July 8, 2005, at B10, *available at* http://chronicle.com/weekly/v51/i44/44b01001.htm (observing that "notably few" nonprofit scandals involved colleges or universities and indicating that many of the reporting requirements being considered by Congress were ill-suited to university culture, which does not subscribe to "top-down accountability").

47. *See* Scott Goldstein, *The War on Corruption Is Just Getting Started*, NJBIZ (New Brunswick, N.J.), Jan. 16, 2006, at 3, *available at* 2006 WLNR 1570040. The complaint alleges that both University Hospital and the faculty practice plan, University Physician Associates (UPA), knowingly billed Medicaid for the same outpatient physician services. Complaint at Attach. A ¶ 2, United States v. Univ. Med. & Dentistry N.J., No. 05-3134 (D.N.J. Dec. 29, 2005). The double billing continued even after outside counsel recommended that UMDNJ notify Medicaid of the double billing and tell UPA that it must cease billing for outpatient clinic services. *Id.* at ¶¶ 5, 9.

48. The U.S. Attorney agreed to recommend deferring the prosecution of the criminal

As a first step of reform, the then-acting governor initiated partial reconstitution of the board;[50] the New Jersey legislature then spent months haggling over proposed legislation that would specifically allocate the power to appoint future boards.[51] But to suggest that a new board or radically altered corporate structure by itself will cure an organization that is clearly infected to its core[52] is simply naïve. The antidotes will be many and varied. Nonetheless, an examination of the troubles uncovered and the solutions employed is enlightening.[53]

---

complaint against UMDNJ for 24 to 36 months, Deferred Prosecution Agreement ¶ 4 (n.d.), http://www.umdnj.edu/about/board/DFAFinalsigned.pdf, but he retained the right to investigate and prosecute any individual trustee, officer, agent, employee or attorney, *id.* at ¶ 22.

49. *See, e.g.*, Brennan, *Exit Plan*, *supra* note 44, at 737, 740; Lisa Brennan, *UMDNJ Paid No-Show Cardiologists to Lure Patients, U.S. Monitor Reports*, 186 N.J. L.J. 702, 702 (2006) [hereinafter Brennan, *No-Show Cardiologists*]; *Medgate?*, 181 N.J. L.J. 647, 647 (2005).

50. *See* Kelly Heyboer, *Lobbyist Withdraws Her Nomination to UMDNJ Board*, STAR-LEDGER (Newark, N.J.), Aug. 16, 2006, at 20, *available at* 2006 WLNR 14186409 [hereinafter Heyboer, *Lobbyist Withdraws*] (noting that, in 2005, "three UMDNJ trustees, including the chairwoman, resigned after the state enacted new ethics rules banning board members and their families from doing any business with the schools they serve"); *see also* Kelly Heyboer, *Ethics Board Finds Four UMDNJ Trustees Can Remain*, STAR-LEDGER (Newark, N.J.), Jan. 21, 2006, at 8, *available at* 2006 WLNR 1165185 (discussing Acting Governor Codey's executive order forbidding board members or their families from having business ties to their schools). UMDNJ is a state entity, created and governed by state statute. N.J. STAT. ANN. ch. 64G (West Supp. 2007). One of the statutes gives the governor considerable influence over the composition of the board of trustees. *See id.* § 18A:64G-4. Senate President Richard Codey was serving as Acting Governor following former Governor James McGreevey's 2004 resignation. Rosa Cirianni, *A New Life, but Paperwork Still to Be Done*, STAR-LEDGER (Newark, N.J.), Nov. 17, 2004, at 20, *available at* 2004 WLNR 20242939.

51. Ted Sherman & Josh Margolin, *Monitor Chastises UMDNJ: University Ordered to Start Looking for a New Leader*, STAR-LEDGER (Newark, N.J.), June 26, 2006, at 13, *available at* 2006 WLNR 11064738. The legislature eventually enacted legislation that created separate boards for the university and hospital. *See* N.J. STAT. ANN. ch. 64G; Angela Stewart, *A Call for Harmony on Hospital Board*, STAR-LEDGER (Newark, N.J.), Aug. 8, 2007, at 19, *available at* 2007 WLNR 15278287. Because the hospital is not a separate corporation, implementation of the statute became bogged down in confusion over the relationship between the two boards and what powers are actually vested in the hospital board. *Cf.* Governor's Statement on A-2900, § 18A:64G-4 annot. (anticipating such confusion).

52. UMDNJ was widely known as a "patronage pit" for the hiring of friends and family of both trustees and politicians. Josh Margolin & Ted Sherman, *How UMDNJ Became a "Patronage Pit,"* STAR-LEDGER (Newark, N.J.), Apr. 4, 2006, at 1, *available at* 2006 WLNR 5644745. The system became so pervasive that the last pre-DPA president established a ranking methodology, giving job applicants scores of 1 to 3 depending upon the strength of their connections. *Id.*

53. Senators Grassley and Baucus of the Senate Finance Committee sought assurances from Governor Corzine that the issues at UMDNJ would be brought under control and that fraud on federal health care plans would discontinue. *See* Letter from Chuck Grassley, Chairman, & Max Baucus, Ranking Member, U.S. Senate Comm. on Fin., to Jon S. Corzine, Governor of N.J. (Jan. 30, 2006), *available at* http://finance.senate.gov/

Months of investigation by the federal monitor confirmed that the nonprofit sector can indeed produce scandals rivaling those of Enron and WorldCom. The revelations of "fraud, waste, influence-peddling and hiring improprieties"[54] resulted in UMDNJ's president,[55] general counsel and other lawyers in the Office of Legal Management,[56] CFO,[57] all three medical school deans,[58] the dental school dean,[59] prominent physician-leaders,[60] a number of

---

press/Bpress/2005press/prb013006.pdf; *see also* Ted Sherman & Josh Margolin, *U.S. Senate Opens Probe of UMDNJ Panel; Says Alleged Medicare Fraud May Hit "Tens of Millions of Dollars,"* STAR-LEDGER (Newark, N.J.), Jan. 31, 2006, at 1, *available at* 2006 WLNR 1715710 (discussing the letter sent by Grassley and Bauer). UMDNJ is not the only university subject to investigation by the Senate; following a multi-month investigation, Senator Grassley sent a letter to the American University Board of Trustees in which he charged it with having "ignored damaging audit findings on lavish spending by the university's former president, disregarded possible Internal Revenue Service sanctions, and proposed retribution against a whistle-blower." Paul Fain, *U.S. Senator Threatens Action Against American University Board*, CHRON. HIGHER EDUC. (Wash., D.C.), May 26, 2006, at A33, *available at* http://chronicle.com/weekly/v52/i38/38a03301.htm.

54. Brennan, *Exit Plan, supra* note 44, at 737.

55. Josh Margolin & Kelly Heyboer, *School Delays Paying Petillo His Severance: Options Sought if Outgoing Leader Is Ever Convicted*, STAR-LEDGER (Newark, N.J.), Jan. 26, 2006, at 3, *available at* 2006 WLNR 1438533. Although UMDNJ's then-president, John Petillo, had "not been accused of any wrongdoing in the ongoing corruption probe," he was encouraged to resign by Governor Corzine, "who wanted the university to get a fresh start with a new leader." *Id.*

56. Lisa Brennan, *UMDNJ Counsel Under Investigation for Complicity in Sinecure Scheme*, 186 N.J. L.J. 884, 884 (2006) [hereinafter Brennan, *Sinecure Scheme*] (reporting that, one year into his tenure, UMDNJ's acting general counsel was placed on leave for his alleged involvement in a fraudulent cardiology kickback scheme). The federal monitor charged that the counsel's office withheld crucial details about the scheme from outside counsel from whom it sought, but never received, a final opinion letter. HERBERT J. STERN, INTERIM REPORT OF THE FEDERALLY-APPOINTED MONITOR FOR THE UNIVERSITY OF MEDICINE AND DENTISTRY OF NEW JERSEY 4–5 (2006), *available at* http://www.umdnj.edu/home2web/federal%20monitor/pdf/report11146c.pdf [hereinafter UMDNJ INTERIM REPORT].

57. Josh Margolin, *Whistleblower Files Suit Against UMDNJ*, STAR-LEDGER (Newark, N.J.), Dec. 23, 2006, at 17, *available at* 2006 WLNR 22433193.

58. Ted Sherman, *Money, Management Woes Bring Probation for UMDNJ*, STAR-LEDGER (Newark, N.J.), June 28, 2006, at 16, *available at* 2006 WLNR 11196732 [hereinafter Sherman, *Money, Management Woes*]. The dean of the School of Osteopathic Medicine resigned amid allegations of abuse of his expense account. Ted Sherman & Josh Margolin, *Report on UMDNJ Exposes Thousands in Misspent Funds*, STAR-LEDGER (Newark, N.J.), Apr. 24, 2006, at 1, *available at* 2006 WLNR 6843235. Other high-ranking academic officers were affected as well, including the School of Osteopathic Medicine's senior associate dean for academic and student affairs. Josh Margolin & Ted Sherman, *A Top Dean at UMDNJ Fired After Raid by FBI*, STAR-LEDGER (Newark, N.J.), June 2, 2006, at 1, *available at* 2006 WLNR 9486981 (noting that this dean "was terminated after an FBI raid on his campus office . . . to halt the destruction of internal documents sought by the U.S. Attorney's Office").

59. *Dental School to Bar 20 from Graduation*, RECORD (Hackensack, N.J.), May 19, 2006, at A3. The dental school dean was forced to resign after it emerged that he had "'cooked

cardiologists,[61] and a state senator working for the university as a lobbyist[62] losing their positions. Others suffered significant pay cuts,[63] and some—one dean and a state senator—even faced indictment.[64] The grand jury investigation continues to this day.[65]

In the waning days of 2006, the federal monitor notified the U.S. Attorney that UMDNJ had violated the terms of its DPA by failing to report an illegal kickback scheme.[66] UMDNJ had allegedly placed eighteen community cardiologists in part-time clinical faculty positions at above-market salaries calculated in part upon the physicians' prospective referral estimates; in return,

---

the books' to ensure his own bonus" and had abused his expense account. *Id.*

60. Josh Margolin, *2 Alleged "No-Shows" to Be Fired at UMDNJ*, STAR-LEDGER (Newark, N.J.), Nov. 17, 2006, at 1, *available at* 2006 WLNR 19989959 [hereinafter Margolin, *2 Alleged No-Shows*] (describing the termination of "two cardiologists who allegedly were given no-show faculty jobs as part of a [kick-back] scheme"); Josh Margolin, *UMDNJ Puts Top Doctor on Leave: Exec Also Implicated in Referrals Scheme*, STAR-LEDGER (Newark, N.J.), Nov. 22, 2006, at 1, *available at* 2006 WLNR 22822797 [hereinafter Margolin, *Top Doctor on Leave*] (reporting that the chairman of medicine, a world-famous infectious disease specialist, was placed on administrative leave during the investigation of his involvement in a cardiology kick-back scheme and subsequently resigned).

61. Margolin, *2 Alleged No-Shows*, *supra* note 60.

62. Jennifer Moroz & Troy Graham, *U.S. Says Politician a UMDNJ No-Show*, PHIL. INQUIRER, Sept. 19, 2006, at B8, *available at* 2006 WLNR 16226298 (reporting a federal monitor's charges that a $35,000 per year "program support coordinator" position was created to hide payments to a state senator to essentially "lobby himself," as chairman of the Senate Budget and Appropriations Committee, on UMDNJ's behalf).

63. Margolin, *2 Alleged No-Shows*, *supra* note 60 (noting that, in addition to the two cardiologists facing termination, "[m]ost of the remaining doctors [in the cardiology program] will have their salaries cut—in most cases, by more than half").

64. Josh Margolin & Ted Sherman, *UMDNJ Suspends Its Top Official in Camden Amid Fiscal Allegations*, STAR-LEDGER (Newark, N.J.), June 13, 2007, at 15, *available at* 2007 WLNR 11045779 [hereinafter Margolin & Sherman, *Top Camden Official*] (noting the indictment of the osteopathic school's dean for "rigging UMDNJ's hiring procedures to put state Sen. Wayne Bryant . . . on the university payroll" for a no-show job); Elise Young, *Senator Indicted in Graft Probe*, RECORD (Bergen County, N.J.), Mar. 30, 2007, at A1, *available at* 2007 WLNR 6062559.

65. *See* Margolin & Sherman, *Top Camden Official*, *supra* note 64; Sherman, *Money, Management Woes*, *supra* note 58. At least one whistleblower suit alleges a cover-up, including destruction of evidence and ignoring grand jury subpoenas. *See* Josh Margolin, *New Whistle-Blower Sues UMDNJ: Fired Finance Official Alleges Higher-Ups Used Complex Plan to Conceal Double-Billing*, STAR-LEDGER (Newark, N.J.), Dec. 1, 2006, at 25, *available at* 2006 WLNR 20775363 [hereinafter Margolin, *New Whistleblower*]. In late 2005, UMDNJ experienced a break-in and theft of documents being prepared for presentation to the grand jury. *Medgate?*, *supra* note 49. The grand jury is investigating UMDNJ's government affairs office's approach to making political contributions to "friends" of the university, among other allegations. *See* Josh Margolin, *UMDNJ Whistleblower Files Suit Over Her Termination*, STAR-LEDGER (Newark, N.J.), Nov. 18, 2006, at 13, *available at* 2006 WLNR 22822798.

66. UMDNJ INTERIM REPORT, *supra* note 56, at 2, 8-9.

the cardiologists had no responsibilities beyond the understanding that they would refer their patients to UMDNJ.[67] The monitor also alleged that UMDNJ took affirmative steps to hide the nature of the deal from investigators.[68] The U.S. Attorney and federal monitor had learned of the arrangement with the cardiologists upon reading a *New Jersey Law Journal* article about a settlement with a whistleblower.[69]

Multiple whistleblower suits suggest that there were employees who spoke up about the improprieties, but whose efforts were at best ignored and at worst punished. At least three of these former employees have sued the university along with its former and current executives; they allege the university fired them for voicing their concerns about behavior ranging from illegal billing and physician kickbacks to political slush funds.[70] Moreover, the university's recent problems are not exclusively legal. The Middle States Commission on Higher Education placed the university on probation for its failure to have

---

67. *Id.* at 5, 8, 11-12.

68. *Id.* at 2.

69. *Id.* The University "settled a lawsuit by its former chief of cardiology for $2.2 million, after he alleged that he was forced to leave the university because of his objections to an illegal scheme giving physicians no-show faculty jobs in exchange for patient referrals." Margolin, *New Whistleblower*, *supra* note 65.

The kickback scandal received much attention from the press. *See, e.g.*, Editorial, *UMDNJ's Violation of Trust*, STAR-LEDGER (Newark, N.J.), Nov. 13, 2006, at 14, *available at* 2006 WLNR 1969872; Josh Margolin, *At UMDNJ, an Attempt to Cover Up $36M Fraud*, STAR-LEDGER (Newark, N.J.), Nov. 12, 2006, at 1, *available at* 2006 WLNR 19653436; Ted Sherman & Josh Margolin, *How UMDNJ Pumped Up its Heart Program*, STAR-LEDGER (Newark, N.J.), Nov. 5, 2006, at 1, *available at* 2006 WLNR 19212954. The kickbacks were apparently part of the University's attempts to salvage its state cardiac surgery license, which was on the verge of revocation, UMDNJ INTERIM REPORT, *supra* note 56, at 2-3; Josh Margolin, *UMDNJ Bribe Charges Face U.S. Scrutiny*, STAR-LEDGER (Newark, N.J.), Nov. 14, 2006, at 27, *available at* 2006 WLNR 19770343, in part due to an inadequate number of procedures, loss of accreditation, and a 5–6% death rate, Patricia Alex, *New Trustees Join Med School in Midst of Financial Struggle*, RECORD (Hackensack, N.J.), June 28, 2006, at A4, *available at* 2006 WLNR 11196396; Brennan, *Sinecure Scheme*, *supra* note 56; Editorial, *supra*, at 14.

70. Margolin, *New Whistleblower*, *supra* note 65 (noting that lawsuits have been brought by a billing manager who "was harassed, demoted, then banished to a makeshift office in the lunchroom before he was suspended" after discussing illegal contract bidding procedures with the federal monitor; a government affairs worker who alleges she was fired after testifying before a grand jury about a political slush fund; and a senior finance official who alleges she was suspended and eventually fired for testifying about the Medicare billing and kickback schemes); *see also* Ted Sherman, *Exec Says He Faced Retaliation at UMDNJ*, STAR-LEDGER (Newark, N.J.), Nov. 29, 2006, at 1, *available at* 2006 WLNR 20632572 [hereinafter Sherman, *Exec Faced Retaliation*] (describing the billing manager whistleblower's allegations that "[t]o conceal the exorbitant contracts, the department was sending fake invoices that billed other university departments for millions of dollars of communications costs that they had, in fact, never incurred" and reporting that the federal monitor discovered the university had paid $35.2 million for a $5.9 million contract).

properly audited its 2005 financial statements,[71] and Moody's Investor Services downgraded its credit rating twice.[72]

Of course, the ultimate question is, "What was the board doing while all of this was going on?" Individual board members were part of the problem, using their positions to obtain employment for friends and family and engaging in conflicted transactions with the university.[73] Furthermore, the board was not active, sometimes because it was not provided with the information needed to fulfill its obligations.[74] The board also had little control of the university's spending, as discovered by the federal monitor who reported that there was "'virtually no oversight' on $104 million a year in no-bid contracts, and [the university] paid $88.3 million to various vendors between 2001 and 2005 without approved purchase orders."[75] Even where policies were in place, evidence suggests that the university managers circumvented required board approval rather easily.[76]

Upon becoming interim president, former Health Care Financing Administration (HCFA) Administrator Bruce Vladeck observed that he was worried the board was barely meeting the minimal requirements of the hospital accrediting agency, the Joint Commission.[77] As part of its DPA, the university agreed to the appointment of a chief compliance officer who would report to the president and board and to the establishment of a board audit by an independent monitor.[78] Newly appointed board members[79] observed that,

---

71. Alex, *supra* note 69.

72. *See* Margolin, *New Whistleblower*, *supra* note 65 (stating that Moody's cited "'financial and reputational risks'" related to investigations and anticipated additional litigation as reasons for downgrading the university's credit rating).

73. Kelly Heyboer & Josh Margolin, *Layoffs, Shifts on Horizon for UMDNJ Execs*, STAR-LEDGER (Newark, N.J.), Apr. 25, 2006, at 1, *available at* 2006 WLNR 6932464 (claiming that the Newark City Council President "repeatedly used his position on UMDNJ's board to get friends and family members jobs at the university . . . [and] helped get one of his campaign contributors a $1-a-year deal to lease space in a UMDNJ building"); *Medical School Trustee Resigns Amid State Ethics Probe*, RECORD (Hackensack, N.J.), July 30, 2006, at A3, *available at* 2006 WLNR 13202154 [hereinafter *Medical School Trustee Resigns*] (discussing the resignation of a trustee who allegedly pressured university officials to hire his brother, for whom the mental health counselor position qualifications were downgraded).

74. *See* Lisa Brennan, *Tough Call: U.S. Attorney Defers Prosecution in Effort to Keep Drug Company, Medical School in Business*, MIAMI DAILY BUS. REV., Apr. 12, 2006, at 10, *available at* 4/12/2006 MIAMIDBR 10 (Westlaw) [hereinafter Brennan, *Tough Call*].

75. Ted Sherman & Josh Margolin, *Citing Report of Fraud, UMDNJ Monitor Calls for New Contract Rules*, STAR-LEDGER (Newark, N.J.), July 21, 2006, at 1, *available at* 2006 WLNR 12576398.

76. *See* Sherman, *Exec Faced Retaliation*, *supra* note 70 (describing a whistleblower's allegation that "contracts were improperly incorporated into existing service agreements, avoiding review by the university's board of trustees").

77. Angela Stewart, *UMDNJ President Asks Trustees to Do More than the Minimum*, STAR-LEDGER (Newark, N.J.), May 24, 2006, at 25, *available at* 2006 WLNR 8924232.

78. Deferred Prosecution Agreement, *supra* note 48, at ¶¶ 7–13.

previously, "the board 'never asserted its authority or assumed the responsibility' to run the university."[80] Perhaps worse was a perception that the board was obstructionist in addressing the U.S. Attorney's concerns; in response to criticism that the federal monitor's mandate was overbroad, the U.S. Attorney indicated that he could not trust the board and feared that it would block reform.[81]

Reform of the UMDNJ board by the state legislature became its own morass, as legislators jockeyed over power to appoint board members and constituencies lobbied to ensure representation.[82] Finally, legislation was enacted and signed by the governor, which provided for a nineteen-member board of trustees for UMDNJ[83] and a nine-member board of directors for University Hospital.[84] Under this legislation, the board of trustees is to be appointed by the governor;[85] no board member may be an employee or paid official of any hospital affiliated with the university.[86] Members may be removed for cause by the Office of the Governor after a public hearing.[87] The University Hospital board of directors comprises nine individuals appointed by the governor, four of whom are appointed with the advice and consent of the state senate.[88] Finally, the act requires the establishment of a trustees' website, which must include "the board's rules, regulations, resolutions, and official policy statements"; a five-day notice of any board or committee meeting;

---

79. Reconstituting the board was a challenge; Acting Governor Codey's attempt to appoint two reformers was blocked by a senator who sought passage of an unrelated needle-exchange bill. Josh Margolin and Susan K. Livio, *"Senatorial Courtesy" Blocks Bid for Reform*, STAR-LEDGER (Newark, N.J.), Jan. 6, 2006, at 12, *available at* 2006 WLNR 356277. A Corzine nominee withdrew in the face of questions about a no-bid lobbying contract her firm had previously had with UMDNJ. Heyboer, *Lobbyist Withdraws*, *supra* note 50.

80. Reisinger, *supra* note 45, at 399.

81. Brennan, *Tough Call*, *supra* note 74.

82. *See generally* Opinion, *A Bad Move*, RECORD (Hackensack, N.J.), Mar. 30, 2006, at L8, *available at* 2006 WLNR 5397804 (expressing concern that proposed legislation for restructuring composition of UMDNJ board would perpetuate patronage).

83. N.J. STAT. ANN. § 18A: 64G-4(a) (West Supp. 2007).

84. *Id.* § 18A:64G-6.1(a).

85. *Id.* § 18A:64G-4(a). Two members will be recommended by the President of the Senate and the Speaker of the General Assembly, respectively, with the remainder requiring the advice and consent of the senate. *Id.* All but three members must be state residents; the board must reflect "the gender, racial and ethnic diversity of the State"; and the bill delimits the geographic regions from which the board members must be appointed. *Id.*

86. *Id.* This presumably means that neither the university president nor any dean or faculty member may sit on the board of trustees; further, it evidently requires physician representation to be a non-academic or to come from a non-affiliated hospital. However, nothing seems to preclude such representation on the University Hospital board of directors. *See id.* § 18A:64G-6.1(a).

87. *Id.* § 18A:64G-4(a).

88. *Id.* § 18A:64G-6.1(a).

minutes of each board and committee meeting; and notice of any contract entered into that was not competitively bid.[89]

It was not long before the interim president himself ran afoul of the federal monitor; Vladek was charged with violating the DPA and "intentionally misle[ading] investigators"[90] about the cardiology kickback program.[91] As a result, the board retained an external consultant to review the charges and initiated a special investigation of the university's cardiology program,[92] headed by the state health commissioner.[93] A new university president began at UMDNJ in July 2007.[94]

C. Lessons Learned

These two stories highlight only some of the problems that can occur in the nonprofit sector. Nonprofit managers operate with any number of incentives to present financial statements that show significant assets,[95] to attribute more expenses to charitable rather than administrative activities,[96] and, in the case of hospitals, to create cost reports that show maximum reimbursable expenses from federal health programs and charity care.[97] There is no reason to believe that independent boards will avoid repetition of these corporate disasters. Nevertheless, such troubles might be prevented by raising the bar on governance structures, requiring monitoring directors, obtaining regular outside audits, and imposing rigorous transparency.

---

89. *Id.* § 18A:64G-6.2.

90. Margolin, *Top Doctor on Leave*, *supra* note 60.

91. Brennan, *No-Show Cardiologists*, *supra* note 49; *see* UMDNJ INTERIM REPORT, *supra* note 56, at 2 ("UMDNJ has violated the terms of the DPA not only in failing to alert us as to the existence of the whistleblower, but also in failing to provide us with relevant documents and information concerning the serious allegations contained in this lawsuit.").

92. Brennan, *No-Show Cardiologists*, *supra* note 49.

93. Margolin, *2 Alleged No-Shows*, *supra* note 60.

94. Ana M. Alaya, *New Chief Is Upbeat at UMDNJ*, STAR-LEDGER (Newark, N.J.), July 20, 2007, at 23, *available at* 2007 WLNR 13900203.

95. Michelle H. Yetman & Robert J. Yetman, The Effects of Governance on the Financial Reporting Quality of Nonprofit Organizations 2 (Aug. 25, 2004), http://ssrn.com/abstract=590961. A variety of motivations drive the desire to show maximum assets, including pressure from regulators, a desire to expand staff, and aspirations for greater compensation. *Id.*

96. *See id.* at 3 (a more favorable charitable spending ratio presents a more efficient organization to donors and regulators); *see also* Callen et al., *supra* note 10, at 509 ("[T]hose organizations most dependent on outside donations try hardest to appear efficient.").

97. *See, e.g.*, Greaney & Boozang, *supra* note 12, at 23 (describing the difficulties encountered by a New York nonprofit hospital when it faced a decline in third-party reimbursements and inpatient admissions).

III. WHAT IS BOARD INDEPENDENCE AND WHAT IS IT TRYING TO ACCOMPLISH?

A. An Overview of the Role of Independence in Governance

The recent trend of board "independence" among policymakers and governance gurus is intended to enhance a board's monitoring functions, thereby improving the entity's overall performance.[98] Theorists claim that independence produces superior oversight of management, reduces resistance to leadership change, and more effectively aligns board behavior with shareholder interests.[99] However, while boards with independent directors may be more effective in performing certain functions (for example, replacing the CEO and making takeover bids) and may positively affect governance overall,[100] it remains unestablished how many independent directors must be on a board to achieve these positive effects.[101]

Significantly, much of the emerging empirical data remains equivocal on the value of independent boards. Researchers disagree about the long-term effects of board independence on profitability and other areas where independence was presumed to make a positive difference. Studies are split on the correlation between board independence and firm profitability; some even

---

98. *See generally* Barry D. Baysinger & Henry N. Butler, *Revolution Versus Evolution in Corporation Law: The ALI's Project and the Independent Director*, 52 GEO. WASH. L. REV. 557 (1984) (discussing the controversy arising from the ALI's recommendation in the late 1970s that public corporations have independent boards, and observing that, in response to calls for reform, major corporations began to change their governance structures during the 1970s); Irwin Borowski, *Corporate Accountability: The Role of the Independent Director*, 9 J. CORP L. 455, 455–56 (1984) (noting that director independence was a significant priority of Harold Williams, chair of the SEC from April 18, 1977 to March 1, 1981, who advocated "plac[ing] the independent director at the heart of the accountability system").

99. *See* Donald C. Langevoort, *The Human Nature of Corporate Boards: Law, Norms, and the Unintended Consequences of Independence and Accountability*, 89 GEO. L.J. 797, 801 (2001) ("[A]n independent director is one who actually takes the monitoring task for the benefit of the shareholders and/or other constituencies seriously."); Note, *Beyond "Independent" Directors: A Functional Approach to Board Independence*, 119 HARV. L. REV. 1553, 1561-62 (2006) [hereinafter *Beyond "Independent" Directors*].

100. *See* PAUL W. MACAVOY & IRA M. MILLSTEIN, THE RECURRENT CRISIS IN CORPORATE GOVERNANCE 37 (2003) ("Our position is that the empirical evidence seems to suggest . . . that reform efforts are having some impact on current governance practice. While this does not negate [Roberta] Romano's claim that such efforts have not affected firm performance, a critical evaluation of the studies upon which her opinions are based is warranted."); *id.* at 95-96 (presenting independent leadership as a critical first step in reform).

101. *See* Sanjai Bhagat & Bernard Black, *The Non-Correlation Between Board Independence and Long-Term Firm Performance*, 27 J. CORP. L. 231, 234–35 (2002) (surveying the literature on these points and concluding that studies have not focused on the question of whether boards must have a supermajority of independent directors to achieve these ends).

suggest a relationship between independence and weakened performance.[102] Experts assumed that independence would at least improve boards' performance of monitoring functions.[103] In fact, after specifically analyzing the impact of independence on the performance of audit committees, Roberta Romano has concluded that this presumption may be incorrect.[104] There is insufficient evidence that "requiring audit committees to consist solely of independent directors will reduce the probability of financial statement wrongdoing or otherwise improve corporate performance."[105]

Predictably, the studies themselves are subject to significant criticism. Many argue that their varied and negative outcomes are attributable to inconsistency in the factors used to define director independence and firm performance.[106] Further, some researchers attribute negative results to the fact that so-called "director independence" may not amount to meaningful board autonomy.[107]

---

102. *Compare* MACAVOY & MILLSTEIN, *supra* note 100, at 95–96 (claiming that independent boards improve corporate performance), *and* Ira M. Millstein & Paul W. MacAvoy, Essay, *The Active Board of Directors and Performance of the Large Publicly Traded Corporation*, 98 COLUM. L. REV. 1283, 1317 (1998) (observing that the correlation between United States corporations with active and independent boards of directors and higher economic profit supports the reasonable assumption that corporate governance matters to corporate performance), *with* Baysinger & Butler, *supra* note 98, at 572–73 (reporting that, although increased board independence corresponds with increased financial performance, "the addition of independent directors to a corporate board is subject to both diminishing marginal increases and absolute declines in relative performance"), Bhagat & Black, *supra* note 101, at 263 (reporting that firms with more independent directors do not experience increased profitability, and may actually perform worse), and Lawrence D. Brown & Marcus L. Caylor, Corporate Governance and Firm Performance 7 (Dec. 7, 2004), http://ssrn.com/abstract=586423 (stating that "solely independent audit committees are positively related to dividend yield, but not to operating performance or firm valuation").

103. *See* April Klein, *Audit Committee, Board of Director Characteristics, and Earnings Management*, 33 J. ACCT. & ECON. 375, 376 (2002) (noting that "[t]he common thread running through the SEC and stock exchange proposals" is the assumption that board independence improves earnings management).

104. Roberta Romano, *The Sarbanes-Oxley Act and the Making of Quack Corporate Governance*, 114 YALE L.J. 1521, 1533 (2005).

105. *Id.* Romano also concludes that the "vast majority" of studies have found that auditor independence is not compromised when the auditing firm also provides nonaudit services. *Id.* at 1537; *see also* Brown & Caylor, *supra* note 102, at 7 (reporting that "solely independent audit committees are positively related to dividend yield, but not to operating performance or firm valuation" and that "formal polic[ies] on auditor rotation [are] positively related to return on equity but not to any . . . other . . . performance measures").

106. *See, e.g.*, James D. Cox, *The Role of Empirical Endurance in Evaluating the Wisdom of the Sarbanes-Oxley Act*, 40 U.S.F. L. REV. 823, 836-38 (2006).

107. *See generally* Bhagat & Black, *supra* note 101, at 266–67 (positing that outside directors may not be truly independent because their social ties to the CEO, dependence on the CEO for appointment, or duration of service undermines their independence); *see also* Cox, *supra* note 106, at 836–37 (criticizing Bhagat and Black's failure to "seek more qualitative

Many nonprofits have rushed to establish independent boards despite this debate in the for-profit sector. At most, the available evidence suggests that experimentation with boards of mixed composition is warranted. Although independent directors may perform some discrete functions well, they may lack the level of knowledge about the entity and its business required for making critical decisions. For example, some commentators suggest that independent directors may serve as an important impetus to change strategy, but that inside directors may best know which alternatives to consider.[108] A few commentators suggest that some types of independent directors, such as bankers and lawyers, may be more effective than others, and that the use to which independent directors are put may also make a difference.[109]

Precisely because the research in the for-profit sector remains inconclusive, and because of the paucity of empirical work in the nonprofit sector, promoters of best practices should pause before aggressively pursuing nonprofit governance reform that rests on an independent board.[110] Nonprofits should experiment with a mix of inside and outside directors, including monitoring directors and other directors solicited for their particular expertise and stakeholder status, until it becomes clear what combination works best for each nonprofit.[111]

Even while nonprofits experiment with board composition, they will continue to face the challenge of measuring governance and the relationship between governance and outcomes. Social science literature measuring nonprofit performance has focused on effectiveness (that is, achieving

---

inputs to assess the relative independence of their 'independent' directors"). A recent Harvard commentary on these studies' results comes to a similar conclusion: "[P]utting independent directors on a board is unlikely to have much effect on financial performance if not accompanied by the implementation of structures and procedures to counteract the social and psychological constraints that paralyze many facially independent boards." *Developments in the Law—Corporations and Society*, 117 HARV. L. REV. 2169, 2200 (2004).

108. *See, e.g.*, Bhagat & Black, *supra* note 101, at 264.

109. *See, e.g.*, *id.* at 267.

110. State corporate law has not been a significant influence in achieving board independence. *See generally* Lucian A. Bebchuck & Assaf Hamdani, Essay, *Federal Corporate Law: Lessons from History*, 106 COLUM. L. REV. 1793, 1807–09 (2006) (discussing briefly the evolution of state regulation of board composition and noting that "states generally granted companies freedom in choosing the composition of their boards, and did not require them to appoint independent directors").

111. *See generally* Bhagat & Black, *supra* note 101, at 263–67 (discussing the benefits of both inside and outside directors). Langevoort suggests that the board should be comprised of "a roughly equal number" of inside and independent directors, with "so-called 'gray' directors" serving as mediators if the split between outsiders and insiders becomes contentious. Langevoort, *supra* note 99, at 814–15. Whether director term limits are important also remains an open question. Bhagat and Black speculate that on one hand, long-serving directors may become less vigilant monitors, but that on the other hand, perhaps long-term directors are more independent than newer directors. Bhagat & Black, *supra* note 101, at 266–67. They state more assuredly that "[o]lder directors, at some point, likely become less effective." *Id.* at 267.

articulated goals)[112] and efficiency, defined as "the ratio of administrative expenses to total expenses, the ratio of fundraising expenses to total expenses, and the ratio of program expenses to total expenses."[113] However, few studies examine the relationship between these performance metrics and governance.

These evaluations are not a science, as assessments of board effectiveness vary depending upon the nonprofit constituency surveyed.[114] For example, the stakeholders in colleges differ on the appropriate metric for measuring their success. While students rely on a number of ranking systems, such as *U.S. News & World Report*,[115] universities prefer to consider the percentage of faculty with terminal degrees, number of books published, endowment size, or number of NIH grants or Nobel prizes awarded to faculty.[116] Furthermore, some colleges might stress the rate of graduation within five years, diversity of student body and faculty, availability of need-based scholarships, ROTC enrollment, or post-graduation employment statistics.[117]

Another example would be an urban hospital that runs a perpetual deficit. The deficit could indicate inefficient management or overpayments to insiders; however, it could also indicate the provision of a tremendous amount of unreimbursed charity care, which would at least be mission-consistent.[118] As a third example, an environmental organization's lack of success in protecting rain forests may reflect any number of problems, including insufficient funds,

---

112. Callen et al., *supra* note 10, at 498–99.

113. *Id.* at 495.

114. *Id.* at 499 (observing that "each group measures effectiveness on the basis of criteria and impressions most relevant to it").

115. *See, e.g.*, America's Best Colleges 2008: National Universities, http://www.usnews.com/sections/rankings (follow "National Universities" hyperlink) (last visited Nov. 25, 2007) (ranking national universities on such factors as academic quality, economic and ethnic diversity, student debt and amount of need-based aid awarded, acceptance rate, transfer rate, and average age of the student body).

116. A significant board responsibility in endowed contexts is investment of charitable assets, FREMONT-SMITH, *supra* note 16, at 431, the success of which is measurable. Each year, the Chronicle of Higher Education publishes a chart comparing universities' investment performance. *See, e.g.*, Maria Di Mento, *How Endowments of 247 Major Nonprofit Organizations Performed*, CHRON. HIGHER EDUC. (Wash., D.C.), June 2, 2006, at B3, *available at* http://chronicle.com/weekly/v52/i39/39b00301.htm.

117. Only recently have the federal government and accreditors pushed universities to rate performance by means of student graduation rates; this approach remains controversial. *See* Paul Bradley, *Measuring Performance*, COMMUNITY C. WKLY., June 18, 2007, at 6; Arthur Levine, *Higher Education's New Status as a Mature Industry*, CHRON. HIGHER EDUC. (Wash., D.C.), Jan. 31, 1997, at A48, *available at* http://chronicle.com/che-data/articles.dir/art-43.dir/issue-21.dir/21a04801.htm.

118. Partly out of concern about retaining federal tax-exempt status and as an attempt to aid entities in pursuing their mission, the Catholic Health Association has developed a metric for the definition, measurement, and reporting of charity care. *See* Cath. Health Ass'n U.S., Executive Summary, http://www.chausa.org/Pub/MainNav/ourcommitments/CommunityBenefits/Resources/TheGuide/ (last visited Nov. 19, 2007).

poor leadership and strategy, or simply the impossibility of overcoming the political power and strength of those opposing its goals.

The current justifications for board independence and the varied definitions of director independence are insufficient. Consequently, nonprofit entities should experiment with board makeup, including the appointment of monitoring directors, who should comprise whatever proportion of the board is deemed appropriate by the particular organization. This conception could change as more data become available to guide the development of more specific best practices. Definitions of independence have become so narrow that they frequently are irrelevant to certain types of nonprofits. This suggests the need for alternative terminology that more effectively conveys the goals being sought. Nevertheless, because extant notions of independence provide the backdrop to the entire conversation about corporate governance, they are the focus of the following discussion.

Even in the business sector, the concept of director independence does not have a single meaning; a commentator helpfully categorizes the definitions into "three broad categories: the independent director as 'disinterested outsider,' the independent director as 'objective monitor,' and the independent director as 'unaffiliated professional.'"[119] The "disinterested outsider" category captures definitions that "focus on the director's lack of financial ties to the corporation."[120] This makes the category underinclusive—particularly in the nonprofit setting—because it fails to address directors' "soft ties to management,"[121] such as those that emerged in the Grasso case. This focus on compensated relationships also ignores the unique noneconomic characteristics of many nonprofits; for example, because UMDNJ is a state entity, directors' political connections were as important as any financial ties. On the other hand, a board wholly made up of "objective monitors" seems counterintuitive in a nonprofit context. Such a board would appoint directors irrespective of allegiance to the organization's mission or religious inspiration, thus prioritizing compliance over pursuit of mission. This would undermine not

119. *Beyond "Independent" Directors*, *supra* note 99, at 1555; *see also* Ronald J. Gilson & Reinier Kraakman, *Reinventing the Outside Director: An Agenda for Institutional Investors*, 43 STAN. L. REV. 863, 873 (1991) (criticizing typical outside directors as "more independent of shareholders than they are of management").

120. *Beyond "Independent" Directors*, *supra* note 99, at 1555.

121. *Id.* at 1555–57. Fitch's definition might similarly appear underinclusive—"A director is defined as independent when his or her seat on the board is the sole connection to the company"—but its discussion further acknowledges the complexities of the personal and professional relationships that frequently exist between key executives and directors. FITCH RATINGS, EVALUATING CORPORATE GOVERNANCE: THE BONDHOLDERS' PERSPECTIVE 6 (2004) [hereinafter FITCH RATINGS, EVALUATING CORPORATE GOVERNANCE], *available at* http://www.fitchratings.com/corporate/reports/report_frame.cfm?rpt_id=203150. The focus, Fitch urges, should be on "the spirit of independence." *Id.* at 9.

only the particular entity's purpose for being, but also nonprofits' societal role of providing opportunities for civic involvement.[122]

The final, "unaffiliated professional" model of independent director resembles a typical independent audit committee; it involves expert directors with access to their own staff and counsel.[123] The subjective aspiration of this model is "to bring a high degree of rigor and skeptical objectivity to the evaluation of company management and its plans and proposals."[124] Again, this model is unsuited for the nonprofit sector. Its conceptual vision is antithetical to the nonprofit purpose of bringing together citizens desiring a forum for community activism or special involvement in a religious organization or affinity group. Given this overview, the following discussion surveys current definitions of director independence.

B. Defining Director Independence in the Business Corporation

Long before Senator Sarbanes and Congressman Oxley[125] or the exchanges themselves entered the fray of corporate governance, Institutional Shareholder Services (ISS)[126] and the Council of Institutional Investors (CII) preferred companies with independent boards, prompting them to publish criteria for use by their respective constituencies.[127] In short, many of the changes induced by the Sarbanes-Oxley Act were not novel; in fact, by 2002, several of these structural changes to corporate governance had already been proven not particularly effective in achieving the goals sought.[128] Nonetheless, the

---

122. *See generally* Reiser, *Dismembering Civil Society*, *supra* note 18 (discussing the relationship between board accountability and "the nonprofit sector's role in constructing and maintaining civil society").

123. *Beyond "Independent" Directors*, *supra* note 99, at 1559.

124. Langevoort, *supra* note 99, at 798.

125. Senator Paul Sarbanes and Representative Michael Oxley sponsored major corporate governance reform legislation that was signed into law by President Bush on July 30, 2002. Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, 116 Stat. 745 (2002) (codified as amended in scattered sections of 11, 15, 18, 28, & 29 U.S.C.).

126. ISS produces an annual corporate governance rating of 7500 U.S. and global corporations, called a Corporate Governance Quotient. National Association of Corporate Directors, Frequently Asked Questions in Corporate Governance, http://www.nacdonline.org/FAQ/details.asp?faq=1 (last visited Nov. 19, 2007). While originally created for investors, the ratings are now relied upon by investment banks, insurers, and hedge funds, as well as industry peers engaged in benchmarking. *Id.*

127. Even now, they remain stricter than the Securities and Exchange Commission rules or the listing standards of the NYSE and NASDAQ. One obvious practical problem created by these various standards is their divergence. *See generally* David B. H. Martin, *Considering Director Independence*, SM015 A.L.I.-A.B.A. 241 (2006) (Westlaw) (comparing the various standards in detail).

128. *See* Clark, *supra* note 7, at 255-56; Romano, *supra* note 104, at 1526. *See generally* James S. Linck, Jeffrey M. Netter & Tina Yang, The Effects and Unintended Consequences of the Sarbanes-Oxley Act, and Its Era, on the Supply and Demand for Directors (Feb. 14, 2007),

Sarbanes-Oxley Act sought "1) to strengthen the independence of auditing firms, 2) to improve the quality and transparency of financial statements and corporate disclosure, 3) to enhance corporate governance, 4) to improve the objectivity of research, and 5) to strengthen the enforcement of the federal securities laws."[129] Most relevant to this discussion is that Sarbanes-Oxley seeks to strengthen the monitoring aspect of governance by assigning specific functions to independent directors and heightening expectations of diligence.[130] Audit committees,[131] which must be composed of independent directors, bear the greatest burden, as they are responsible for ensuring the integrity of companies' internal controls and the independence of the auditing process.[132]

Sarbanes-Oxley nudged the exchanges to institute significant governance changes for listed entities. Again, much focus was placed upon improving governance through the assignment and execution of certain responsibilities by independent directors and requiring that board's independent directors meet regularly outside the presence of management.[133] The commentary to the NYSE Corporate Governance Standards states that a board of independent directors "will increase the quality of board oversight and lessen the possibility of damaging conflicts of interest."[134] While there are differences among the NYSE, American Stock Exchange, and NASDAQ, all three entities generally require that member companies establish both a compensation committee and an independent audit committee; both the compensation and director nomination processes must involve a majority of independent directors.[135]

---

http://ssrn.com/abstract=902665 (examining the repercussions of Sarbanes-Oxley).

129. Linck et al., *supra* note 128, at 4.

130. *See* Clark, *supra* note 7, at 267.

131. Public companies have been required since 1999 by the various exchanges' listing standards to have audit committees. Cox, *supra* note 106, at 829.

132. The audit committee is responsible for the selection, oversight and compensation of the company's independent auditors, and has the power and financial authority to retain counsel and other appropriate advisors specific to the committee. 15 U.S.C. § 78j-1 (2002); *see* Standards Relating to Listed Company Audit Committees, 68 Fed. Reg. 18,788 (Apr. 16, 2003) (codified at 17 C.F.R. pts. 228, 229, 240, 249, and 274). This is designed to ensure independence of the auditors from the client corporation's management. *See* Cox, *supra* note 106, at 830–31 ("The audit committee is now required to review with the auditor the critical accounting judgments, estimates, and choices made by management."). The listing rules of both the NYSE and NASDAQ respond to inappropriate earnings management and distorted reporting by requiring financial expertise of audit committee members. Klein, *supra* note 103, at 375–76.

133. NYSE, LISTED COMPANY MANUAL § 303A.03 (Supp. 2003 & 2004). The Governance Standards require that a majority of a listed company's directors must be independent. *Id.* § 303A.01(a). A listed company must disclose in its annual proxy statement the names of directors it has identified as independent. *Id.* § 303A.02 cmt. The Council of Institutional Investors (CII) and Institutional Shareholder Services (ISS) criteria also address the percentage of board and committee meetings attended by directors, and the number of other boards on which directors sit. Martin, *supra* note 127, at 250–51.

134. NYSE, *supra* note 133, § 303A.01 cmt..

135. AM. STOCK EXCHANGE, AMEX COMPANY GUIDE §§ 803-805 (2007), *available at*

The NASDAQ standards narrowly define who qualifies as an independent director; the definition excludes as "insiders" not only officers, employees, and family members who do business with the listed company or a subsidiary, but also "any other individual having a relationship which, in the opinion of the [company]'s board of directors, would interfere with the exercise of independent judgement in carrying out the responsibilities of a director."[136] The standards further identify specific characteristics precluding independence: (1) employment by the company, its parent, or its subsidiary, at present or within the past three years;[137] (2) receipt of payments from the company under certain enumerated circumstances;[138] (3) a family relationship with one who is or was in the prior three years an executive officer of the company, its parent, or its subsidiary;[139] (4) affiliation in certain enumerated ways with an organization to which the company has made significant payments or donations;[140] and (5) a relationship with the company's outside auditor.[141]

The most common critique of current definitions of independence involves their failure to address situations in which a director's close personal relationship with a CEO or other senior manager might interfere with the ability to act independently.[142] Problematically, many people accept directorships for reasons wholly unrelated to serving as a good director or corporate monitor;

---

http://wallstreet.cch.com/AMEX/CompanyGuide (follow hyperlink for Part 8: Corporate Governance); NASDAQ, MANUAL § 4350(c)-(d) (n.d.); NYSE, *supra* note 133, §§ 303A.05–.07; *see* Clark, *supra* note 7, at 268–70.

136. NASDAQ, *supra* note 135, § 4200(a)(15); *see id.* § IM-4200. Under the NYSE listing standards, the board must "determine[] that the director has no material relationship with the listed company (either directly or as a partner, shareholder, or officer of an organization that has a relationship with the company)." NYSE, *supra* note 133, § 303A.02. The Business Roundtable's approach is similarly expansive; it excludes from qualifying as independent those directors who have "relationships with the corporation or its management [or other directors]—whether business, employment, charitable, or personal—that may impair, or appear to impair, . . . independent judgment." BUSINESS ROUNDTABLE, PRINCIPLES OF CORPORATE GOVERNANCE 14 (2005), *available at* http://www.businessroundtable.org/pdf/CorporateGovPrinciples.pdf.

137. NASDAQ, *supra* note 135, § 4200(a)(15); *see* Gilson & Kraakman, *supra* note 119, at 873 (noting the presumption that individuals who do not have a "personal financial stake in retaining management . . . act as shareholder surrogates," who will run company "in the long-term best interests of its owners").

138. NASDAQ, *supra* note 135, § 4200(a)(15).

139. *Id.*; *see* Borowski, *supra* note 98, at 460 (noting that "former SEC chairman Harold "Williams defined an independent board member as one who has no familial or business relationship to the corporation or its management").

140. NASDAQ, *supra* note 135, § 4200(a)(15).

141. *Id.* The Rules require the board to make an affirmative determination that independent directors do not have the kind of disqualifying relationships identified in Rule 4200. *Id.* § 4350(c)(1). Directors serving on the audit committee are subject to even more stringent qualifications. *Id.* § 4350(d)(2).

142. *See Developments in the Law*, *supra* note 107, at 2198–99 (noting that the impact of social relationships on director independence merits close scrutiny (citing as an example *In re* Oracle Corp. Derivative Litigation, 824 A.2d 917 (Del. Ch. 2003))).

they may seek the prestige, stipend, or connections that accompany the position. Moreover, even directors satisfying the definition of independence may be subject to influence—especially if management controls their appointment and continuing service.[143] Directors who serve for long periods may also lose objectivity as they become normed to the ethos of the board and corporation or develop social and professional connections with other board members and the CEO.[144] Studies suggest that directors' outside affiliations may also affect their behavior; for example, board members who are themselves officers of major corporations tend to behave as they would have their ideal director behave, and therefore are unlikely to be activist directors.[145] None of these issues are effectively addressed by the various for-profit conceptions of director independence. More importantly, these definitions fail to articulate the ideal characteristics that should inhere in directors of mission-driven nonprofit entities.

C. Extrapolating Independence to the Nonprofit Sector

As a foundation, it is helpful to understand who currently populates nonprofit boards.[146] Nonprofit boards tend to be larger than for-profit boards,

---

143. *See* Gilson & Kraakman, *supra* note 119, at 884 ("Although outside directors are financially independent, they are traditionally selected by management and remain socially and ideologically tied to management . . . ."); *cf.* Bainbridge, *supra* note 6, at 9–10 (noting that, "as board power increases relative to the CEO . . . newly appointed directors become more demographically similar to the board").

144. *Cf.* Bhagat & Black, *supra* note 101, at 266 (speculating that lengthy tenures could reduce monitors' vigilance); *Developments in the Law*, *supra* note 107, at 2198-2200 (describing the effects of "financial, social, and psychological constraints" on supposedly independent directors).

145. *See* Gilson & Kraakman, *supra* note 119, at 884.

146. An overview of the nonprofit sector is useful here. "The 1.4 million charitable, religious, scientific, educational, and cultural organizations that comprise what is generally considered the 'nonprofit' sector hold assets worth more than $2 trillion and receive annually an estimated $241 billion in support from individuals, corporations, and foundations." FREMONT-SMITH, *supra* note 16, at 1. This accounting does not include churches and church-related organizations, which are not required to seek federal tax exemption. *Id.* at 9. In 1998, the IRS estimated that there are about 354,000 churches and other religious organizations. *Id.* Nonprofit volunteers "provide the equivalent of 9 million full-time staff members"; an additional 11.7 million paid employees work in the charitable sector. PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 10. However, most nonprofits are small: "only 4 percent of all charitable organizations have annual budgets of more than $10 million." *Id.* at 11. "Health care organizations account for the largest amount of assets and receive over half of the revenue of the sector, while expending 54.3% of wages and salaries." FREMONT-SMITH, *supra* note 16, at 9. Thirty thousand nonprofit organizations educate about twenty percent of all college and ten percent of all elementary and high school students. Universities hold the largest endowments in the country. *Id.* at 10. The nonprofit sector comprises almost six percent of all entities legally established in the United States. *Id.* at 7. The majority of financial support for the nonprofit sector comes from fees paid by consumers, donations, and government grants. PANEL ON THE

primarily because of the desire to attract potential donors.[147] While most best practices discourage large boards, suggesting that they lead to less engagement and oversight,[148] a recent national study of over five thousand nonprofits found no negative relationship between board size and engagement.[149] Rather, the study found a positive relationship between size and fundraising, public education about the entity's mission, and involvement in public policy initiatives.[150] Another study suggests that "the larger the board, the less efficient the organization is with respect to fundraising expenses," unless these organizations include a "larger proportion of major donors on the board," in which case their fundraising efforts are significantly more efficient.[151]

Despite nonprofit boards' tendency to be larger than their for-profit counterparts, the overwhelming majority of nonprofits report difficulty in recruiting board members.[152] Compensation of nonprofit directors is almost nonexistent; at most, 10% of the largest entities provide compensation for board service.[153] Whether large or small, nonprofit boards are not racially or ethnically diverse, which suggests that diversity is not a priority;[154] this insularity seems inconsistent with tax exempt entities' legal obligation to ensure that they provide community benefit. While most boards are relatively gender-balanced, the proportion of women serving on boards declines dramatically as size increases; thus, in nonprofits with over $40 million in expenses, women constitute only 29% of directors.[155] Yet a study by Francie Ostrower of the

---

NONPROFIT SECTOR, *supra* note 4, at 12.

147. *See* FRANCIE OSTROWER, URB. INST. CTR. ON NONPROFITS & PHILANTHROPY, NONPROFIT GOVERNANCE IN THE UNITED STATES 17 (2007), *available at* http://www.urban.org/UploadedPDF/ 411479_Nonprofit_Governance.pdf.

148. *See e.g.*, IRS, Good Governance Practices, *supra* note 5, at 1.

149. OSTROWER, *supra* note 147, at 17.

150. *Id.*

151. Callen et al., *supra* note 10, at 510.

152. OSTROWER, *supra* note 147, at 16.

153. *Id.* at 11. Nothing suggests a relation between compensation and engagement overall; however, compensation has been found to be "negatively associated with levels of board activity in fundraising, community relations, and educating the public about the organization and its mission." *Id.*

154. *Id.* at 18. According to Ostrower,

> On average, 86 percent of board members are white, non-Hispanic; 7 percent are African-American or black; and 3.5 percent are Hispanic/Latino .... Medians convey even greater homogeneity—96 percent for white members and zero for African-Americans and Hispanics. Fifty-one percent of nonprofit boards are composed *solely* of white, non-Hispanic members.

Id. A substantial percent of board members of organizations that serve significant numbers of minority clients are wholly white. *Id.*

155. *Id.* at 19 (noting also that "emphasis on financial skills and reputation in the community as recruitment criteria were negatively associated with the percentage of women, as was being in the health field, reliance on endowment funding, and location in a metropolitan statistical area").

Urban Institute suggests that "[g]ender diversity [is] positively associated with activity in fundraising, planning, community relations, and educating the public about the organization."[156]

A fair proportion of boards, about 19–26%, include people who are related to each other.[157] Many nonprofit boards include directors who also serve on the boards of business corporations (ranging from 31% of directors among the smallest nonprofits to 80% among the largest);[158] this figure is positively related to the percentage of nonprofit entities implementing Sarbanes-Oxley best practices.[159] Because the independent board is being encouraged aggressively in the nonprofit sector through the requirements of ratings agencies[160] and grantors (which have significant influence over nonprofits' practices),[161] increasingly fewer entity employees serve on nonprofit boards.[162] According to Ostrower, "only a minority of boards" actively participate in many of the activities examined in her study, "including fundraising (29 percent), monitoring the organization[']s programs and services (32 percent), monitoring the board's own performance (17 percent), planning for the future (44 percent), community relations (27 percent), and educating the public about the organization and its mission (23 percent)."[163]

With this background, we can consider the meaning and desirability of nonprofit board independence. Extant definitions of nonprofit director independence are rather anemic, with significant focus on the absence of a financial relationship between the director and the entity. For example, the Revised Model Nonprofit Corporation Act provides an optional section that

---

156. *Id.* at 16.

157. *Id.* at 20. Boards that emphasize friendship or acquaintance with a current director in their recruiting of new directors score negatively on almost all measures of board performance. *Id.* at 16.

158. *Id.* at 20.

159. *Id.* at 4.

160. FITCH RATINGS, SARBANES-OXLEY ACT: VOLUNTARY COMPLIANCE VIEWED AS A BEST MANAGEMENT PRACTICE 1 (2004) (stating that adoption of key SOX provisions by colleges and universities will be looked upon favorably as a rating consideration and citing management as a factor in universities' financial success and board oversight). Fitch advises that the corporate governance criteria considered by bondholders include "[b]oard independence and quality," "[t]he presence of related-party transactions," "integrity of the audit process," "[e]xecutive compensation relative to company performance," and "[d]iffering ownership structures." FITCH RATINGS, EVALUATING CORPORATE GOVERNANCE, *supra* note 121, at 1. Fitch also states the belief that "corporate governance can have a material impact on . . . credit quality" and acknowledges that weak governance can impair financial position. *Id.* at 2.

161. *See* discussion *infra* note 197.

162. Callen et al., *supra* note 10, at 501. In a survey of over 5100 nonprofits, the Urban Institute found that boards with paid CEOs were more professional, but that the presence of a voting CEO on a board "was *negatively* associated with having an outside audit, a conflict of interest policy, a document retention policy, and a whistleblower policy." OSTROWER, *supra* note 147, at 4–5.

163. *Id.* at 12.

would prohibit more than 49% of a public benefit corporation's board from being "financially interested persons."[164] In response to a Senate request for a study and report on a variety of nonprofit governance issues, Independent Sector has taken the position that

> At least one-third of the . . . governing board should be independent: that is, individuals who have not received compensation or material benefits directly or indirectly from the organization in the previous 12 months, whose compensation is not determined by other board or staff members, and who is not related to someone who received such compensation from the organization.[165]

The Senate Finance Committee internal recommendations are more realistic about accomplishing the intended goal of monitoring management.[166] They define an independent member "as free of any relationship with the corporation or its management that may impair or appear to impair the director's ability to make independent judgments."[167] The recommendations

---

164. REV. MODEL NONPROFIT CORP. ACT § 8.13(a) (1987); *accord* CAL. CORP. CODE § 5227 (West 1990); ME. REV. STAT. ANN. tit. 13-B, § 713-A (1964); *see* N.H. REV. STAT. ANN. § 7:19(II) (2003). According to the drafters, the provision is optional because "[l]egitimate public benefit corporations might have difficulty in finding active and competent directors who ha[ve] no financial interest in the corporation." REV. MODEL NONPROFIT CORP. ACT § 8.13 cmt. The Revised Model Act defines "financially interested persons" as:

> (1) Individuals who have received or are entitled to receive compensation, directly or indirectly, from the corporation for services rendered to it within the previous 12 months, whether as full- or part-time employees, independent contractors, consultants or otherwise, excluding any reasonable payments made to directors for serving as directors; or
>
> (2) Any spouse, brother, sister, parent or child of any such individual.

Id. § 8.13(b). This differs from director conflict of interest transactions, where the nonprofit enters into a transaction with an entity in which one of the directors of the nonprofit has a direct or indirect interest. *See id.* § 8.31; *see also* ALI DRAFT PRINCIPLES, *supra* note 3, § 310 cmt. (c)(4) (distinguishing between positional independence and transactional conflicts).

165. PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 7.

166. The Senate Finance Committee has focused on a wide range of corporate behavior, including abuse of the charitable corporate form and attendant tax-exemption by donor-advised funds, supporting organizations, credit counseling organizations, and parties to tax shelter transactions; heightening penalties for abusive behavior by tax-exempt entities, such as excessive compensation; establishing review standards for nonprofit conversions; expanding tax-exempts' 990 reporting requirements; allowing states to prosecute violations of federal tax law; enhancing disclosure requirements of related organization and insider transactions; increasing transparency of information to the public; and more explicitly articulating the duties of nonprofit directors and bases for removal. *See* U.S. SENATE COMM. ON FIN., TAX EXEMPT GOVERNANCE PROPOSALS 1–14 (Staff Discussion Draft 2006), http://www.senate.gov/~finance/hearings/testimony/2004test/062204stfdis.pdf.

167. *Id.* at 13.

would also preclude any compensated member from serving as chair or treasurer of the board.[168]

The definition offered in the draft of the ALI Principles of the Law of Nonprofit Organizations rejects the business model of independence as unsuited to most nonprofits; rather, the Draft Principles recommend a board composed of individuals with sufficient objectivity to assure management oversight.[169] The ALI draft addresses the concept of director independence in the comments accompanying its Duty of Loyalty provision.[170] Specifically, the comments reject what is referred to as "external board-member independence," noting that the tradition of appointing founders and major donors to a nonprofit board precludes this form of independence; to elect otherwise, the comments suggest, may result in failure of accountability to key constituencies.[171] The comments instead encourage "internal independence," which can be achieved by appointing a majority of directors who are non-management and uncompensated, who can thereby ensure management oversight.[172] The comments on committee membership additionally recognize that state laws may require certain committees, such as audit committees, to be composed exclusively of independent members.[173]

Although the ALI approach is more suitable to the nonprofit sector than any other attempt has been, it may go both too far and not far enough. First, as discussed previously, the rationale for a majority of independent directors is not empirically supported; further experimentation and study are required. Second, the proposed Principles fail to address either director expertise[174] or the need to allow nonprofits to create boards in the context of the particular nonprofit's size and available resources.

Third, consistent with the move away from including paid employees on nonprofit boards, the Draft Principles anticipate the inclusion of staff members on board committees, rather than the board itself.[175] The Principles' guidance on this point is confusing, though. They do not explain whether such non-director appointments would have to be required by the entity's bylaws, whether the appointed individual would have a vote at the committee level,

---

168. *Id.*

169. ALI DRAFT PRINCIPLES, *supra* note 3, § 310 cmt. c.

170. *Id.* The "[c]ommentary explores the concept of independence required of charity board members, distinguishing positional independence from accountability; discussing separation of oversight and management; and highlighting transactional conflicts of interest as an area for attention." *Id.* at reporter's memorandum.

171. *Id.* cmt. (c)(2).

172. *Id.* cmt. (c)(3).

173. *Id.* § 325 cmt. (b)(5).

174. *See generally* PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 8 (stating that nonprofits should adhere to recommended practices and state law mandating that boards include directors with some financial expertise).

175. *See* ALI DRAFT PRINCIPLES, *supra* note 3, §§ 310 cmt. (c)(3), 325 cmt. (b)(5); *see also* Callen et al., *supra* note 10, at 513 (noting the trend of higher staff representation on committees than on boards, with the exception of audit committees).

whether the special appointee could participate and vote at a board meeting concerning the issue for which the person was appointed, or whether the legal duties of a director would be imposed upon the special appointee for purposes of the committee's work.

Finally, the Draft Principles perpetuate the notion that nonprofit entities may condition board appointment on minimum financial contributions.[176] Empirical studies support the presumption that donors and corporate leaders benefit nonprofit boards more than just financially; research suggests that nonprofit boards that include major donors tend to be more efficient—that is, administrative expenses are smaller in proportion to program expenses.[177] It is unclear whether this is because donor board members demand efficiency or simply because donors are drawn to more efficient organizations.[178] Donor board participation is not an unmitigated positive, however. Overreliance on a particular director's generosity can result in that person exercising disproportionate suasion, which can be detrimental to the organization and its beneficiaries.[179]

Further, at least for some kinds of nonprofits, boards should not be wholly made up of business leaders and the social elite. For example, financial knowledge is not necessarily concomitant with the skills suited to educating the community about the entity's mission or engaging in public policy debates. For this reason, diversity of perspective on nonprofit boards should extend beyond the notion of the independent versus the insider director. Social service agencies and health care providers—entities that serve a community larger than that reflected at board meetings—face particular challenges in this regard. Nonprofits such as universities and hospitals might also benefit from the inclusion of stakeholders not always represented in the boardroom, such as faculty members and physicians, who have the expertise and institutional commitment to contribute to mission formation and development of a vision for the future.

In sum, the Principles reflect an ethos that likely contributes to the continued homogeneity of boards. Although the Principles reflect both the

176. ALI DRAFT PRINCIPLES, *supra* note 3, § 320 cmt. (g)(2).

177. Callen et al., *supra* note 10, at 515.

178. *Id.*

179. No story better epitomizes this point than the extraordinary involvement of Thomas Monaghan (the philanthropist who made millions as the founder of Domino's Pizza) in the much acclaimed Ave Maria School of Law in Ann Arbor, Michigan, which he founded as a conservative institution devoted to the Catholic faith. Susan Hansen, *Our Lady of Discord: A Schism over Giving Away a Pizza Fortune*, N.Y. TIMES, July 30, 3006, § 3, at 1. With the law school in its nascent yet successful early years, Monaghan decided to establish Ave Maria University in a rather remote area of Florida and precipitously determined that the Michigan law school would move to the Florida campus as well. *Id.* Describing the fight over whether to move the law school to the new site of Ave Maria University in Florida, the president of the law school's alumni association "criticized Mr. Monaghan's insistence on operating the school like a private business and . . . the board's failure to stand up to him." *Id.*

concern that independence is often too narrowly defined for nonprofits and the need to ensure nonprofits' ability to appoint directors from key constituencies, these aspirations are narrowly imagined.

As discussed earlier, boards generally do not spend a significant amount of time fundraising,[180] which raises questions about the legitimacy of donation capacity as a criterion for board membership. Nonetheless, funder grooming is likely to remain an important goal for nonprofits, and inclusion of at least some donors as directors clearly benefits the entity. Nevertheless, board membership should not be the only reward for an actual or potential financial contribution. Engagement of current and potential donors might be just as effectively accomplished by making annual giving a criterion for membership on an advisory board, the creation of which is wholly embraced by the ALI draft.[181] Advisory boards enable nonprofits to confer a prestigious role upon donors who may not be available for or suited to the full breadth of board responsibilities. They also provide an ideal mechanism to groom and vet prospective board members pursuant to other criteria that indicate qualified directors.

Having critiqued the current governance reform movement, the next tasks undertaken in this Article are to understand how boards work and to propose a different vision. To some, this vision will prove unsatisfactory because it urges experimentation, creativity, and significantly more study of the nonprofit sector before committing to any particular governance model. Simply too few data exist upon which to make confident public policy recommendations. That being said, the prescription for moving forward does include the appointment of monitoring directors, closing of statutory governance loopholes, and a much more rigorous approach to transparency.

IV. STICKING WITH THE BASICS: ADDRESSING THE STATUTORY FLAWS OF GOVERNANCE STRUCTURE

Structuring a nonprofit board that will fulfill the myriad expectations imposed upon it in today's environment is part art and part science; the precise formula for success remains undiscovered, as confirmed by empirical data. Yet there are some obvious flaws in states' statutory schemes regulating governance structure, the correction of which might strengthen governance more than the current obsession with board independence. This section begins with a brief description of the theory behind the corporate board's purpose, upon which statutory frameworks are based, and suggests that the theory is removed from the true environment in which many nonprofits operate. The discussion then identifies specific deficiencies that, if remedied, might more directly resolve some of the structural governance flaws inherent in the current statutory framework. Finally, the section concludes with a specific focus on a best

180. *See supra* text accompanying note 163.

181. ALI DRAFT PRINCIPLES, *supra* note 3, § 320 cmt. (b)(3).

practice of monitoring directors and a proposal for a much more rigorous statutory vision of transparency in the nonprofit sector.

A. The Board's Purpose

The Revised Model Nonprofit Corporation Act and most states' nonprofit statutes derive from business corporation statutes, with an aim to easing administration by making the schemes as similar as possible.[182] The problem, of course, is that for-profits and nonprofits are dramatically different animals and should be governed by statutory schemes that reflect their differences.

The for-profit board theoretically operates on behalf of the firm's shareholders as the ultimate legal authority and holder of responsibility for the entity's activities.[183] The board exists, so the explanation goes, because the true owners of the firm, shareholders, cannot perform these functions themselves. To fill the void, the owner-shareholders delegate decisionmaking authority to "the firm's board of directors, who in turn delegate most operational decisions to the firm's senior officers."[184] Because the board retains ultimate responsibility for the firm's performance, however, it fulfills two basic roles: "management and monitoring."[185] These functions include hiring and

---

182. The drafters of the Revised Model Nonprofit Corporation Act relied upon the Revised Model Business Corporation Act for the ministerial portions of the statute, but adopted the California Nonprofit Corporation Act's classification method. Lizabeth A. Moody, *The Who, What, and How of the Revised Model Nonprofit Corporation Act*, 16 N. KY. L. REV. 251, 265–66 (1989). Moody also notes that many states' nonprofit acts simply substitute the word "member" for "shareholder" in the corresponding business organization act. *Id.* at 270; *see also* Harry G. Henn & Jeffrey H. Boyd, *Statutory Trends in the Law of Nonprofit Organizations: California, Here We Come!*, 66 CORNELL L. REV. 1103, 1104 (1981) ("The law historically has given nonprofit organizations, like Cinderellas, the hand-me-downs of their half-siblings, the business organizations.").

183. The Business Roundtable, a source of corporate best practices in both the business and nonprofit sectors, defines board oversight responsibilities to include: (1) "[p]lanning for management development and succession," (2) "[u]nderstanding, reviewing and monitoring the implementation of the corporation's strategic plans," (3) "[u]nderstanding and approving annual operating plans and budgets," (4) "[f]ocusing on the integrity and clarity of the entity's financial statements and financial reporting," (5) "[a]dvising management on significant issues facing the corporation," (6) "[r]eviewing and approving significant corporate actions," (7) "[r]eviewing management's plans for business resiliency," (8) "[n]ominating directors and committee members and overseeing effective corporate governance," and (9) "[o]verseeing legal and ethical compliance." BUSINESS ROUNDTABLE, *supra* note 136, at 8-10.

184. HENRY HANSMANN, THE OWNERSHIP OF ENTERPRISE 35 (1996). This delegation creates "agency costs," which include the board's costs of obtaining the information necessary to understand the firm's operations, perform its functions well, and implement corrective changes where necessary. *Id.* Additionally, where board oversight proves insufficient and management engages in detrimental opportunistic behavior, the results of such behavior are also considered agency costs. *Id.*

185. Clark, *supra* note 7, at 278.

determining the compensation for the CEO, assessing senior management's performance, approving significant transactions or business decisions, and guiding the firm's strategic direction.[186]

By contrast, the nonprofit corporation does not have "owners," as that term is commonly understood;[187] theorists claim that this results in significantly less monitoring.[188] Consequently, the board of directors is charged with overcoming the disconnect among donors, management, and the intended beneficiaries of the nonprofit's activities.[189] However, the assumption that nonprofits operate without owners or significantly involved stakeholders grossly oversimplifies the complex environment in which nonprofits operate.

On the one hand, the agencies primarily vested with oversight of charitable entities—attorney general offices on the state level and the IRS at the federal level[190]—do a rather poor job.[191] The fact remains, though, that attorneys

---

186. Susanna M. Kim, *Dual Identities and Dueling Obligations: Preserving Independence in Corporate Representation*, 68 TENN. L. REV. 179, 209–13 (2001). Observers disagree on the board's proper role with respect to strategic planning: "Some envision a limited strategic role for the board, with directors merely providing advice to, or acting as a 'sounding board' for, senior management.... Others assert that the board should play a more active, substantial role, deliberating and deciding major strategic issues as a collective body." *Developments in the Law*, *supra* note 107, at 2184–85.

187. HANSMANN, *supra* note 184, at 17. According to Hansmann, "The defining characteristic of a nonprofit organization is that the persons who control the organization—including its members, directors, and officers—are forbidden from receiving the organization's net earnings." *Id.* This is commonly referred to as the "nondistribution constraint." *See id.* at 18.

188. *See, e.g.*, Geoffrey A. Manne, *Agency Costs and the Oversight of Charitable Organizations*, 1999 WIS. L. REV. 227, 227-28. Scholars attribute the low degree of monitoring to the even higher agency costs in the nonprofit context, the lack of owners with a financial investment in oversight, and the fact that oversight by donors is "prohibitively expensive." *Id.* at 232, 234.

189. *See id.* at 231–32 (noting the "customary separation between a nonprofit's patrons and its beneficiaries" and the consequent role of the board and officers). This disconnect is generally referred to as "information asymmetry." *Id.* This all assumes, of course, that the philanthropist is motivated by an interest in the goals of the nonprofit, as opposed to seeking a tax deduction, the ensuring admission of an unqualified child into a college, or acquiring the esteem attached to philanthropy, in which case there is little incentive for oversight. *Id.* at 234. Where philanthropy is mission-inspired, particularly in the context of religion, agency costs arise from the extraordinary difficulty facing donors who desire to confirm that their contributions are being used efficiently in the manner promised by the donee. *Id.* at 232. Even if the donors remains active in the charity, the difficulties in information gathering and assessment are enormous. Even donors who are able obtain accurate information about how many clients were served, how many client contacts occurred, or the number of food bags distributed by the nonprofit have learned little about the effects of their donations. Another donor type will find it very challenging to confirm whether the beneficiary of her largess is indeed teaching or behaving consistently with the ideology that attracted the donor to the agency in the first place.

190. The legally enforceable nondistribution constraint embodied in both state corporate

general have great difficulty obtaining information about nonprofit corporations' internal operations,[192] and most states are functionally and financially incapable of dealing with anything but the most egregious nonprofit behavior.[193] Even the multi-decade dysfunction of UMDNJ's governance only came to light in connection with the Justice Department's investigation of federal health program reimbursement issues, rather than state monitoring of its own agency.[194]

Effective monitoring is also circumscribed by rules of standing to sue. In most states, those with the power to sue a nonprofit for mismanagement or other troublesome behavior are the attorney general (who probably lacks the resources) and the entity's directors (who are unlikely to sue each other); in these jurisdictions, neither donors nor beneficiaries have standing.[195] Accordingly, Professor Manne summarizes the state of the nonprofit sector as follows: "The charitable sector of the economy, large, barely amenable to suit, and ineffectively reined in by the nondistribution constraint and the fiduciary rules under corporate and trust law, is largely unaccountable to anyone.

---

law and the rules governing federal tax-exempt status purportedly assures the donor that the entity's managers are not being over-paid or diverting income to themselves and intimates. *See* HANSMANN, *supra* note 184, at 17–18. Until recently, however, there has been almost no monitoring or enforcement. *See, e.g.*, Stephanie Strom, *Questions About Some Charities' Activities Lead to a Push for Tighter Regulation*, N.Y. TIMES, Mar. 21, 2004, at A23 [hereinafter Strom, *Questions About Activities*] (noting Sen. Grassley's concerns about insufficient Congressional oversight). In 2004, the IRS implemented the Executive Compensation Compliance Initiative, pursuant to which it issued compliance check letters to about 2000 tax-exempt organizations and found that "significant reporting issues exist[ed]." IRS, REPORT ON EXEMPT ORGANIZATIONS EXECUTIVE COMPENSATION COMPLIANCE PROJECT—PARTS I & II 1 (2007), http://www.irs.gov/pub/irs-tege/exec._comp._final.pdf. Over 30% of the organizations amended their Forms 990. *Id.*

191. Enforcement in New York is generally regarded as the exception; it was New York Attorney General Spitzer who ultimately challenged the NYSE board over CEO Grasso's compensation package. *See supra* Part II.A.

192. The Senate Finance Committee's staff discussion proposals seek to facilitate greater federal-state sharing and allow states to prosecute nonprofits' violations of federal tax laws. U.S. SENATE COMM. ON FIN., *supra* note 166, at 6–7.

193. *See* Manne, *supra* note 188, at 251 (quoting Mary Grace Blasko, Curt S. Crossley & David Lloyd, *Standing to Sue in the Charitable Sector*, 28 U.S.F. L. REV. 37, 38–39 (1993)). At least traditionally, the politics of prosecuting certain church-affiliated or otherwise powerful nonprofit may also be unappealing to an attorney general with grander political aspirations. *See id.*

194. *See supra* Part II.B.

195. *But see* Stern v. Lucy Webb Hayes Nat'l Training Sch. for Deaconesses & Missionaries, 367 F. Supp. 536, 540 (D.D.C. 1973) (allowing health care service purchasers to pursue class action breach of fiduciary duty claims against trustees and directors); City of Paterson v. Paterson Gen. Hosp., 235 A.2d 487, 495 (N.J. Super. Ct. Ch. Div. 1967) (allowing "parties especially interested" to compel performance by a nonprofit organization).

Fiduciary rules fail to operate effectively without enforcement and their enforcement against nonprofits is grossly inadequate."[196]

On the other hand, legal academics' commentary about the nonprofit sector ignores the most significant influences on nonprofits' behavior: granting agencies,[197] credit ratings companies, [198] insurers and bondholders,[199] feeder organizations,[200] donors,[201] and institutional members, all of which frequently become intricately involved with the nonprofit and impose conditions that significantly affect governance and operations. One study concludes that

---

196. Manne, *supra* note 188, at 252.

197. *See generally* Katherine O'Regan & Sharon Oster, *Does Government Funding Alter Nonprofit Governance? Evidence from New York City Nonprofit Contractors*, 21 J. POL'Y ANALYSIS & MGMT. 359 (2002) (discussing the use of government contracting decisions to influence nonprofit governance practices). O'Regan & Oster report that boards of nonprofits that receive funding from the City of New York behave differently than those that do not; specifically, boards of government funding recipients spend significantly less time fund-raising and more time on financial monitoring and advocacy. *See id.* at 374. While it was clear that board members of recipient entities themselves donate less, it is unclear whether it is because these boards are comprised of people with less means or because government funding is displacing private contributions by directors. *Id.* at 370. Other results suggested that government agencies may not be leveraging their positions to the extent they might to ensure adoption of governance monitoring mechanisms, *id.* at 366, and that boards of recipient agencies may expend more energy on issues of interest to the funding agency, displacing attention from other issues, *see id.* at 368.

198. For example, credit ratings agency Fitch noted in a report on corporate governance:

> [B]ond investors can punish self-interested behavior by raising a company's risk premium and limiting its access to competitive funding markets. . . . Financial covenants protect bondholders by requiring the financial management of the company to meet specified trigger ratios. Negative covenants might in some cases help curb actions that could harm bondholders by giving the creditor control over mergers and acquisitions, future investments, additional borrowings, and payment of dividends, among other examples.

FITCH RATINGS, EVALUATING CORPORATE GOVERNANCE, *supra* note 121, at 5.

199. Bondholders and underwriters "provide significant oversight and monitoring . . . of borrowers," from performing initial due diligence procedures to providing continuous monitoring throughout the life of the loan. Yetman & Yetman, *supra* note 95, at 16–17. Securities laws ensure the integrity of this market by requiring that all municipal bond issuers file annual financial and operating reports. *Id.* at 17. Bondholders' interests can be quite different than those of equityholders, due primarily to their "differing contractual structures and risk profiles." FITCH RATINGS, EVALUATING CORPORATE GOVERNANCE, *supra* note 121, at 4. "[T]he maximum potential economic return for bondholders is essentially fixed," and bondholders take a senior position in bankruptcy; consequently, bondholders' interests are in servicing debt and long-term recovery. *Id.*

200. *See, e.g.*, Yetman & Yetman, *supra* note 95, at 14.

201. The empirical literature is mixed on the question of the influence of major donors on nonprofit governance. Some claim that donors can fill the shoes of shareholders in terms of monitoring and demanding efficiency; however, others have found that major donors have little incentive to get involved with governance issues once they have given their money. Callen et al., *supra* note 10, at 494 (surveying the literature on this topic).

entities and individuals that have a financial relationship with or require audits of a nonprofit have a greater influence over the quality of financial reporting and attendant governance than does state regulation.[202] This study reaches three other conclusions about the relationship between governance and the quality of nonprofit financial reporting: monitoring by auditors has the strongest positive influence on nonprofits' financial reporting; "financial statements of nonprofits that have issued municipal bonds are more reliable for decision makers;" and state oversight and reporting laws "will only modestly improve nonprofit financial reporting quality."[203]

Additionally, as discussed above, major donor presence on boards is associated with lower administrative expenses as a proportion of total expenses.[204] Corporate legal scholars also overlook that some commercial nonprofits, such as hospitals and universities, operate in thriving markets that may impose some discipline on behavior (although the political difficulty of closing hospitals provides them some insulation from these market forces).[205]

---

202. *See* Yetman & Yetman, *supra* note 95, at 6, 36. The United Way is an example of an organization that requires audits by entities to which it gives in excess of a certain amount of funding. *Id.* at 14.

203. *Id.* at 6. California is the only state to have successfully enacted a mini-SOX law, the Nonprofit Integrity Act of 2004. S. 1262, 2003-2004 Leg., Reg. Sess. (Cal. 2004). It requires nonprofit entities with annual revenue in excess of $2 million to undergo annual audits by an independent accountant, overseen by a board audit committee. *Id.* The audit must be disclosed to the public and the California attorney general. *Id.* Other states have enacted less comprehensive statutes, with none engaging in significant regulation of governance structure. *See, e.g.*, HAW. REV. STAT. ANN. § 414D-140 (LexisNexis 2004) (authorizing the attorney general to remove directors who breach their fiduciary duties); HAW. REV. STAT. § 414D-233 (Supp. 2006) (requiring nonprofits to notify the attorney general of disposition of assets); ME. REV. STAT. ANN. tit. 9 § 5004 (Supp. 2006) (requiring charities to file audited financial statements as part of their license renewal applications); MASS. GEN. LAWS ANN. ch. 12, § 8F (West Supp. 2006) (requiring public charities with annual revenues exceeding $100,000 to file audited financial statements with the Public Charities Division); N.H. REV. STAT. ANN. § 7:28 (Supp. 2006) (requiring charities with annual revenues in excess of $500,000 to file audited financial statements with the attorney general). Then-New York Attorney General Spitzer, one of the first to propose state-SOX legislation, was unsuccessful in accomplishing passage of his bill. Strom, *Questions About Activities*, *supra* note 190. In 2005, Colorado, Connecticut, and Kansas enacted legislation addressing charitable organization financial disclosure and auditing. S. 205, 65th Gen. Assemb., 1st Reg. Sess. (Colo. 2005); S. 946, 2005 Gen. Assemb., Jan. Sess. (Conn. 2005); S. 121, 81st Leg., Reg. Sess. (Kan. 2005). More significant legislation was defeated in Texas, Minnesota, and North Carolina. C. Mark Pickrell & Terri Wagner Cammarano, Am. Health Law. Ass'n, Regulation of Nonprofit Governance: The Crest of the First Wave and the Shape of the Next, 6–7 (Dec. 2006) (unpublished member briefing, on file with the *Tennessee Law Review*).

204. Callen, et al., *supra* note 10, at 509.

205. Because so much of health care is subsidized by the government, the market is generally an insufficient tool to achieve ideal distribution of resources; thus, government intervention is required. New York is the most recent example. The recommendations of the New York Commission on Healthcare Facilities in the 21st Century, which was mandated by the

The take-away of the discussion so far is that state regulation and oversight have had little impact on nonprofit governance, with the exception of the handful of state statutes requiring financial audits of nonprofits with certain levels of revenue.[206] Although the strength and sufficiency of nonregulatory influences on nonprofit board behavior have yet to be determined, they probably hold more sway than government regulation. Without further empirical data, the safest operational assumption is that most of the oversight power and responsibility for ensuring mission integrity is vested in the nonprofit board. The following discussion argues that some of the formation options available to nonprofits offer an easy opt-out of the structures necessary to attaining strong governance.

B. Governance Structure in the Public Benefit Corporation

The Revised Model Nonprofit Corporation Act, which became final in 1988, recognizes three forms of nonprofit corporations:[207] public benefit,[208] mutual benefit,[209] and religious corporations.[210] These categories correspond

---

legislature to recommend changes for streamlining the state's hospitals, prompted significant litigation and protest regarding the commission's decisions to close nine hospitals and reconfigure 48 others. Joel Stashenko, *Panel Affirms Legality of Hospital Closing Process*, 238 N.Y. L.J. (2007), *available at* LegalTrac, Document No. A166641853.

206. *See* OSTROWER, *supra* note 147, at 6.

207. Michael C. Hone, *Introduction* to THE REVISED MODEL NONPROFIT CORPORATION ACT, at xix, xxiv–xxx (1987). By separating nonprofits into categories, the Revised Model Nonprofit Corporation Act followed the lead adopted by New York and California. *See* CAL. CORP. CODE §§ 5059–5061 (West 2007); N.Y. NOT-FOR-PROFIT CORP. LAW § 201 (McKinney 1970). Marion Fremont-Smith provides a concise history behind the Model Act and tracks which states have adopted the original and various revised versions. FREMONT-SMITH, *supra* note 16, at 151-52. Some states have not adopted the Revised Model Act, *cf. id.* at 151 (describing the history of nonprofit corporation statutes and the Revised Model Act); furthermore, some still recognize the corporation sole, which has no board of directors, *see generally* James B. O'Hara, *The Modern Corporation Sole*, 93 DICK. L. REV. 23 (1988) (discussing this form of corporate structure); Patty Gerstenblith, *Associational Structures of Religious Organizations*, 1995 BYU L. REV. 439, 454-62 (same).

208. Public benefit nonprofits are entities that "hold themselves out as doing good works, benefiting society or improving the human condition." Hone, *supra* note 207, at xxiv. Although public benefit nonprofits may compensate for the reasonable value of goods and services, they may not operate in a way that provides private economic benefit to their founders, officers, directors or controlling persons; thus, they may not distribute their profits or dividends and members do not have an equity interest in the assets of the nonprofit. *Id.* at xxiv–xxv. While organizations that qualify for federal tax exemption under Internal Revenue Code § 501(c)(3) must be incorporated as a public benefit corporation, 501(c)(3) status is not a necessary condition, since organizations that engage in lobbying may qualify as a public benefit corporation but not for federal tax exemption. *See id.* at xxv.

209. "Mutual benefit corporations hold themselves as benefitting [sic], representing and serving a group of individuals or entities." *Id.* at xxviii. "Members may have an economic interest in mutual benefit corporations . . . and they may receive distributions" upon dissolution.

with the goals of the particular nonprofit. Because mutual benefit corporations are most analogous to shareholder corporations,[211] this Article focuses on public benefit corporations, saving analysis of religious corporations for a companion piece.[212]

Every public benefit corporation must have a board of directors that exercises authority over the affairs and management of the corporation.[213] The method of selecting the board of directors is to be specified in the articles of incorporation; options include election, designation, appointment, or some combination of the three.[214] The details of statutory formation options reveal the potential to undermine board power.

First, the Revised Model Act's allowance for formation flexibility enables incorporators to vest the board's power in a single person, thereby eviscerating any pretense of having an active and engaged board.[215] Specifically, the Act provides:

Id.; *see also* REV. MODEL NONPROFIT CORP. ACT § 13.02 (1987) (allowing distributions upon dissolution).

210. As a general proposition, the rules that apply to the public benefit corporation also apply to the religious corporation, though not quite as extensively and with greater flexibility to avoid any possibility of First Amendment infringement. Hone, *supra* note 207, at xxix–xxx.

211. *See* Ira Mark Ellman, *Another Theory of Nonprofit Corporations*, 80 MICH. L. REV. 999, 1000 (1982) (noting that "nonprofit" is not necessarily synonymous with "charitable" and defining "mutual benefit nonprofits" as "those structured to satisfy customers' needs").

212. Interestingly, the Exposure Draft of the Revised Model Act did not include "religious corporations," which was added after the drafters received significant commentary objecting to its exclusion and arguing for the necessity of separate treatment. Moody, *supra* note 182, at 259–61; *see id.* at 267.

213. REV. MODEL NONPROFIT CORP. ACT § 8.01 (1987).

214. *Id.* § 8.04 & cmts. 1–2.

215. *See id.* § 8.01. This problem can also occur in states that allow the board to consist of as few as one director; the Revised Model Act provides that a board must have at least three directors. *Id.* § 8.03.

Marion Fremont-Smith suggests that this interpretation of nonprofit legislation is wrong: "Directors may not, however, abdicate their duty to direct, and they may be chargeable with losses resulting from failure to participate. They may not agree to place authority in one person or under one group of members, or agree that a director be only nominal." FREMONT-SMITH, *supra* note 16, at 162. I suspect that we are both right. Fremont-Smith cites *Ray v. Homewood Hospital, Inc.*, 27 N.W.2d 409 (Minn. 1947), in which three sets of individuals made a series of self-serving commitments to each other regarding the directorship and management of a nonprofit hospital, including an agreement that one of the three pairs would be appointed to the board as long as they promised in return not to take any action, *id.* at 410. The entire relationship fell apart when these "nominal" board members began asserting themselves, and one set of incorporators was ousted from their positions as trustee and president. *Id.* In this context, no one can disagree with the court's statements that

> [d]irectors may not agree to exercise their official duties for the benefit of any individual or interest other than the corporation itself, and an agreement by which individual directors, or the entire board, abdicate or bargain away in advance the judgment which the law contemplates they shall exercise over the affairs of the corporation is contrary to public

> The articles may authorize a person or persons to exercise some or all of the powers which would otherwise be exercised by a board. To the extent so authorized any such person or persons shall have the duties and responsibilities of the directors, and the directors shall be relieved to that extent from such duties and responsibilities.[216]

For example, these powers could be delegated to the CEO, who would presumably self-monitor. A structure that allows execution of board responsibilities by a single person, or even a few directors in some instances, undermines the entire purpose of a multi-member board, which is to ensure mission fidelity, operational oversight, and monitoring of management.[217] By

---

policy and void. They may not agree to abstain from discharging their fiduciary duty to participate actively and fully in the management of corporate affairs. The law does not permit the creation of a sterilized board of directors.

Id. at 411 (citations omitted). My point is that boards may quite lawfully delegate a significant amount of power to committees; in such circumstances, the guidance ensuring responsibility for oversight and consequences is scant. Fremont-Smith further voices this concern: "In all jurisdictions in which the corporate standard of care for nonprofit corporations has been adopted, a director's right to rely on information, reports, and statements prepared by other directors, officers, employees of the corporation, or outside experts in discharging his duties is also recognized." FREMONT-SMITH, *supra* note 16, at 205 (citing REV. MODEL NONPROFIT CORP. ACT § 8.30 (1987)).

216. REV. MODEL NONPROFIT CORP. ACT § 8.01(c). Persons to whom this kind of authority is delegated must behave in accordance with the standards of behavior, such as the duties of care and loyalty, impressed upon directors. *Id.* § 8.01 cmt; *see also id.* §§ 8.30–8.33. The Revised Model Act strives for flexibility, specifically noting the intent to accommodate entities that shift significant powers from the board to a representative assembly or convention when such entities are in session, and recognizing that some nonprofit boards are heavily involved in fundraising. *Id.* § 8.01 cmt. The proposed third edition of the Model Act is substantially similar:

> (a) Some, but less than all, of the powers, authority or functions of the board of directors of a nonprofit corporation under this [act] may be vested by the articles of incorporation or bylaws in a designated body.
>
> . . . .
>
> (c) To the extent the powers, authority, or functions of the board of directors have been vested in a designated body, the directors are relieved from their duties and liabilities with respect to those powers, authority, and functions.

MODEL NONPROFIT CORP. ACT THIRD EDITION § 8.12 (Proposed Exposure Draft 2006) (alteration in original). One important, but vague, difference between the Revised Model Act and the proposed third edition of the Model Act is that the proposed version provides for delegation of "[*s*]*ome, but less than all*, of the [board's] power, authority or functions." *Id.* (emphasis added).

217. Scholars of behavioral psychology and economics support this legal framework, concluding that a multi-member board offers a better leadership structure for corporations than does an individualistic model. *See* Bainbridge, *supra* note 6, at 12; Marleen A. O'Connor, *The Enron Board: The Perils of Groupthink*, 71 U. CIN. L. REV. 1233, 1237 (2003). Groups actually make decisions more quickly than individuals, and although data are mixed regarding whether group decisions are better than those made by the most superior individual in the group, consensus decisionmaking seems to exceed that of the average individual. *See* Bainbridge,

comparison, the ALI Draft Principles prefer that every director bear governance responsibility.[218] This is the better model.

Even if authority for oversight does vest in every director, committee structure and delegation can cause directors difficulty in obtaining the necessary information about or involvement with certain matters crucial to the exercise of that authority. In modest- to large-size nonprofits, committees are crucial to nimbleness and efficiency.[219] Most entities of any size create executive committees to make time-sensitive decisions between meetings without convening the entire board.[220] Yet when the executive committee is wholly vested with significant powers with no duty to report to the board, and the board has no oversight responsibility, the diversity of perspectives and skills achieved by careful composition of the board can be rendered useless.

Delegation of specialized tasks to committees based upon directors' expertise is entirely appropriate,[221] but committee assignments can be manipulated to achieve the agenda of management or a minority faction of the board.[222] Outcomes can be predetermined, controversy avoided, or improprieties hidden by placing the right people on the right committees. The

---

supra note 6, at 14–18. Groups are also more effective at reining in the overconfident, uncritical, or overly invested officer or director. *Id.* at 30. This is supported by research data showing that, while individuals are frequently superior to groups in creative activities and individuals generate more ideas when working alone, groups are superior for evaluative tasks, which are a board's primary activities. *Id.*

218. ALI DRAFT PRINCIPLES, *supra* note 3, § 320 cmt. (b)(2).

219. Unless otherwise specified in the articles or bylaws, boards may create one or more committees and may appoint directors to serve on them. REV. MODEL NONPROFIT CORP. ACT § 8.25(a). Although the Revised Model Act anticipates "board appointment" of committees, *see id.*, many nonprofits operate by soliciting interest of directors, then presenting a draft of committee assignments to the board for a vote.

220. The Revised Model Act anticipates and permits board actions without a board meeting as long as such actions have unanimous written consent. *Id.* § 8.21. The ALI Draft Principles represent a significant step forward in this area and could be further improved by explication of transparency obligations. The comments to the ALI Draft Principles observe that "[i]n general, the executive committee acts for the board between regular board meetings, and may exercise all powers of the board unless expressly limited by statute or the organizational documents." ALI DRAFT PRINCIPLES, *supra* note 3, § 325 cmt. (b)(2). The comments further elaborate that the executive committee should not usurp the board, may be reversed by the board, and should not be allowed inordinate power. *Id.*

221. Committees are useful for "developing long-term plans, supervising investments, raising money, making grants, and evaluating management." REV. MODEL NONPROFIT CORP. ACT § 8.25 cmt.

222. *See* HANSMANN, *supra* note 184, at 42 ("[D]elegation can . . . produce seriously inefficient outcomes by empowering committee members to impose their own idiosyncratic preferences on the group as a whole"); Michael Meyers, Letter to the Editor, *Inside, the ACLU Shows Hypocrisy*, USA TODAY, June 23, 2006, at 11A (noting that the ACLU's Executive Director "informs all committee appointments, including the composition of the committee [its President] appointed that has proposed further restrictions on board members' public and internal speech").

appointment and operation of the New York Stock Exchange's compensation committee is a notorious example of how a committee can be organized to benefit management:

> [I]n Grasso's later years as Chairman—years when his compensation reached very high levels—he had a hand in selecting the Board members who decided his compensation.
>
> Not only did Grasso have significant input in the selection of Board members throughout his tenure; he also had the unfettered authority to select which Board members served on the Compensation Committee and likewise, to select the Committee Chair.... Several members of the Committee during Grasso's tenure had friendships or personal ties or relationships with Grasso....
>
> ....
>
> He had a strong influence in who was selected as members of the Nominating Committee and the Board, and he personally selected which Board members served on the Compensation Committee.[223]

The ALI Draft Principles also address the issue of committee delegation. Proposed section 325 provides that, while the governing board may delegate to committees the "authority to perform the board's functions and obligations,"[224] such delegation is subject to the board's ultimate oversight responsibility.[225] This oversight mechanism works, however, only if another mechanism exists for consistent and complete communication between committees and the full board. Bylaws should specifically address what types of matters may be resolved by the executive committee and what types of matters require full board involvement, even if a special meeting is required.[226] Further, bylaws should require the executive committee to communicate in a detailed and timely fashion with the rest of the board about matters handled at the executive committee level.[227] They should also describe the board's right to access

---

223. People *ex rel.* Spitzer v. Grasso (*Grasso III*), No. 401620/04, 2006 WL 3016952, at *5 (N.Y. Sup. Ct. Oct. 18, 2006) (citations omitted).

224. ALI DRAFT PRINCIPLES, *supra* note 3, § 325(a)(1).

225. *Id.* § 325(a). The board must determine the manner of their supervision and monitoring of such delegation. *Id.* § 320(b)(2)–(3).

226. State statutes may also provide guidance on these roles:

> [N]onprofit-corporation statutes typically prohibit a committee of the board from taking four specific actions: (1) authorizing distributions; (2) approving (or proposing to members) action that must be approved by members, such as dissolution, merger, or the sale, pledge, or transfer of all or substantially all of the corporation's assets; (3) electing, appointing, or removing directors, or filling vacancies on the board or any board committee; and (4) adopting, amending, or repealing the articles or bylaws.

Id. § 325 cmt. (b)(3); *see also* MODEL NONPROFIT CORP. ACT THIRD EDITION § 8.25(e) (Proposed Exposure Draft 2006) (proposing essentially the same restrictions).

227. For an example of the importance of such provisions, see Stephanie Strom, *Rift at A.C.L.U., on Fund-Raising and Leadership*, N.Y. TIMES, Dec. 8, 2005, at A1 ("The internal friction has roiled the organization, which is unaccustomed to scrutiny of its operations, and

information that is disclosed to the executive committee but not to the entire board.[228] These issues would easily be resolved by requiring each committee to post its minutes on the corporate web page.[229]

Public benefit corporations may have self-perpetuating boards or boards appointed in whole or in part by "members." Each model presents its own governance challenges. Board selection in the nonmember entity is relatively straightforward. The Revised Model Act provides that, unless otherwise indicated in the nonprofit's articles of incorporation, the board of a corporation without members shall select its own directors;[230] this is referred to as the "self-perpetuating" board.[231] In this context, directors who are appointed may serve indefinite terms unless restricted by the articles or bylaws, while elected directors are limited to a maximum term of five years.[232] Smaller nonprofits and nonprofits reliant on a single significant donor must resist the board acting as the alter ego of the entity's founders or original board members. Fidelity to mission must guide all decisions, and the board is obliged to monitor management. Founding directors should consider imposing term limits upon themselves and including on the initial board at least one or two nonfounding directors with prior nonprofit board experience—directors who might also serve as "monitoring directors."

Ideally, new board members of medium to large nonprofits with self-perpetuating boards are selected by a nominating committee that does not include the executive director and is not dominated by the entity's founders. To build a board comprising individuals with a range of perspectives and skill sets, the nominating committee should begin the search and selection process by creating a list of criteria and desirable qualifications. In doing so, it should take into account the backgrounds and expertise of the current directors and the benefits of stakeholder representation. The appointment process should include an interview with the nominating committee. Additionally, any new board member's tenure obviously should begin with an orientation that comports with the myriad best practices available on the subject.

---

prompted members of the executive committee to try to limit access to recordings of board meetings. . . . Citing a need for efficiency, the organization now often insists that board members who want to question [its Executive Director] or other senior staff members first get approval from the executive committee.").

228. *See generally id.* (further illustrating this need).

229. *See infra* Part IV.C.2.

230. REV. MODEL NONPROFIT CORP. ACT § 8.04 (1987). "A director elected by the board may be removed without cause by the vote of two-thirds of the directors then in office or such greater number as is set forth in the articles or bylaws." *Id.* § 8.08(h). In removing a fellow director, the board "must meet the duty of care and . . . loyalty." *Id.* § 8.08 cmt. 2.

231. *Id.* § 8.04 cmt. 2. The majority of charitable corporations have self-perpetuating boards. FREMONT-SMITH, *supra* note 16, at 159; Mulligan, *supra* note 12, at 1986. The self-perpetuating board structure may exacerbate boards' confusion over to whom they are accountable. *Id.*

232. REV. MODEL NONPROFIT CORP. ACT § 8.05(a), cmt.

Membership nonprofit corporations are different. Nonprofit corporations may or may not have members.[233] "There are no precedents and little commentary about the duties of members."[234] Members may or may not have any actual powers,[235] and any such powers may be allotted to them based upon the membership class to which they belong.[236] Furthermore, members have no fiduciary obligations to the corporate entity.[237]

Professor Moody, a member of the drafting committee for the Revised Model Act, provides an explanation for the section on nonprofits' membership:

> [M]embers generally relate to the organization by participation rather than by the financial interest generated by an investment. The Revised Act attempts to define "member" to take into account the governance functions fulfilled by members in nonprofit organizations. Calling some person or category of persons a "member" does not create them as such under the Revised Act. Members are defined as persons who, "pursuant to a provision of a corporation's articles or bylaws, have the right to vote for the election of a director or directors." A person is not a member by reason of any rights the person may have as a delegate; any rights the person may have to designate a director; or any right a person may have as a director of the corporation.
>
> Membership rights are closely connected with the function of members of a nonprofit organization.238

The statutory framework for public benefit corporations presumes that members, because they do not have a financial stake in the corporate enterprise,

---

233. *Id.* § 6.03; LINDA O. SMIDDY & LAWRENCE A. CUNNINGHAM, SODERQUIST ON CORPORATE LAW AND PRACTICE p. 2–4 (3d ed. 2007). The existence and power of members may be enumerated in the articles of incorporation or bylaws. REV. MODEL NONPROFIT CORP. ACT § 6.01.

234. FREMONT-SMITH, *supra* note 16, at 441.

235. The articles of incorporation or bylaws may provide that members do not have voting rights. *See* REV. MODEL NONPROFIT CORP. ACT § 7.21(a). Sometimes, entities designate major donors or "friends" as members. FREMONT-SMITH, *supra* note 16, at 440.

236. REV. MODEL NONPROFIT CORP. ACT § 6.10. The Revised Model Act requires membership organizations to establish fair and reasonable procedures for the termination, expulsion or suspension of a member. *Id.* § 6.21. For further discussion of the powers of nonprofits' members, see Moody, *supra* note 182, at 270–71; Charles H. Steen & Michael B. Hopkins, *Corporate Governance Meets the Constitution: A Case Study of Nonprofit Membership Corporations and Their Associational Standing Under Article III*, 17 REV. LITIG. 209, 220 (1998).

237. Dana Brakman Reiser, *Decision-Makers Without Duties: Defining the Duties of Parent Corporations Acting as Sole Corporate Members in Nonprofit Health Care Systems*, 53 RUTGERS L. REV. 979, 983 (2001) [hereinafter Reiser, *Decision-Makers Without Duties*].

238. Moody, *supra* note 182, at 270–71. The Revised Model Act requires annual membership meetings, § 7.01(a), and dictates that at such meetings, "[t]he president and chief financial officer shall report on the activities and financial condition of the corporation" and act on matters for which notice has been provided, *id.* § 7.01(d). The Act also provides for special meetings and court-ordered meetings. *Id.* §§ 7.02–7.03.

lack an incentive to monitor corporate behavior.[239] Because of this lack of incentive, many commentators analogize members to shareholders with no practical means for oversight. Nonprofit corporate law addresses the resultant presumed void in oversight authority by investing in the state attorney general the power to bring derivative actions on behalf of the nonprofit corporation.[240]

Neither these commentaries nor the law adequately account for the spectrum of member corporations extant today, many of which do not fit the mold anticipated by the drafters. The member corporate form is employed in a variety of distinct ways; these different applications have significant implications for any analysis of the governing laws. The following discussion focuses on the kind of member corporation frequently, though not exclusively, employed in nonprofit health systems. In this model, the parent corporation is the sole member of all of the subsidiary entities that make up the system, which might include multiple hospitals, a self-insurance trust, and even a laundry; Professor Reiser calls this model the "parent-acting-as-sole-corporate-member" (PASCM).[241] In religiously-sponsored entities, the member may be even more remote from the actual operating entities—the parent may itself be a member corporation, in which the sole member is the religious order or diocese that originally founded the parent university, hospital, or social service agency back in the days of uncomplicated corporate structures. Most often, the system of parent and subsidiaries operates as a conduit through which the member fulfills its mission, with the composite entities separately incorporated for licensure, liability, reimbursement, financing, or any number of other legal or financial reasons.

The PASCM uses this structure to set its subsidiaries in pursuit of the overall system mission, reserving certain powers to the PASCM itself and appointing some or all of the subsidiaries' directors;[242] frequently, the subsidiaries' board membership will overlap with the PASCM's board.[243] The goal of the PASCM form is that each subsidiary will operate to the benefit of the system as a whole.[244] This may require any one entity's directors to pursue

---

239. *See* Hone, *supra* note 207, at xxvi.

240. *Id.* at xxvii; *accord* REV. MODEL NONPROFIT CORP. ACT § 1.70; *see also id.* § 6.30(f) (requiring members or directors who bring derivative actions to notify the attorney general). Derivative actions may also be initiated by certain members or any director. *Id.* § 6.30(a).

241. *See generally* Reiser, *Decision-Makers Without Duties*, *supra* note 237 (explaining in detail why this corporate form may be chosen, its pros and cons, and the failures of the law to adequately address the duties of such corporations' members).

242. *See* FREMONT-SMITH, *supra* note 16, at 159 ("This sole-member parent corporation, in addition to retaining the inherent powers of a member to elect directors and approve by-law changes, would have the power to ratify budgets, appoint or elect the officers, and approve changes in activities, dissolutions, or disposition of assets."); Reiser, *Decision-Makers Without Duties*, *supra* note 237, at 991.

243. Nicole Huberfeld, *Tackling the "Evils" of Interlocking Directorates in Healthcare Nonprofits*, 85 NEB. L. REV. 681, 689 (2007).

244. *Id.* It would be interesting to know how many subsidiaries' articles and bylaws include as part of the mission statement fidelity to the overall mission of the parent or system.

a strategic path that is unfavorable to the subsidiary, but favorable to the system overall or to another entity within the system.[245]

For these reasons, it is impossible to imagine how the PASCM model can comply with a best practice demanding independent boards for each system component. The concepts are philosophically inapposite; that a director is appointed by the PASCM suggests a dependence in fact, irrespective of whether the relationship is captured by whatever definition of independence has been borrowed from the available alternatives. By virtue of the method of their appointment, such directors (or at least the PASCM) must assume that they are selected with an expectation of responsiveness to the member's preferences and interests. Thus, a requirement of true independence threatens the PASCM model, a consequence that proponents of board independence do not appear to have considered. Furthermore, advocates of proportionate independence provide no guidance on the relationship between it and the PASCM model, nor any identification of an overarching priority; either independence or systemic cohesion must trump.

It is premature and likely unwise to ring the death knell of the PASCM model, despite the opinions of some attorneys general that a decision by a subsidiary's board that is driven by the needs of the parent or a sibling subsidiary breaches its directors' fiduciary duties.[246] Yet an unyielding requirement of director independence would certainly preclude interlocking directorates and cuts strongly against PASCM-appointed directors.

While the institutional member model has several merits, there is no question that problematic conflicts can arise between unrelated interests of the member and the member corporation, in which case directors should act in the best interest of the entity on whose board they serve rather than the independent interests of the member. A brief example makes the point. Religious orders of nuns are, quite literally, dying out and facing significant challenges in obtaining the resources necessary to sustain their aging sisters. The assets of the entities of which they are members could serve as obvious sources of cash; the member orders could push the boards to monetize the assets of their hospitals or schools, then use the resulting proceeds to care for their sisters. Such action clearly departs from the mission of the school or hospital corporation on whose board the directors serve, as well as the system of which the corporation is a

---

Such an approach would, on the one hand, ameliorate mission fidelity conflicts. Comments to the ALI Draft Principles suggest that "[i]f the purpose of the charity, as declared in its organizational documents, includes the purpose of that member or group, then the board (and board members) of the charity may properly take into account those larger interests in discharging the duty of loyalty." ALI DRAFT PRINCIPLES, *supra* note 3, § 310 cmt. (a)(3). The question would then become whether such a close linkage heightens the risk of piercing the corporate veil between the subsidiary and the parent.

245. Huberfeld, *supra* note 243, at 689.

246. Greaney & Boozang, *supra* note 12, at 78–80 (discussing the position taken by some attorneys general that the directors of a managed care entity breach their fiduciary duties by making decisions to benefit sibling hospital corporations).

part. Since it is this mission that ultimately must govern the directors' behavior, it would be a breach of the duty of loyalty to accede to the member's wishes. On the other hand, an expressed preference by a religiously-affiliated member that the entity directors make decisions consistent with the religion's tenets would be a more reasonable expectation of directors, especially if articulated in the entity's formation documents.

This section illustrates that there are more fundamental ways to address flaws in nonprofit governance structure than requiring board independence. An obvious starting place for strengthening nonprofit governance would be to close the several gaps in the statutory framework that allow boards to circumvent individual and aggregate director responsibility for the organization. More aggressive legislative reforms should await empirical support, be specifically geared to the unique nature of the nonprofit, and proceed on the assumption that boards serve as the primary nonprofit oversight mechanism. Having identified corporate law's feeble contribution to strong nonprofit governance, the next section discusses what else might be done to strengthen governance.

C. Strengthening Governance Without Requiring Independent Boards

Nonprofit organizations have several options for strengthening governance short of radically restructuring board composition. Some of the following suggestions may very well be inappropriate for, or require accommodation to suit, the small nonprofit.[247] Nevertheless, they should serve as useful starting points for experimentation.

1. Structure Governance to Enhance Active and Effective Leadership

Any conversation about nonprofit boards must begin with a clear articulation of what constitutes good governance—what goals the boards should aim to achieve. In the nonprofit setting, a board should strive first and foremost to identify its purposes; namely, to answer the question, "Why does the entity exist?" This is generally referred to as the entity's "mission." A mission should be relevant, clearly articulated, and at least potentially achievable. A nonprofit board must be capable of: overseeing effective and efficient accomplishment of the entity's mission while ensuring that the mission remains relevant to the changing environment; selecting and supervising management; properly stewarding the entity's resources; and ensuring compliance with applicable legal and accounting standards.

Although it might be easy enough to identify the goals of good governance, it is much more difficult to identify the elements required to achieve ideal governance, including the composition of the board of directors. The nonprofit sector comprises incredibly diverse entities, and the sector's existence serves

247. *See, e.g.*, Ellen P. Aprill, Essay, *What Critiques of Sarbanes-Oxley Can Teach About Regulation of Nonprofit Governance*, 75 FORDHAM L. REV. 765, 769 (2007) (discussing the need for "special small firm rules" for certification and disclosure requirements).

important and fundamental roles in society beyond the mere provision of services. These factors make it very difficult to measure either the quality of governance or the success of the nonprofit entity. The varied and amorphous nature of many nonprofits' goals makes it difficult to identify a standard that is helpful in determining whether those goals have been accomplished. Judgments regarding a nonprofit's level of success can vary, depending upon which constituency group is making the assessment.

To further confuse the mix, it remains unclear to whom nonprofit boards are ultimately accountable, with the most realistic answer being that they are accountable to multiple stakeholders—beneficiaries, donors, taxpayers, bondholders, licensing agencies, the IRS, and the state attorney general.[248] The good governance movement, with its focus on corporate compliance, risks restricting boards' attention to the demands of regulatory adherence to the exclusion of serving beneficiaries. A board composed of too many "experts" as a result of the push for independent boards can further exacerbate this problem, as it may be more natural for such directors to focus on finances and regulations, rather than the enterprise of charity.

Although a generic formula for board composition that will ensure ideal governance is impossible to articulate, this impossibility does not preclude the development of guidelines that are adaptable to each specific organization. These guidelines should focus primarily on mission accomplishment and secondarily on compliance. Because particular board functions, such as oversight of management and forcing a change of a weak CEO, are likely to be performed most effectively by outside directors,[249] this Article recommends the inclusion of at least some formally designated "monitoring directors." The remaining directors, though, should not take such denomination to relieve them of oversight responsibility.

Thus, as a general matter, nonprofit boards should consist of multiple members who are committed to the mission and who have the skills required for leadership; additionally, boards should include at least one member designated as a monitoring director.[250] Funding and rating agencies should be

248. *See supra* notes 197-203 and accompanying text.

249. *See* Langevoort, *supra* note 99, at 801 ("[A]n independent director is one who actually takes the monitoring task for the benefit of the shareholders and/or other constituencies seriously."); *see also Beyond "Independent" Directors*, *supra* note 99, at 1561 (discussing the NYSE's requirements that "outside directors meet separately from management to evaluate its performance and to consider shareholder concerns," thereby "creating a structural mechanism to combat managerial domination").

250. Presumably nonprofits are best able to identify directors with allegiance to their mission. Including some number of independent directors on boards helps ensure that boards are faithfully performing the monitoring function. Religious members and the PASCM can retain control over the corporate mission by reserving the power to articulate the mission and to approve any changes to the statement of purpose in the articles of incorporation. Sister Melanie DiPietro, Founder, Seton Ctr. For Church-Related Nonprofit Corps., Nonprofit Corporations Governance Issues, Presentation at the Seton Center for Church-Related Nonprofit Corporations Inaugural Conference: Catholic Charities and the Diocese: Autonomy and Relatedness (Nov. 17,

more circumspect in their imposition of governance conditions. Although data show that the requirement of an audit is tremendously beneficial in producing reliable financial statements, other governance requirements appear to produce few tangible benefits. After conducting successful experimentation with board composition and gaining experience with directors of different backgrounds, each nonprofit will eventually be able to amend its bylaws to reflect its ideal board formula.

The characteristics of directors suited for the monitoring role logically start—but do not end—with the concept of independence; monitoring directors must not be related to an employee or themselves employed or compensated by the corporation on whose board they serve, nor by a parent or member institution. Further, the monitoring director should not be selected, nominated, or appointed by the member institution; however, the member's concurrence in the director's appointment is likely reasonable. Nevertheless, the exclusionary qualifications of the monitoring director should be more robust, in that they should encompass any social or psychological relationship likely to interfere with unfettered judgment or willingness to express a different perspective. Consequently, the monitoring director should not be the subordinate in a power relationship with the board chair or a religious superior who serves on the board. Additionally, a director who has served as an officer of the entity, parent, or member institution should not qualify as the monitoring director.

Boards should publicly identify, on an annual basis, who among their directors are the monitoring directors; they should also document their analysis and conclusions about whether relationships exist that arguably could interfere with a director's monitoring functions. At a minimum, at least one monitoring director should serve on the executive, compensation, and audit/compliance committees; preferably, a different monitoring director would be assigned to each. Board and committee meeting minutes should specifically note when and why a monitoring director dissents from a vote. Finally, monitoring directors should be subject to term limits in order to combat the norming that will naturally result from the relationships created with other board members over time.

Much of the empirical literature focuses on the independence-undermining biases that result from directors' loyalty or feelings of obligation to the CEOs who have appointed them.[251] Best practices in the for-profit setting increasingly urge that the CEO not serve as the board chair for precisely this reason. The nonprofit sector should address this, as well. Presumably, a nonprofit could further ameliorate the loyalty phenomenon by appointing directors through a nominating committee (that does not include the CEO) whose recommendations are then approved by the board as a whole.[252] Prior to

---

2006). One observer has suggested that nonprofits' members issue an annual report on whether they believe the entity is faithfully adhering to its mission. *Id.*

251. *See, e.g.*, O'Connor, *supra* note 217, at 1245 ("Social psychology indicates that the independent directors are likely to feel strong loyalty to their appointer.").

252. *See generally Developments in the Law*, *supra* note 107, at 2194–95 (discussing the

considering or receiving names of potential directors, the nominating committee should, as with any corporate officer position, outline the qualifications desired in the prospective director. For example, the board might specify a need for financial experts sufficiently qualified to staff the audit and finance committees.[253] Nonprofit nominating committees should also publish open board positions in venues that might produce qualified candidates, such as the *Chronicle of Higher Education*, a hospital trade magazine, or a church bulletin.

2. Redefine Transparency

A strong and qualified board will work only if it knows what is really going on inside the organization. A nonprofit's connection and service to its constituencies—and to the public—will occur only if the community has easy access to the inner workings and policy decisions being made by the nonprofit. Therefore, nonprofits should conceive of transparency much more broadly than they have to date.

Transparency has been emphasized in the for-profit world. Much of Sarbanes-Oxley and the stock exchanges' listing standards require transparency of matters deemed important to investor decisionmaking about stock ownership and information relevant to confidence in corporate leadership.[254] Director independence is intended to ensure shareholders' access to necessary information from the perspective of a knowledgeable insider who is sufficiently objective to translate the information for the shareholders' benefit.[255]

Transparency should be at least as important in the nonprofit sector; voice in nonprofit governance is rendered meaningless without attendant access to information. Robust transparency has multiple layers: transparency between management and the board; between board committees and the full board; between the entity and members; between the nonprofit and those charged with oversight of it, namely the IRS and the state attorney general; between the entity and its constituencies, including donors, beneficiaries, and bondholders; and between the entity and the public.

NYSE and NASDAQ standards' provisions that deal with issues of independent director loyalty in the nomination process).

253. *Cf.* Daniele Marchesani, *The Concept of Autonomy and the Independent Director of Public Corporations*, 2 BERKELEY BUS. L.J. 315, 327 (2005) (discussing the importance of appointing competent directors who "have the experience and knowledge necessary for the job").

254. *See, e.g.*, Sarbanes-Oxley Act of 2002, Pub. L. No. 107-204, tit. IV, 116 Stat. 745, 785 (2002) (codified as amended in scattered sections of 15 U.S.C.) ("Enhanced Financial Disclosures").

255. *Beyond "Independent" Directors*, *supra* note 99, at 1562–64. "This informational transparency function separates into two distinct but complementary components: information forcing and information validation." *Id.* at 1563.

Transparency between management and a board, and within a board, should be complete; complete transparency allows the nonprofit to take advantage of the expertise and perspectives brought by various directors and to ensure that directors possess the information necessary to fulfill their fiduciary duties, including mission preservation. The board's monitoring function requires director insight into the entity's activities, with attendant access to all relevant information and processes. Further, board and committee members should have an opportunity to help shape the meeting agenda. Consequently, board bylaws should provide for a call for agenda items, which may include a request for background information, from the entire board or committee membership in advance of the meeting.[256] The board should have a staff person assigned directly to this task and retain the ability to request that this staffer collect information, have a report prepared, or obtain the presence of any senior or middle manager, without requiring mediation by the CEO or other top manager.[257]

Mission primacy requires that beneficiaries have access not only to information about available services, but also to information about quality or legal problems related to the delivery of those services. Mission integrity entitles donors and taxpayers to know to what end their support is being dedicated; thus, it also requires the nonprofit's complete transparency with the public at large, save for information subject to attorney-client privilege.[258]

A presumption of transparency in the nonprofit sector would represent a dramatic shift in ethos and perspective. Various government agencies, Independent Sector, and similar entities regularly tout transparency.[259] For

---

256. *See* Marchesani, *supra* note 253, at 337.

257. *See id.* at 336–37 (describing boards' need for direct access to information and the potential problems that arise when CEOs control information flow).

258. Total transparency will also enable the market for philanthropic dollars to operate at optimal efficiency. *See* Callen et al., *supra* note 10, at 509 (noting that donors are less likely to contribute to administratively inefficient organizations).

259. *See, e.g.*, IRS, Good Governance Practices, *supra* note 5, at 6 (emphasizing organizational transparency by recommending that the entity's 990s, updated financial statements, and annual reports be posted on the charity's website); MD. ASS'N NONPROFIT ORGS., STANDARDS FOR EXCELLENCE: AN ETHICS AND ACCOUNTABILITY CODE FOR THE NONPROFIT SECTOR 8, 24 (n.d.), *available at* http://www.marylandnonprofits.org/html/standards/documents/Booket507Revised.pdf (stating that nonprofits "should provide the public with information about their mission, program activities, and finances," but giving little specific direction regarding transparency); U.S. DEP'T TREASURY, ANTI-TERRORIST FINANCING GUIDELINES: VOLUNTARY BEST PRACTICES FOR U.S.-BASED CHARITIES 3–4 (2006) (recommending in part that the charity make public the names of its board members and their salaries, a list of its "five highest paid or most influential employees" and their total compensation packages, a list of the board members and key employees of any subsidiaries or affiliates that receive funds from the charity, and, upon request, records such as an annual report and financial statements); *see also* Press Release, U.S. Dep't Treasury, Treasury Updates Anti-Terrorist Financing Guidelines for Charitable Sector (Sept. 29, 2006), http://www.treas.gov/press/releases/hp122.htm (last visited Nov. 20, 2007) (announcing the

example, Independent Sector's Panel on the Nonprofit Sector has advocated that "the public must have access to accurate, clear, timely, and adequate information about the programs, activities, and finances of all charitable organizations."[260] The actual norms of the nonprofit sector are reflected in the ALI Draft Principles, which provide that directors must have access to any information that is necessary to fulfilling their fiduciary duties.[261] Yet they must also,

> except as may be waived by board policy, preserve the confidentiality of information obtained as a member of the governing board, but may disclose to members of the board, to the attorney general, or to a court otherwise confidential organizational information that [they] believe[] appropriate to prevent, mitigate, or remedy harm to the charity.[262]

This approach is wholly inadequate; thus, it is incumbent upon states to enact legislation requiring pervasive transparency by nonprofit corporations, modeled after the open meeting laws that apply to governmental entities.[263]

Regardless of one's perspective on nonprofits—whether one subscribes to the view of attorneys general that all nonprofit assets belong to the public, or the opinion of tax scholars that nonprofits are beneficiaries of significant public subsidies in the form of their tax exemptions, or the belief that nonprofits are primarily duty-bound to their members or beneficiaries—charities have a special place in our civil society. Little rationale exists to facilitate nonprofits' withholding of information about purpose, the use of funds, levels of success achieved, or challenges met in pursuing their mission.

In fact, the current UMDNJ Trustee web page[264] provides a stark contrast to the inadequate concept of transparency embraced by the ALI. It includes a full list of board members and their biographies, and it features click-throughs for committee composition, board bylaws and resolutions, meeting minutes, and

---

release of the updated guidelines and encouraging nonprofits "to take a proactive risk-based approach" to governance and transparency). *See generally* PANEL ON THE NONPROFIT SECTOR, *supra* note 4 (making various recommendations for strengthening transparency and accountability).

In comparison, the proposals of the Senate Finance Committee are rather anemic; they would require disclosure of the exempt organizations' financial statements, certain audits, 990s and other tax returns, applications for tax-exempt status, and the entity's IRS determination letter. U.S. SENATE COMM. ON FIN., *supra* note 166, at 10–11. The Committee's proposals would further require the organizations to post the most recent five years' worth of these documents on their websites. *Id.*

260. PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 21.

261. ALI DRAFT PRINCIPLES, *supra* note 3, § 340 cmt. a; *see also id.* § 340.

262. *Id.* § 340(b).

263. For examples of state open meetings laws, see MD. CODE ANN., STATE GOV'T §§ 10-501 to 10-512 (LexisNexis 2004); WIS. STAT. ANN. §§ 19.80 to 19.98 (West 2003).

264. About UMDNJ: UMDNJ Board of Trustees, http://www.umdnj.edu/about/about04_trustees.htm (last visited Nov. 24, 2007).

contracts;[265] every federal monitor report is available, as well.[266] The UMDNJ model represents the more appropriate approach to transparency in the nonprofit sector.

Because of the unique role the nonprofit plays in our society, confidentiality should have little place in its operations. Health care providers' quality of care issues should be known to the public. Any charitable entity's discovery of fraud or other noncompliance with the law, as well as its proposed remedy, should be reported to law enforcement—and is likely to become public in any case. If a charity's auditors have made a "going concern" determination, that information should not be limited to bondholders and state licensing agencies; it is also important to employees, vendors, beneficiaries, and donors.[267] Issues such as imminent insolvency, closure, and potential acquisition are important to numerous constituencies of both the acquirer and the failing entity, including creditors and the state, and should not be a secret. An advocacy organization's internal debates about whether to take particular positions on issues should be known by its members, who may then assert their own views or transfer their loyalty and membership to an organization that more closely reflects their perspectives.

Legal advice that is subject to attorney-client privilege would be an exception to this presumption of transparency. Employee termination, especially termination involving a "negotiated separation" that addresses future employment references, is one of the more challenging kinds of information and may be subject to a contractual confidentiality provision, if not attorney-client privilege. On the other hand, the termination of a physician's medical staff privileges due to quality of care problems, or a teacher's termination for an inappropriate relationship with or harassment of a student, might reasonably be considered information to which a variety of constituencies, including prospective employers, patients, and students, are morally entitled. Similarly challenging is whether the proprietary information of commercial nonprofits, such as universities and hospitals, should be subject to this pervasive disclosure approach. No evidence indicates that state hospitals and universities have been adversely affected by transparency; in fact, some public universities and hospitals are among the nation's finest and most successful, which suggests that complete confidentiality is not essential to market success.

The most common and convenient method by which much of society obtains information today is through the internet. The current norm that entities

---

265. *Id.*

266. Federal Monitor: University of Medicine and Dentistry of NJ, http://www.umdnj.edu/home2web/federalmonitor/ (last visited Nov. 24, 2007).

267. The Auditing Standards Board's Statement on Auditing Standards No. 59 requires an auditor to document when she has "substantial doubt . . . about the entity's ability to continue as a going concern and . . . is not satisfied that management's plans are enough to mitigate these concerns." Elizabeth K. Venuti, *The Going-Concern Assumption Revisited: Assessing a Company's Future Viability*, CPA J., May 2004, at 40, 40, *available at* http://www.nysscpa.org/cpajournal/2004/504/essentials/p40.htm.

make information about themselves available to the public "upon request" is outmoded. Every nonprofit organization should have a web page that facilitates the kind of transparency anticipated by this discussion. It should include the nonprofit's mission statement; its articles of incorporation and bylaws; its audit committee charter; the identities of members of the board and each board committee, designating which directors are independent or monitoring (including the organization's definition of "independent"); and the compensation of directors, officers, and the five highest-compensated employees.[268] It should also reveal the most recent five years of audited financial statements, accompanied by the attendant management letters[269] and Forms 990;[270] detailed information about any donor-advised funds that the charity owns, and how and to whom such funds are distributed;[271] prices the entity charges for its services; and any other reports that are statutorily required to be filed by nonprofits with any state or federal agency. The website should also include the outcomes of legal actions concluded in the last five years against the entity, its employees, and its agents, including settlements, deferred prosecution agreements, or any other comparable arrangements. Finally, the

---

268. In universities and hospitals, the highest paid employees are not necessarily the officers; coaches' and surgeons' salaries frequently exceed those of the presidents.

269. The management letter identifies issues that are not required to be disclosed in the annual financial report; it includes concerns and suggestions noted by the auditor during the audit, especially observations or "reportable conditions" deemed pertinent to management's assessment of the company's financial condition and the reliability of its accounting systems. *E.g.*, Young v. Lepone, 305 F.3d 1, 5 & n.3 (1st Cir. 2002) (quoting *Monroe v. Hughes*, 31 F.3d 772, 774 (9th Cir. 1994), for the proposition that "[a] reportable condition is generally regarded as a weakness in the design or operation of the internal control structure that, in the auditor's judgment, reflects a significant shortcoming that 'could adversely affect the organization's ability to record, process, summarize, and report financial data consistent with the assertions of management in the financial statements'").

270. Financial transparency seems the easiest and least controversial. Board members, donors, government agencies, beneficiaries, and lenders all have a legitimate interest in a nonprofit entity's audited financial statements. Many nonprofits' Forms 990 are available on Guidestar, and many nonprofits make their financial statements available upon request or as part of their annual reports. *See* Guidestar, http://www.guidestar.org (last visited Nov. 24, 2007). Guidestar is a nonprofit organization that facilitates nonprofits' transparency by providing a searchable database about nonprofits operating in the United States, based upon IRS Forms 990, the IRS Business Master File, and information volunteered by the organizations. *See* GuideStar—About GuideStar and Philanthropic Research, Inc., http://www.guidestar.org/about/index.jsp?source=dnabout (last visited Nov. 24, 2007); GuideStar—Frequently Asked Questions, http://www.guidestar.org/help/faq.jsp#data (last visited Nov. 24, 2007). A Listening Post Project survey found that 74% of respondents disclose their financial statements upon request; 70% distribute financial statements to funders, and 54% publish financial statements in their annual reports. LESTER M. SALAMON & STEPHANIE L. GELLER, NONPROFIT GOVERNANCE AND ACCOUNTABILITY 5 (2005), http://www.jhu.edu/listeningpost/news/pdf/comm04.pdf. Interestingly, though, only 9% post their financial statements on their own web sites. *Id.*

271. PANEL ON THE NONPROFIT SECTOR, *supra* note 4, at 42.

site should provide a hotline number for reporting concerns about fraud or other illegalities by the entity or its agents, the name and number of the entity's corporate compliance officer, and, if applicable, evidence of current licensure and accreditation status with expiration dates (including information about any aspect of the entity that is on licensure or accreditation probation, suspension, or revocation). The entity should post all reports and statements within thirty days of availability so that viewers can determine the timeliness of completion.

Transparency will contribute significantly to nonprofit public benefit corporations' fulfillment of their ideal role in our society: as entities that can offer opportunities for democratic participation at every level. Such transparency, if sincerely pursued, might positively affect governance simply by virtue of making every act and decision open to public inspection.

V. CONCLUSION

The ongoing move toward board and director independence in the nonprofit sector appears to be a movement without a clear goal, supported by little evidence that independence has accomplished improvements in the business sector. Rather than becoming caught up in the corporate trends of the for-profit realm, reformers should invest more effort in obtaining empirical data about nonprofits' governance and performance. Before enacting legislation, states should consider that government regulation seems to have had little positive effect on nonprofit governance and that, in any case, states have limited ability to enforce such laws. States should strategically choose where to regulate. In doing so, they should close the loopholes that exist in current statutory formulations for strong governance, require audits for entities whose expenses exceed a certain amount, and significantly expand transparency requirements—requirements that should be satisfied, at least in part, by website postings.

The government is not wholly irrelevant to effecting the ethos of a nonprofit's operations. As discussed above, all funders, including government grantors and bonding agencies, can influence governance; for example, they may require audited financial statements, transparency, and self-assessment to determine whether programs are effective and mission-relevant. Nonetheless, externally imposed operational requirements should not allow compliance to overshadow mission and stewardship of assets.

The keys to good governance and the ideal composition of nonprofit boards remain undefined. At minimum, boards should include at least a few directors whom they publicly identify as monitoring directors and should increase their efforts toward transparency. Nonprofits should experiment in diversifying the skill sets and individual perspectives that comprise their boards until they discover the most suitable formulation to meet their twin needs of good governance and community benefit.

GUNS AND GAY SEX: SOME NOTES ON FIREARMS, THE SECOND AMENDMENT, AND "REASONABLE REGULATION"

GLENN HARLAN REYNOLDS*

Professor Adam Winkler has published an interesting article on judicial review and the Second Amendment.[1] Winkler's analysis is useful, and his central point is sound. Winkler observes that even if the individual rights "Standard Model" of Second Amendment interpretation[2] is widely adopted by federal courts, most firearms regulations will withstand judicial scrutiny.[3] In fact, the outcome that he predicts is virtually inevitable for political reasons, regardless of constitutional doctrine. Even without factoring in political pressure, Winkler correctly observes that the individual rights interpretation, conscientiously applied, will permit the vast majority of gun control laws to withstand constitutional scrutiny.[4]

Nonetheless, Winkler's analysis of decisions, mostly under state constitutional rights to arms, omits a key line of state cases. These cases date from the nineteenth to the twenty-first century and shed considerable light on how courts might, and perhaps should, interpret an individual right to arms under the federal Constitution's Second Amendment.

In this short Essay, I will describe this line of cases, their influence on the United States Supreme Court's only twentieth century decision on the Second Amendment,[5] and how they affect Winkler's analysis. I will then suggest some approaches that courts may use when applying the individual right to arms. The result is not entirely inconsistent with Winkler's conclusions, but it provides a more nuanced view. I will conclude with a few thoughts regarding the District of Columbia Circuit's recent individual-rights decision in the Second Amendment case of *Parker v. District of Columbia*[6] and the importance of institutional trust in constitutional interpretation.

---

* Beauchamp Brogan Distinguished Professor of Law, The University of Tennessee College of Law; B.A., The University of Tennessee, 1982; J.D., Yale Law School, 1985.

1. Adam Winkler, *Scrutinizing the Second Amendment*, 105 MICH. L. REV. 683 (2007).

2. *See generally* Glenn Harlan Reynolds, *A Critical Guide to the Second Amendment*, 62 TENN. L. REV. 461 (1995) (explaining that the Standard Model portrays the Second Amendment as conveying upon the people an individual right to own arms).

3. Winkler, *supra* note 1, at 687.

4. *Id.* at 733.

5. United States v. Miller, 307 U.S. 174 (1939).

6. 478 F.3d 370 (D.C. Cir.), *cert. granted*, 128 S. Ct. 645 (2007) (No. 07-290); *see also* Posting of Lyle Denniston to SCOTUSblog, http://www.scotusblog.com/movabletype/archives/2007/07/second_amendmen.html (July 16, 2007, 10:57 EDT) (reporting upon the appeal).

I. AN INDIVIDUAL RIGHT TO ARMS

After Winkler's article addresses a wide variety of state right-to-arms cases, it concludes:

> Forty-two states have constitutional provisions guaranteeing an individual right to bear arms and, tellingly, the courts of every state to consider the question apply a deferential "reasonable regulation" standard in arms rights cases. No state's courts apply strict scrutiny or any other type of heightened review to gun laws. Under the standard uniformly applied by the states, any law that is a "reasonable regulation" of the arms right is constitutionally permissible.[7]

This statement, while not exactly inaccurate, is incomplete. One of the best known and most important lines of state right-to-arms cases does not comfortably fit this characterization.[8] What is more, this line of cases influenced the United States Supreme Court's decision in *United States v. Miller*,[9] the Court's only modern Second Amendment decision to date. These cases come from my home state of Tennessee. I have discussed them at greater length elsewhere,[10] but a short summary will suffice for the purposes of this Essay.

The first case is *Aymette v. State*,[11] a right-to-arms case that actually does not involve guns at all. Aymette was the proud owner of an "Arkansas toothpick," a large and scary knife, which he claimed the right to carry on his person anywhere he chose to go.[12] At the time, the Tennessee constitution provided "[t]hat the free white men of this State have a right to keep and to bear arms for their common defence."[13] According to the court, Mr. Aymette interpreted this provision to give:

> to every man the right to arm himself in any manner he may choose, however unusual or dangerous the weapons he may employ, and, thus armed, to appear wherever he may think proper, without molestation or hindrance, and that any law regulating his social conduct, by restraining the use of any weapon or regulating the manner in which it shall be carried, is beyond the legislative competency to enact, and is void.[14]

---

7. Winkler, *supra* note 1, at 686–87.

8. Andrews v. State, 50 Tenn. 141 (3 Heisk. 165) (1871); Aymette v. State, 21 Tenn. 152 (2 Hum. 154) (1840).

9. 307 U.S. at 178 (citing *Aymette*, 21 Tenn. at 156 (2 Hum. at 158)).

10. Glenn Harlan Reynolds, *The Right to Keep and Bear Arms Under the Tennessee Constitution: A Case Study in Civic Republican Thought*, 61 TENN. L. REV. 647 (1994).

11. 21 Tenn. 152 (2 Hum. 154) (1840).

12. *Id.* at 152–53 (2 Hum. at 155–56).

13. TENN. CONST. of 1834, art. I, § 26.

14. *Aymette*, 21 Tenn. at 153 (2 Hum. at 156).

The Tennessee Supreme Court disagreed. After a review of English and American history and case law,[15] it held that the purpose of the right to arms was to enable the people to resist government tyranny and "to keep in awe those who are in power."[16] The court found that weapons like the Arkansas toothpick did not promote this end.[17] Thus, it held:

> The object, then, for which the right of keeping and bearing arms is secured is the defence of the public. The free white men may keep arms to protect the public liberty, to keep in awe those who are in power, and to maintain the supremacy of the laws and the constitution. The words "bear arms," too, have reference to their military use, and were not employed to mean wearing them about the person as part of the dress. As the object for which the right to keep and bear arms is secured is of general and public nature, to be exercised by the people in a body, for their common defence, so the arms the right to keep which is secured are such as are usually employed in civilized warfare, and that constitute the ordinary military equipment. If the citizens have these arms in their hands, they are prepared in the best possible manner to repel any encroachments upon their rights by those in authority. They need not, for such a purpose, the use of those weapons which are usually employed in private broils, and which are efficient only in the hands of the robber and the assassin. These weapons would be useless in war. They could not be employed advantageously in the common defence of the citizens. The right to keep and bear them is not, therefore, secured by the constitution.
>
> . . . .
>
> The Legislature, therefore, have a right to prohibit the wearing or keeping [of] weapons dangerous to the peace and safety of the citizens, and which are not usual in civilized warfare, or would not contribute to the common defence.[18]

Through this decision in *Aymette*, the Tennessee Supreme Court established: (1) the Tennessee constitution did protect an individual right to arms; (2) that right was intended largely as a protection against tyranny and to effectuate the right of revolution contained elsewhere in the Tennessee constitution; and (3) the arms protected were those of "the ordinary military equipment" and not such weapons as were useful only for crimes and brawling.[19] It also held that the *wearing or carrying* of arms was different from the *keeping and bearing* of arms, with the former being subject to more regulation than the latter.[20]

After the Civil War, the Tennessee constitution was amended to make this point clear. That language remains in the Tennessee constitution today: "[T]he citizens of this State have a right to keep and to bear arms for their common

15. *Id.* at 154–56 (2 Hum. at 156–58).
16. *Id.* at 156 (2 Hum. at 158).
17. *Id.* at 159 (2 Hum. at 161).
18. *Id.* at 156–57 (2 Hum. at 158–59).
19. *Id.*
20. *Id.*

defense; but the Legislature shall have power, by law, to regulate the wearing of arms with a view to prevent crime."[21] Conveniently for our purposes, this produced a new case almost immediately. In the 1871 case of *Andrews v. State*,[22] the defendants challenged a Tennessee gun control law as unconstitutional under the Tennessee constitution's right-to-arms clause.[23] Although the language in the Tennessee constitution had changed, the court's interpretation remained consistent.

The defendants in *Andrews* were charged with violating a statute making it illegal "for any person to publicly or privately carry a dirk, sword-cane, Spanish stiletto, belt or pocket pistol or revolver."[24] In this case, the defendants had a revolver. The *Andrews* court, citing *Aymette*, once again held that the right to keep and bear arms protected such weapons as were part of the ordinary military equipment.[25] The court held that if the defendants' revolver was determined to be the kind ordinarily used by the military, the statute would be unconstitutional as applied to them.[26]

At the same time, the Tennessee Supreme Court rejected the Attorney General's argument that the right to keep and bear arms is a mere "political right" existing for the benefit of the state, and arms thus are subject to unlimited regulation by the state.[27] The court stated:

> In this we think [the Attorney General] fails to distinguish between the nature of the right to keep, and its necessary incidents, and the right to bear arms for the common defense. Bearing arms for the common defense may well be held to be a political right, or for protection and maintenance of such rights, intended to be guaranteed; but the right to keep them, with all that is implied fairly as an incident to this right, is a private individual right, guaranteed to the citizen, not the soldier.[28]

---

21. TENN. CONST. art. I, § 26.

22. 50 Tenn. 141 (3 Heisk. 165) (1871).

23. *Id.* at 143 (3 Heisk. at 167).

24. *Id.* at 146–47 (3 Heisk. at 171) (quoting Act of June 11, 1870, Tenn. Pub. Acts 28).

25. *Id.* at 153–54 (3 Heisk. at 179–80).

26. *Id.* at 159–60 (3 Heisk. at 186–87). The Tennessee Supreme Court remanded the case to the trial court for "evidence as to what character of weapon is included in the designation 'revolver.'" *Id.*

27. *Id.* at 154–55 (3 Heisk. at 180–81).

28. *Id.* at 156 (3 Heisk. at 182). The court also found certain penumbral aspects to the right to keep and bear arms:

> The right to keep arms, necessarily involves the right to purchase them, to keep them in a state of efficiency for use, and to purchase and provide ammunition suitable for such arms, and to keep them in repair. And clearly for this purpose, a man would have the right to carry them to and from his home, and no one could claim that the Legislature had the right to punish him for it, without violating this clause of the Constitution.
>
> But farther than this, it must be held, that the right to keep arms involves, necessarily, the right to use such arms for all the ordinary purposes, and in all the ordinary modes usual

Yet the court pointed out that the right to keep and bear arms was distinct from the right to carry them:

> It is insisted, however, by the Attorney General, that, if we hold the Legislature has no power to prohibit the wearing of arms absolutely, and hold that the right secured by the Constitution is a private right, and not a public political one, then the citizen may carry them at all times and under all circumstances. This does not follow by any means, as we think.
>
> While the private right to keep and use such weapons as we have indicated as arms, is given as a private right, its exercise is limited by the duties and proprieties of social life, and such arms are to be used in the ordinary mode in which used in the country, and at the usual times and places. Such restrictions are implied upon their use as are thus indicated.
>
> . . . .
>
> . . . If the citizen is possessed of a horse, under the Constitution it is protected and his right guaranteed, but he could not, by virtue of this guaranteed title, claim that he had the right to take his horse into a church to the disturbance of the people; nor into a public assemblage in the streets of a town or city, if the Legislature chose to prohibit the latter and make it a high misdemeanor.
>
> The principle on which all right to regulate the use in public of these articles of property, is, that no man can so use his own as to violate the rights of others, or of the community of which he is a member.
>
> So we may say, with reference to such arms, as we have held, he may keep and use in the ordinary mode known to the country, no law can punish him for so doing, while he uses such arms at home or on his own premises; he may do with his own as he will, while doing no wrong to others. Yet, when he carries his property abroad, goes among the people in public assemblages where others are to be affected by his conduct, then he brings himself within the pale of public regulation, and must submit to such restriction on the mode of using or carrying his property as the people through their Legislature, shall see fit to impose for the general good.[29]

Andrews suggests that the carrying of weapons in public is subject to "reasonable regulation," but regulation of the ownership, normal repair, practice with, and transport of weapons, as well as the purchase of ammunition, is subject to a higher degree of scrutiny.[30] The *Andrews* court found that the purpose of arming the citizenry against a potentially tyrannical federal government underlies both the Tennessee constitution's right-to-arms provision and the Second Amendment to the United States Constitution.[31] Regulation

---

> in the country, and to which arms are adapted, limited by the duties of a good citizen in times of peace; that in such use, he shall not use them for violation of the rights of others, or the paramount rights of the community of which he makes a part.

Id. at 153 (3 Heisk. at 178–79).

29. *Id.* at 155–56, 158–59 (3 Heisk. at 181–82, 185–86).

30. *Id.* at 153 (3 Heisk. at 178–79).

31. *Id.* at 151–53 (3 Heisk. at 177–78).

must not unreasonably chill this purpose. The court's analysis is generally in accordance with the maxim, popular among nineteenth century courts, of *sic utere tuo ut alienum non laedas* ("you should use what is yours so as not to harm what is others'").[32] Even the reasonable regulation of weapons *carrying* has some limitations:

> It is insisted by the Attorney General, as we understand his argument, that this clause confers power on the Legislature to prohibit absolutely the wearing of all and every kind of arms, under all circumstances. To this we can not give our assent. *The power to regulate, does not fairly mean the power to prohibit; on the contrary, to regulate, necessarily involves the existence of the thing or act to be regulated. . . .*
>
> But the power is given to regulate, with a view to prevent crime. The enactment of the Legislature on this subject, must be guided by, and restrained to this end, and bear some well defined relation to the prevention of crime, or else it is unauthorized by this clause of the Constitution.[33]

In other words, prohibition is not regulation. In addition, regulation based solely on the legislature's (or even a substantial sector of the public's) general dislike of guns and gun owners would not be justified. Regulation must be supported by a "well defined relation to the prevention of crime," not simply by some voting bloc's prejudice.[34]

These are old cases, but they remain good law. Indeed, the language from *Aymette* about "the ordinary military equipment"[35] was quoted by the United States Supreme Court in *United States v. Miller*,[36] which suggests that the Supreme Court found this line of cases to be at least relevant to analysis under the federal Constitution's Second Amendment.

A twenty-first century case in this line, *Stillwell v. Stillwell*,[37] illustrates how courts apply these cases today. This case also sheds further light on the reasonable regulation of firearms and its proper sphere. *Stillwell v. Stillwell*, as the name suggests, involved a divorced couple.[38] The former wife sought and received an order from the family court that barred her ex-husband from

---

32. *See* Glenn H. Reynolds & David B. Kopel, *The Evolving Police Power: Some Observations for a New Century*, 27 HASTINGS CONST. L.Q. 511, 511–12 (2000); ERNST FREUND, THE POLICE POWER 6 (Arno Press 1976) (1904); *see also id.* at 60–61 ("Effective judicial limitations of the police power would be impossible, if the legislature were the sole judge of the necessity of the measures it enacted. . . . The question of reasonableness usually resolves itself into this: [I]s regulation carried to the point where it becomes prohibition, destruction, or confiscation?").

33. *Id.* at 154–55 (3 Heisk. at 180–81) (emphasis added).

34. *Id.* at 155 (3 Heisk. at 181).

35. Aymette v. State, 21 Tenn. 152, 156 (2 Hum. 154, 158) (1840).

36. 307 U.S. 174, 178 (1939).

37. No. E2001-00245-COA-R3-CV, 2001 WL 862620, at *1 (Tenn. Ct. App. July 30, 2001).

38. *Id.* at *1.

carrying a gun or from having any firearms in the house that were not locked up when their child was visiting.[39] Many would regard this order as a reasonable regulation.

The Tennessee Court of Appeals, though, found that *both* Tennessee's right to arms and its right to privacy granted the ex-husband a fundamental right to possess arms.[40] This is significant given that the right to privacy undoubtedly constitutes a fundamental right under Tennessee law.[41] The court stated:

> We certainly cannot overemphasize the need for extreme caution with firearms at all times, especially when children are or may be present. Nevertheless, absent a showing of risk of substantial harm to the child, we conclude that the portion of the Trial Court's order restricting Father's possession of a firearm in the presence of his child was in error, and vacate that portion of the Trial Court's order. Absent a risk of substantial harm to the child, the wisdom of Father's decision is not for the Trial Court or this Court to determine. The Trial Court made no finding of risk of substantial harm, and neither can we based upon the record before us.[42]

The key point is that "absent a risk of substantial harm to the child," the "wisdom" of a father's decision is "not for the Trial Court or this Court to determine."[43] While this standard may certainly be characterized as a "reasonable regulation" standard, it is hardly deferential to the regulation of firearms.[44]

Further, the *Stillwell* court's linkage of the right of privacy and the right to bear arms[45] is interesting for reasons that go beyond the facts of this case. *Stillwell* suggests that the standards that Tennessee courts have applied to regulation in privacy cases likely also would provide useful standards for courts puzzling over the nature of "reasonable regulation" in the firearms context.

II. THE TENNESSEE PRIVACY CASES AND REASONABLE REGULATION

Although both appear to protect a fundamental right, the Tennessee right to arms and the Tennessee right of privacy differ in one respect: The right to arms has a clear textual basis,[46] while the right of privacy is a judicial construct,

---

39. *Id.* at *2.

40. *Id.* at *4.

41. *See infra* Part II.

42. *Stillwell*, 2001 WL 862620, at *4.

43. *Id.*

44. And, again, it is consistent with the *sic utere* principle that reasonable regulation requires harm to others. *See supra* note 32 and accompanying text.

45. *Stillwell*, 2001 WL 862620, at *4 ("[W]e believe the constitutional rights under the Second Amendment of the United States Constitution as well as Article I, Section 26 of the Tennessee Constitution are worthy of the same protection as is the constitutional right to privacy . . . .").

46. TENN. CONST. art. I, § 26.

based on penumbral reasoning.[47] Tennessee's right of privacy has its roots in several clauses, most notably article I, section 8 of the Tennessee constitution[48]—which does the work of the federal Due Process Clause—and article I, sections 1 and 2. These two sections of Tennessee's Declaration of Rights provide that all governmental power stems from the people and that the people have the right (perhaps even the duty) to rebel against a government that is arbitrary and oppressive:[49]

> Sec. 1. All power inherent in the people—Government under their control. That all power is inherent in the people, and all free governments are founded on their authority, and instituted for their peace, safety, and happiness; for the advancement of those ends they have at all times, an unalienable and indefeasible right to alter, reform, or abolish the government in such manner as they may think proper.[50]

> Sec. 2. Doctrine of nonresistance condemned. That government being instituted for the common benefit, the doctrine of non-resistance against arbitrary power and oppression is absurd, slavish, and destructive of the good and happiness of mankind.[51]

A series of cases not involving arms have highlighted the absurdity of interpreting the Tennessee constitution as empowering the passage of arbitrary and oppressive laws, given that these provisions establish a right of revolt against arbitrary power and oppression.[52] Both the Tennessee right of privacy and the right to arms are grounded in the right of revolution.[53] It would be a

---

47. *See generally* Brannon P. Denning & Glenn Harlan Reynolds, Essay, *Comfortably Penumbral*, 77 B.U. L. REV. 1089 (1997) (discussing the concept of penumbral reasoning and its role in constitutional jurisprudence); Glenn Harlan Reynolds, *Penumbral Reasoning on the Right*, 140 U. PA. L. REV. 1333 (1992) (arguing that penumbral reasoning is not a strictly liberal doctrine and frequently has been used by conservative judges).

48. This is often referred to as the "law of the land" clause.

49. This right is rather explicit, but in case there are any doubts, the Tennessee Supreme Court has made clear that it interprets the Declaration of Rights this way. *See* Davis v. Davis, 842 S.W.2d 588, 599 (Tenn. 1992) ("Indeed, the notion of individual liberty is so deeply embedded in the Tennessee Constitution that it, alone among American constitutions, gives the people, in the face of governmental oppression and interference with liberty, the right to resist that oppression even to the extent of overthrowing the government."); *see also* Otis H. Stephens, Jr., *The Tennessee Constitution and the Dynamics of American Federalism*, 61 TENN. L. REV. 707, 710 (1994) (stating that these provisions "clearly assert the right of revolution"); *cf.* Cravens v. State, 256 S.W. 431, 432 (Tenn. 1923) (emphasizing the importance of retaining a spirit of resistance against despotism).

50. TENN. CONST. art. I, § 1.

51. TENN. CONST. art. I, § 2.

52. *See, e.g.*, *Davis*, 842 S.W.2d at 599; *see also, e.g.*, Stillwell v. Stillwell, No. E2001-00245-COA-R3-CV, 2001 WL 862620, at *4 (Tenn. Ct. App. July 30, 2001).

53. *Compare Davis*, 842 S.W.2d at 599 ("in the face of governmental oppression and interference with liberty, [the Tennessee constitution gives] the right to resist that oppression

strange constitution indeed that empowered the government to behave in ways that would justify a revolution.[54]

Perhaps the clearest example of this formulation came in *Campbell v. Sundquist*,[55] a case striking down Tennessee's law against homosexual sodomy as a violation of the Tennessee privacy right.[56] Essentially, the state contended that the sodomy statute constituted a reasonable regulation of sexual behavior:

> First, the [Homosexual Practices] Act discourages activities which cannot lead to procreation. Second, the Act discourages citizens from choosing a lifestyle which is socially stigmatized and leads to higher rates of suicide, depression, and drug and alcohol abuse. Third, the Act discourages homosexual relationships which are "short lived," shallow, and initiated for the purpose of sexual gratification. Fourth, the Act prevents the spread of infectious disease, and fifth, the Act promotes the moral values of Tennesseans.[57]

The court found these justifications for the regulation unpersuasive, suggesting that they not only failed to constitute the kind of compelling interest needed to satisfy strict scrutiny, but they in fact failed even to satisfy the more relaxed rational-basis test.[58] Absent a showing of any substantial risk of harm to others, the court concluded that mere public disapproval of homosexual activities, culture, or lifestyle could not justify the ban:

> It may be asked whether a majority, believing its own happiness will be enhanced by another's conformity, may not enforce its moral code upon all.[59] ... The threshold question in determining whether the statute in question is a valid exercise of the police power is to decide whether it benefits the public generally. The state clearly has a proper role to perform in protecting the public from inadvertent offensive displays of sexual behavior, in preventing people from being forced against their will to submit to sexual contact, in protecting minors from being sexually used by adults, and in eliminating cruelty to animals. To assure these protections, a broad range of criminal statutes constitute valid police power exercises, including proscriptions of indecent exposure, open lewdness, rape, involuntary deviate sexual

---

even to the extent of overthrowing the government"), *with* Aymette v. State, 21 Tenn. 152, 156 (2 Hum. 154, 158) (1840) ("The free white men may keep arms to protect the public liberty, to keep in awe those who are in power, and to maintain the supremacy of the laws and the constitution.").

54. While this notion may seem odd today, it seemed less so during the first century of the nation's existence. *See* Reynolds & Kopel, *supra* note 32, at 511.

55. 926 S.W.2d 250 (Tenn. Ct. App. 1996).

56. *Id.* at 266. This case was decided before the U.S. Supreme Court held in *Lawrence v. Texas*, 539 U.S. 558 (2003), that sodomy laws are unconstitutional.

57. *Campbell*, 926 S.W.2d at 262.

58. *See id.* at 262–65.

59. *Id.* at 265 (quoting Commonwealth v. Wasson, 842 S.W.2d 487, 502–03 (Ky. 1992) (Combs, J., concurring)).

> intercourse, indecent assault, statutory rape, corruption of minors, and cruelty to animals. The statute in question serves none of the foregoing purposes and it is nugatory to suggest that it promotes a state interest in the institution of marriage.[60]

As in *Stillwell*, the absence of a substantial, as opposed to a theoretical, risk of harm to others seems to have been the key failing in the statute.[61] This same reasoning could apply to gun control laws. Winkler seems to think that almost any gun control law could be characterized as preventing harm,[62] but upon closer inspection, this is not so clear.

III. RETHINKING WINKLER'S "REASONABLE REGULATION"

The cases above, both gun-related and otherwise, have a common thread: a sort of cost-benefit analysis where courts presume the benefits of the rights, while they look skeptically upon the purported costs that provide the underlying justification for regulation. The mere presence of guns in *Stillwell* was not evidence of enough risk to outweigh the right to bear arms and to raise children without state interference.[63] Likewise, in *Campbell*, claims that homosexual sodomy posed a risk of AIDS and other infectious disease received no particular deference:

> We agree that the State certainly has a compelling interest in preventing the spread of infectious disease among its citizens, however, the Homosexual Practices Act is not narrowly tailored to advance this interest. The statute prohibits all sexual contact between people of the same gender even if the people involved are disease free, practicing "safe sex," or engaging in sexual contact which does not contribute to the spread of disease.[64]

Factoring these important cases into Winkler's analysis, "reasonable regulation" analysis must protect against the tendency of legislatures to seek *un*reasonable regulation for reasons of political prejudice or irrational fear.[65]

---

60. *Id.* at 265 (quoting Commonwealth v. Bonadio, 415 A.2d 47, 49–50 (Pa. 1980)).

61. *Id.* at 263; *see* Stillwell v. Stillwell, No. E2001-00245-COA-R3-CV, 2001 WL 862620, at *4 (Tenn. Ct. App. July 30, 2001).

62. "[G]un laws are generally motivated by legitimate public safety concerns rather than invidious purposes . . . ." Winkler, *supra* note 1, at 727.

63. *Stillwell*, 2001 WL 862620, at *4.

64. *Campbell*, 926 S.W.2d at 263.

65. As this analysis suggests, Robert Bork's notorious "equal gratifications" argument is insufficient to support regulation of contraceptive use absent some tangible harm. *See generally* Robert H. Bork, *Neutral Principles and Some First Amendment Problems*, 47 IND. L.J. 1 (1971) (discussing his theory in detail). Tennessee courts, at least, do not see Bork's argument as having more force where firearms are concerned. *See* Glenn Harlan Reynolds, *Sex, Lies and Jurisprudence: Robert Bork, Griswold, and the Philosophy of Original Understanding*, 24 GA. L. REV. 1045, 1069–70 (1990) (discussing Bork and regulations based on aesthetics or morality

Courts must distinguish between the reasonable and the unreasonable, as the Tennessee courts have done.[66]

The divergence between the holdings in the cases discussed in this Essay and those discussed by Winkler illustrates the comments of Judge Alex Kozinski in *Silveira v. Lockyer*: "Judges know very well how to read the Constitution broadly when they are sympathetic to the right being asserted."[67] One might add that judges are demonstrably less willing to read the Constitution broadly when they are unsympathetic to the right being suppressed, which might include interpretation of the right to arms. While this brief Essay cannot conclusively settle the issue, below are some thoughts on how a "reasonable regulation" approach might work in the context of more thorough Second Amendment analysis.

First, "reasonable regulation" often can be used to cover the true intentions of regulators who actually intend to extinguish or seriously undermine the right at issue. Courts are rightly suspicious of such possibilities in the context of other rights, such as free speech, abortion, sodomy, birth control, or the dormant commerce clause. To ensure that hostile authorities cannot bypass constitutional protections merely by asserting a public safety justification, the courts have employed various presumptions, tests, and simple judicial skepticism. Judicial review of laws and regulations governing the right to keep and bear arms should invoke the same degree of skepticism, rather than allowing judges to credulously accept that any law regulating guns is inherently a law intended to promote "public safety." Other rights—such as free speech, the right against self-incrimination, or the right to privacy, to name a few—have costs as well as benefits. Likewise, the right to arms may have costs, in terms of limiting how far some "public safety" regulations can go, along with its benefits. As seen in the Tennessee cases described above, courts must second-guess safety justifications where such second-guessing is necessary to protect a textually secured right.[68]

Second, if properly interpreted, the right to arms is less individualistic than some other rights. It is not less individualistic in the sense of the individual versus the collective, but rather, it is less individualistic in the sense that it is more instrumental than expressive. At both the federal and state levels, the right to arms stems from concerns about self defense and the defense of public liberty.[69] Thus, while regulations of speech that turn on questions of style and

---

rather than tangible harm).

66. *See* Andrews v. State, 50 Tenn. 141 (3 Heisk. 165) (1871); Aymette v. State, 21 Tenn. 152 (2 Hum. 154) (1840); *Stillwell*, 2001 WL 862620, at *1.

67. Silveira v. Lockyer, 328 F.3d 567, 568 (9th Cir. 2003) (Kozinski, J., dissenting from denial of rehearing en banc).

68. *See Andrews*, 50 Tenn. at 158–59 (3 Heisk. at 185–86); *Stillwell*, 2001 WL 862620, at *4.

69. Indeed, as Don Kates has demonstrated, the Framers saw these two functions as one and the same. Don B. Kates, Jr., *The Second Amendment and the Ideology of Self-Protection*, 9 CONST. COMMENT. 87, 92 (1992).

aesthetics may still violate the First Amendment's free speech guarantee, which in modern conception is about individual expression, the Second Amendment's right to arms is about capabilities more than expression. For example, a ban on characteristics of guns that make them look "too military" without impairing their actual function might not violate the Second Amendment, but similar limitations on expressive characteristics in the area of film or television might violate the First Amendment.

Finally, judicial latitude may prove costly if overexercised. The public is aware of the truism, noted above, that was articulated by Judge Kozinski when he stated that judges are more likely to read the Constitution broadly when they agree with the ultimate conclusion.[70] Americans rightly expect their courts to treat constitutional rights with a degree of respect. Since the Second Amendment has a clear textual basis and a broad base of popular support, cavalier treatment of this right would likely prove costly to the public image of the judicial system as a whole. Though the term "reasonable regulation" might be stretched to encompass a wide array of intrusive limitations on that right, this stretching should not exceed limits that might be thought reasonable by the public, even if those limits do not command the sympathies of judges and politicians.

When comparing the holdings of cases in which judges are sympathetic to the right in question to the holdings of cases in which they are not, the inequality becomes apparent. Courts will face difficulty in justifying a more sympathetic treatment of rights that lack textual support in the Constitution than of the right to arms, which is specifically enumerated in the Bill of Rights. Professor Mike O'Shea makes that point with regard to the case of *Parker v. District of Columbia*:[71]

> It's not often that the Supreme Court takes up the core meaning of an entire Amendment of the Bill of Rights, in a context where it writes on a mostly clean slate from the standpoint of prior holdings. If the Court takes the case, then October Term 2007 becomes The Second Amendment Term. *Parker* would swiftly overshadow, for example, the Court's important recent cert grant in the Guantanamo cases.
>
> How many Americans would view *District of Columbia v. Parker* as the most important court case of the last thirty years? The answer must run into seven figures. The decision would have far-reaching effects, particularly in the event of a reversal.
>
> . . . .
>
> . . . [T]here is a way more straightforward comparison that a whole lot of average Americans would be making. That's a comparison between the Court's handling of the *enumerated* rights claim at issue in *Parker*, and its demonstrated willingness to embrace even *non*-enumerated individual rights that are congenial to the political left, in cases like *Roe* and *Lawrence*. "So

70. *Silveira*, 328 F.3d at 568 (Kozinski, J., dissenting from denial of rehearing en banc).

71. 478 F.3d 370 (D.C. Cir.), *cert. granted*, 128 S. Ct. 645 (2007) (No. 07-290).

> the Constitution says *Roe,* but it doesn't say I have the right to keep a gun to defend my home, huh?"[72]

This difficulty is troubling and potentially politically explosive. The faith of the public is especially important to the federal judiciary because it is a branch of government that possesses neither the sword nor the purse. If federal courts are to retain this faith, the public must see them as faithfully obeying the commands of the Constitution. Though the public generally pays limited attention to most legal issues, Professor O'Shea is correct that cases like *Parker* will receive considerably more scrutiny.[73] Once the courts fall under the public eye, the way they handle the reasonableness portion of "reasonable regulation" will be particularly important.

We should expect courts to treat the regulation of gun ownership with the same skepticism previously applied to the regulation of gay sex[74] and communist propaganda.[75] In some sense, this proposition may seem unnerving to both gun rights supporters and opponents. Still, if Judge Kozinski's comment is not to become an epitaph for the legitimacy of judicial review, such an expectation is necessary.

---

72. Posting of Mike O'Shea to Concurring Opinions, http://www.concurringopinions.com/archives/2007/07/the_second_amen_1.html (July 16, 2007, 19:10 EDT).

73. *Id.*

74. *See, e.g.*, Campbell v. Sundquist, 926 S.W.2d 250, 263 (Tenn. Ct. App. 1996); *see also, e.g.*, Lawrence v. Texas, 539 U.S. 558, 577–78 (2003).

75. *See* Lamont v. Postmaster General, 381 U.S. 301, 308–09 (1965) (Brennan, J., concurring); *see also id.* at 305–07 (majority opinion).

FAMILY LAW—PARENTAL RIGHTS—PROTECTION OF PARENTAL RIGHTS IN CUSTODY AND TERMINATION OF PARENTAL RIGHTS CASES IN TENNESSEE

In re Adoption of A.M.H., 215 S.W.3d 793 (Tenn. 2007)

I. INTRODUCTION

A.M.H. was born on January 28, 1999 in Memphis, Tennessee to a couple that had recently emigrated from China.[1] Knowing that they would be unable to care for their newborn child, the parents petitioned the Shelby County juvenile court for temporary foster care.[2] The juvenile court arranged for a local adoption agency to provide A.M.H. with three months of foster care,[3] and on February 24, 1999, the parents signed an "interim care agreement" with the agency to place A.M.H. into foster care at the appellees' home.[4] The agreement signed by the parents specified that it did not terminate their parental rights, and they reached a separate oral agreement with the foster parents allowing them to visit the child once a week at the foster parents' residence.[5]

After placing the child in foster care, the parents frequently visited her, "bringing food and gifts and taking photographs" each time.[6] However, after experiencing increased financial difficulties, the parents expressed that they wished to send their daughter to China to live with relatives.[7] Opposed to the removal of A.M.H. from their home, the foster parents reached an oral agreement with the parents whereby A.M.H. would remain in the foster parents' care until she reached the age of majority and the biological parents

---

1. *In re* Adoption of A.M.H., 215 S.W.3d 793, 797 (Tenn. 2007). A.M.H.'s father was on a student visa and enrolled in a doctoral program at the University of Memphis where he had obtained a graduate assistantship. *Id.* A.M.H.'s mother was unmarried to A.M.H.'s father at the time of her immigration; she nonetheless obtained a visa as the father's wife. *Id.* A.M.H.'s mother spoke very little English and the court used interpreters throughout the proceedings. *Id.*

2. *Id.* Just before A.M.H.'s birth, a student at the University of Memphis filed a complaint against the father alleging attempted rape. *Id.* As a result, the university fired the father from his position as a graduate assistant, leaving the parents of A.M.H. with virtually no income and no health insurance. *Id.*

3. *Id.* at 798.

4. *Id.* at 797-98.

5. *Id.* at 798.

6. *Id.*

7. *Id.* In April of 1999, the father was arrested on the attempted rape charge, and the University of Memphis fired him from the part-time position he had obtained there following the termination of his graduate assistantship. *Id.* A jury acquitted the father in February 2003. *Id.* at 804.

would retain their parental rights.[8] Following this oral agreement, the foster parents sought legal custody of A.M.H., and on June 4, 1999, the biological and foster parents signed a "Petition for Custody" and a "Consent Order Awarding Custody" prepared by a juvenile court officer.[9] Included in the consent order was a guardianship provision enabling the foster parents to obtain health insurance for A.M.H.[10] Throughout the custody proceedings, A.M.H.'s mother was insistent that the custody order was only temporary.[11]

The relationship between A.M.H.'s biological parents and her foster parents subsequently became strained, and after less than six months, the biological parents expressed to a juvenile court officer that they wished to regain custody.[12] In May 2000, they returned to juvenile court where they alleged a change in circumstances and sought custody of A.M.H.[13] After a hearing, the juvenile court denied their petition for modification of the custody agreement in June 2000.[14]

In January 2001, an altercation occurred between A.M.H.'s parents and the foster parents regarding visitation, and the foster parents called the police to resolve the dispute.[15] The police officer instructed the biological parents to leave the foster parents' home.[16] Thereafter, they never attempted to visit or otherwise maintain a relationship with A.M.H., but they wrote to the juvenile court just eighteen days after the incident and returned to the court in April 2001 in an attempt to regain custody.[17] On June 20, 2001, exactly four months and five days after the parents last visited A.M.H., the foster parents filed a

---

8. *Id.* at 798.

9. *Id.* at 798-99.

10. *Id.* at 799.

11. *Id.* The foster parents, the biological father, and the adoption agency's attorney met prior to the proceedings to discuss the legal consequences of transferring custody. *Id.* at 798. The attorney for the adoption agency testified that he informed A.M.H.'s father that "'unless everybody consents to give the custody back . . . anybody that gives up even temporary custody takes a risk that . . . the court may not give custody back." *Id.* at 798-99. A.M.H.'s biological mother was not present at the meeting. *Id.* at 798.

12. *Id.* at 800. At trial, A.M.H.'s mother testified through an interpreter that she felt "tricked" into signing the custody order, and the father testified that he felt "intimidated and threatened" by the foster father. *Id.*

13. *Id.*

14. *Id.* at 801. The Court Appointed Special Advocate (CASA) recommended at the hearing that the foster parents retain custody and that the biological parents be allowed supervised visitation twice a week for four hours each visit. *Id.*

15. *Id.*

16. *Id.* The record is unclear as to what exactly the police officer told A.M.H.'s parents. The officer testified that he would have told them not to return to the foster parents' residence "that day;" however, the foster parents' answers to interrogatories stated that the parents were not ever allowed to return to their home. *Id.*

17. *Id.* at 801-02.

petition for adoption and termination of parental rights on abandonment grounds.[18]

On February 23, 2004, the trial on the petitions for custody, adoption, and termination of parental rights began.[19] Several experts testified for both sides, offering psychological assessments of A.M.H. and her biological parents.[20] The chancery court ruled that the parents had abandoned A.M.H. by willfully failing to visit or to provide financial support for her in the four months immediately preceding the foster parents' petition for termination of parental rights.[21] On those grounds, the court held that it would be in the child's best interests to terminate parental rights and to allow her to remain with her foster parents.[22] The court denied the biological parents' petition for modification of the custody arrangement and held the foster parents' petition for adoption in abeyance.[23]

The Tennessee Court of Appeals reversed the chancery court's holding that A.M.H.'s parents had abandoned her by willfully failing to support her; however, it affirmed the termination of parental rights on the grounds that they had willfully failed to visit her for the four consecutive months preceding the foster parents' petition.[24] The court also held that the termination of parental rights was in the best interests of A.M.H.[25] On appeal to the Supreme Court of Tennessee, *held*, reversed.[26] A parent who is actively pursuing custody of a child through legal recourse cannot be said to have willfully abandoned that child. Further, a parent must have knowledge of the consequences of a custody transfer in order to render it legally binding. *In re Adoption of A.M.H.*, 215 S.W.3d 793 (Tenn. 2007).

In re A.M.H. presented the Supreme Court of Tennessee with two core issues regarding parental rights.[27] The first concerned whether the parents of

---

18. *Id.* The petition, which sparked chancery proceedings spanning thirty-two months, claimed that A.M.H.'s parents had abandoned her by willfully failing to visit and willfully failing to financially support her. *Id.* at 803. The petition was later amended to include additional grounds for termination, including lack of paternity, which was later proven false by a DNA test. *Id.* The other alleged grounds for termination—the parent's mental incompetence, and the persistence of conditions preventing A.M.H.'s reunification with her parents—were addressed by psychologists at trial. *Id.* at 803, 805.

19. *Id.* at 804.

20. *Id.* at 805-06.

21. *Id.* at 806. The chancery court based its ruling on abandonment and denied the other grounds for termination offered by the foster parents. *Id.* at 806-07.

22. *Id.* at 806. The court found that the parents were "manipulative and dishonest people who appeared to have no intent to raise A.M.H. but have used the child from birth for financial gain and to avoid deportation." *Id.*

23. *Id.* at 807.

24. *Id.* Justice Kirby dissented, stating that she would reverse the termination of parental rights because the failure to visit was not willful. *Id.*

25. *Id.*

26. *Id.* at 813.

27. *Id.* at 796. A third core holding in the case dealt with jurisdiction. The foster parents

A.M.H. had abandoned her by willfully failing to visit her under the Tennessee statutory provision for termination of parental rights.[28] The second issue was whether the foster parents could retain custody and guardianship of A.M.H. under the consent order if the biological parents' rights were not terminated.[29] As to the first issue, the court held that even though the parents did not visit A.M.H. for four months and were not legally prevented from doing so, because they attempted to regain custody through legal recourse, their failure to visit could not constitute a "willful failure to visit" as grounds for abandonment.[30] Moving to the second issue, the court held that because A.M.H.'s biological parents were unaware of the consequences of the custody order, their consent to the change in custody was invalid.[31] The court never assessed the best interests of the child in its analysis, nor did it directly address whether the biological parents had shown a significant change in their circumstances to warrant revoking custody from the foster parents.[32] Rather, the court asserted the parental "superior rights doctrine" and returned the child to her birth parents more than seven years after she first had been placed in the custody of her foster parents.[33]

---

argued that the Supreme Court of Tennessee did not have jurisdiction to hear the case on account of the statute of repose provision under Tennessee Code Annotated section 36-1-113, which states:

> After the entry of the order terminating parental rights, no party to the proceeding, nor anyone claiming under such party, may later question the validity of the termination proceeding by reason of any defect or irregularity therein, jurisdictional or otherwise, but shall be fully bound thereby, except based upon a timely appeal of the termination order as may be allowed by law; and in no event, for any reason, shall a termination of parental rights be overturned by any court or collaterally attacked by any person or entity after one (1) year from the date of the entry of the final order of termination. This provision is intended as a statute of repose.

TENN. CODE ANN. § 36-1-113(q) (2005). The appellees argued that because the appeal was not completed within one year, the court did not have jurisdiction to hear the case. *In re A.M.H.*, 215 S.W.3d at 807-08. Applying principles of statutory construction, the court determined that the statute was unambiguous and held that a statute of repose "does not limit the time for appellate courts to hear and rule on a case that has been appealed timely;" rather, it "limits the time within which an action may be filed." *Id.* at 808. In this case, the biological parents made a timely appeal and therefore, the statue did not apply to bar the appeal. *Id.*

28. *Id.* at 809; *see* TENN. CODE ANN. §36-1-113(g)(1) (2005) (providing that abandonment is a ground for termination of parental rights); *id.* § 36-1-102(1) (defining abandonment as willful failure to visit for four consecutive months). The court addressed this issue as a question of law, which is reviewed "de novo with no presumption of correctness." *In re A.M.H.*, 215 S.W.3d at 810.

29. *Id.* at 811.

30. *Id.* at 810-11.

31. *Id.* at 811-12.

32. *See id.* at 811-13. The court merely indicated that "[t]he evidence at trial showed that the parents ha[d] overcome many obstacles to achieve financial stability." *Id.* at 813.

33. *Id.* at 812-13.

II. PROTECTION OF PARENTAL RIGHTS IN TENNESSEE

Parental rights are "a parent's rights to make all decisions concerning his or her child, including the right to determine the child's care and custody . . . ."[34] The importance of parental rights is "deeply rooted in our nation's history,"[35] and thus, parental rights have long been afforded constitutional protection under the Fourteenth Amendment of the United States Constitution.[36] In addition to the federal jurisprudential protections of parental rights, the state of Tennessee recognizes that parental rights embody a fundamental liberty interest protected by the Tennessee Constitution.[37] The Supreme Court of Tennessee has held that the state constitution "give[s] natural parents a presumption of 'superior parental rights' regarding the custody of their children."[38]

A. Protection of Parental Rights in Termination of Parental Rights Proceedings in Tennessee

Although parents are afforded rights to the custody of their children that are superior to those of non-parents, parental rights are not absolute and are "subject to limitation . . . if it appears that the parental decisions will jeopardize the health or safety of the child, or have a potential for significant social burdens."[39] Yet, because termination of parental rights constitutes the deprivation of a constitutionally protected liberty, petitioners attempting to terminate the parental rights of others must also prove by clear and convincing evidence that the statutory grounds for termination exist.[40] In some cases, the Department of Children's Services must prove by clear and convincing evidence that it made reasonable efforts to reunite the child with the biological parent.[41] The Department's failure to provide reasonable opportunities for

---

34. BLACK'S LAW DICTIONARY 521 (8th ed. 2005).

35. 32 AM. JUR. PROOF OF FACTS 3D *Parental Rights* § 3 (2007) (citing Bellotti v. Baird, 443 U.S. 622 (1979)).

36. *See* Santosky v. Kramer, 455 U.S. 745, 753 (1982) (citing Quilloin v. Walcott, 434 U.S. 246, 255 (1978); Smith v. Org. of Foster Families, 431 U.S. 816, 845 (1977)). The United States Supreme Court stated that the Due Process Clause protects "a fundamental liberty interest of natural parents in the care, custody, and management of their child . . . ." *Id.*

37. TENN. CONST. art. I, § 8 n.9; *see also* Hawk v. Hawk, 855 S.W.2d 573, 579 (Tenn. 1993) ("Tennessee's historically strong protection of parental rights and the reasoning of federal constitutional cases convince [the court] that parental rights constitute a fundamental liberty interest under Article I, Section 8 of the Tennessee Constitution.") .

38. Blair v. Badenhope, 77 S.W.3d 137, 141 (Tenn. 2002).

39. *Hawk*, 855 S.W.2d at 578 (quoting Wisconsin v. Yoder, 406 U.S. 205, 233-34 (1972)); *see* Tracey B. Harding, Note, *Involuntary Termination of Parental Rights: Reform is Needed*, 39 BRANDEIS L.J. 895 (2001).

40. *Santosky*, 455 U.S. at 769; 32 AM. JUR. PROOF OF FACTS 3D *Parental Rights* § 3 (2007).

41. *In re* M.B., No. M2006-02063-COA-R3PT, 2007 WL 1034676, at *3-4 (Tenn. Ct. App. Mar. 30, 2007).

parental rehabilitation renders any termination of parental rights invalid.[42] In addition to establishing the grounds for termination by clear and convincing evidence, Tennessee law also mandates that any termination of parental or guardianship rights must be in the best interests of the child.[43]

Tennessee Code Annotated section 36-1-113(g) sets forth several grounds for termination, including "abandonment by the parent or guardian."[44] However, abandonment existed as a basis for termination of parental rights prior to statutory codification.[45] When determining whether abandonment had occurred at common law, courts looked to factors such as the parents' failure to financially support the child, their surrender of the child for an indefinite period of time, and whether they had consented to the termination of their rights by virtue of adoption or a guardianship order.[46] Furthermore, at common law, abandonment had to be intentional, and neglect alone was not sufficient to constitute intent to abandon.[47]

Currently, the Tennessee termination statute defines abandonment as occurring when:

> For a period of four (4) consecutive months immediately preceding the filing of a proceeding or pleading to terminate the parental rights of the parent(s) or guardian(s) of the child who is the subject of the petition for termination of parental rights or adoption, that the parent(s) or guardian(s) either have willfully failed to visit or have willfully failed to support or have willfully failed to make reasonable payments toward the support of the child.[48]

The Tennessee General Assembly has defined "willful failure to support" as occurring when the parent intentionally fails to support the child for four consecutive months, taking into account the parent's circumstances and financial means.[49] "Willful failure to visit" occurs under the statute when a parent establishes no more than "minimal or insubstantial contact with the child."[50]

---

42. *Id.* at *7.

43. TENN. CODE. ANN. § 36-1-113(c) (2005).

44. *Id.* § 36-1-113(g)(1).

45. *See* Verna Lilburn, *Abandoment as Grounds for the Termination of Parental Rights*, 5 CONN. PROB. L.J. 263, 277 (1991).

46. *Id.*

47. *Id.*

48. TENN. CODE ANN. § 36-1-102(1)(A)(i) (2005).

49. *See id.* §§ 36-1-102(1)(B), (D). Under the statute, "willfully failed to support" or "willfully failed to make reasonable payments toward such child's support" means "the willful failure, for a period of four (4) consecutive months, to provide monetary support or the willful failure to provide more than token payments toward the support of a child." *Id.* § 36-1-102(1)(D). "Token support" means that "the support, under the circumstances of the individual case, is insignificant given the parents means." *Id.* § 36-1-102(1)(B).

50. *See id.* §§ 36-1-102(1)(C), (E). "Willfully failed to visit" means "the willful failure, for a period of four (4) consecutive months, to visit or engage in more than token visitation."

Termination of parental rights does not automatically follow a finding of abandonment. Rather, the petitioner still must introduce evidence to show that termination of parental rights is in the best interests of the child concerned.[51] The state legislature has determined a number of factors that courts should consider when evaluating the best interests of a child.[52] For example, one Tennessee court found that when a father failed to visit his child for four months on account of his drug addiction, there was no longer any "meaningful relationship" between him and the child, and therefore, it was in the child's best interests that he be freed for adoption.[53]

In order to uphold the federal and state constitutional protections of parental rights, courts generally have strictly construed the termination statutes in favor of the biological parents.[54] For example, the Tennessee Supreme Court held in *In re D.L.B.* that when calculating the four month abandonment period, a court must look only to the four months immediately preceding the filing date of the petition presently under consideration; the court may not look to the dates of any earlier petitions.[55] The Tennessee Supreme Court also determined that parts of the termination statute were unconstitutional because, in effect, they created a presumption of abandonment without considering whether a parent's actions were intentional.[56] Tennessee courts have consistently found that when

---

Id. § 36-1-102(1)(E). "Token visitation" means that "the visitation, under the circumstances of the individual case, constitutes nothing more than perfunctory visitation or visitation of such an infrequent nature or of such short duration as to merely establish minimal or insubstantial contact with the child . . . ." *Id.* § 36-1-102(1)(C).

51. *See id.* § 36-1-113(c)(2).

52. Such factors include whether there has been a change in the parent's circumstances to make it safe for the child to live in the parent's home; whether the parent has maintained regular visitation and developed a meaningful relationship with the child; how the move will affect the child's emotional or psychological condition; whether the parent's home is a healthy and safe environment (which involves an assessment of the presence of criminal activity and controlled substances in the home); and whether the parent has consistently paid child support. *See id.* § 36-1-113(i).

53. *In re* B.P.C., No. M2006-02084-COA-R3-PT, 2007 WL 1159199, at *4 (Tenn. Ct. App. Apr. 18, 2007).

54. 43 C.J.S. *Infants* § 21 (2004).

55. *In re* D.L.B., 118 S.W.3d 360, 366 (Tenn. 2003). In this case, D.L.B. was placed into custody with the Tennessee Department of Children's Services after her mother admitted to using cocaine throughout her pregnancy. *Id.* at 363. CASA filed a petition for termination of parental rights in juvenile court, and some time later, D.L.B.'s foster parents filed a similar petition in chancery court. *Id.* at 364. The first petition filed in juvenile court was dismissed, but the chancery court granted the second petition for termination of parental rights on the basis of abandonment, which it calculated from the four month period preceding the first petition for termination. *Id.* The Supreme Court of Tennessee held that the chancery court was required to use the four months preceding the petition then before the court, which was the foster parents' petition. *Id.* at 366. Using this time period, there was no statutory abandonment based on willful failure to visit the child. *Id.*

56. *In re* Swanson, 2 S.W.3d 180, 188 (Tenn. 1999) (stating that the Tennessee statute failed to "allow for the type of individualized decision-making which must take place when a

a parent has been prevented from contacting, visiting, or otherwise pursuing a relationship with the child, there can be no willful failure to visit sufficient to constitute abandonment because the element of intent will not have been satisfied.[57]

B. Protection of Parental Rights in Custody Proceedings in Tennessee

Custody of a child is a parental right, protected as a fundamental liberty under both the federal and state constitutions.[58] As in termination of parental rights cases, Tennessee courts have consistently protected parents' rights to the care and custody of their children and have held that these rights may not be revoked by the state unless the parent is legally unfit to care for the child.[59] Accordingly, the Tennessee Supreme Court has held that parents cannot be deprived of custody without a showing that continuation of their custody would result in substantial harm to the child.[60] However, the court has also recognized that parental rights are not always paramount in deciding custody of a child.[61] In addition to protecting the fundamental liberties of parents, a child's best interests must be served when determining whether to modify a valid, existing custody order.[62]

In *Blair v. Badenhope*, the Tennessee Supreme Court held that biological parents may not assert a presumption of superior parental rights to modify an existing custody order, even if their initial decision to relinquish custody was voluntary.[63] The court found that when parents have validly consented to a transfer of custody, their rights are no longer superior to those of non-parents and the court will focus instead on the best interests of the child.[64] When a valid transfer of custody has occurred, "the child's interest in a stable and

---

fundamental constitutional right is at stake").

57. *See In re* F.R.R., III., 193 S.W.3d 528, 530 (Tenn. 2006) (citing *In re* D.A.H., 142 S.W.3d 267, 277 (Tenn. 2004)).

58. Blair v. Badenhope, 77 S.W.3d 137, 141 (Tenn. 2002).

59. *Id.* (indicating that the protection of natural parents' fundamental rights provided by the Tennessee Constitution is "now well-settled" in custody disputes); *In re* Askew, 993 S.W.2d 1 (Tenn. 1999) (holding that deprivation of a mother's custody of her child abridged her fundamental right to privacy without a determination that the mother's custody would result in substantial harm to the child); Davis v. Davis, 842 S.W.2d 588, 600 (Tenn. 1992) (stating that individual privacy rights protected by the "liberty clauses of the Tennessee Declaration of Rights" include parental rights).

60. *In re* Adoption of Female Child, 896 S.W.2d 546, 548 (Tenn. 1995).

61. *Blair*, 77 S.W.3d at 148 ("Though strong in many respects, no aspect of the fundamental right of parental privacy is absolute, and a parent . . . may not . . . invoke [the] doctrine [of superior parental rights] to modify a valid custody order.").

62. *Id.* However, a best interests analysis may not be performed until the court determines that substantial harm to the child would result by allowing the biological parents to retain custody. *In re Female Child*, 896 S.W.2d at 548.

63. *Blair*, 77 S.W.3d at 148.

64. *Id.*

secure environment is *at least* as important, and probably more so, than the parent's interest in having custody of the child returned[;]" thereafter, the biological parents are no longer afforded privileged rights.[65]

Parents may voluntarily transfer care and custody of their child to a non-parent through a consent order.[66] Unlike the termination of parental rights, which is absolute and may not subsequently be retracted,[67] parents who have chosen to relinquish custody may regain it later.[68] Once custody has been validly transferred, Tennessee courts apply a two-step test to determine whether the biological parents should prevail in a custody modification dispute.[69] The parents first must demonstrate a material change in circumstances warranting a modification of the custody agreement.[70] The courts have defined a change in circumstances as one that "affects the child's well-being in a meaningful way."[71] If the court finds that there are changes in circumstances to merit modification of the custody arrangement, the biological parents seeking custody must prove that a change of custody would not result in substantial harm to the child.[72] In *In re Adoption of Female Child*, a mother signed a consent order granting custody and guardianship to foster parents with the mistaken understanding that the custody order was only temporary and was necessary to secure medical insurance for her child.[73] The Tennessee Supreme Court held that the consent order transferring custody to the foster parents was invalid because a finding of substantial harm to the child had never occurred.[74] Unless such a finding has been made, it is improper for the court to engage in a general evaluation of the "best interests of the child."[75]

For a consent order transferring custody from parent to non-parent to be valid, it must be voluntary and the parent must have "knowledge of the consequences" of the decision.[76] In *Blair v. Badenhope*, the Tennessee Supreme Court asserted that "[w]here a natural parent voluntarily relinquishes custody without knowledge of the effect of that act, then it cannot be said that these rights were accorded the protection demanded by the Constitution."[77] Therefore, while a parent seeking to modify a valid, existing custody order normally will not enjoy a presumption of superior parental rights,[78] the

65. *Id.*
66. *See Blair*, 77 S.W.3d at 141
67. Harding, *supra* note 39, at 899.
68. *See Blair*, 77 S.W.3d at 141.
69. *Id.*
70. *Id.* at 146 (quoting Millet v. Andrasko, 640 So.2d 368, 371 (La. Ct. App. 1994)).
71. *Id.* at 150 (citing Hoalcraft v. Smithson, 19 S.W.3d 822, 829 (Tenn. Ct. App. 1999)).
72. *Id.* at 148.
73. *In re* Adoption of Female Child, 896 S.W.2d 546, 546 (Tenn. 1995).
74. *See id.* at 547-48.
75. *Id.* at 547.
76. *Blair*, 77 S.W.3d at 147 n.3.
77. *Id.* at 147-48 n.3.
78. *See supra* text accompanying notes 63-65.

application of the superior rights doctrine is justified if the parent did not have knowledge of the consequences when initially consenting to the change in custody.[79] Consent orders that are accomplished by fraud or without notice will also trigger a presumption of a parent's superior rights.[80] Similarly, when "the natural parent cedes only temporary and informal custody to the non-parents," the biological parents will retain their superior rights.[81]

III. TENNESSEE AFFORDS PARENTAL RIGHTS SUPERIORITY THROUGHOUT ALL JUDICIAL PROCEEDINGS INVOLVING CARE AND CUSTODY OF A CHILD

In *In re A.M.H.*, the Tennessee Supreme Court found that A.M.H.'s parents had not abandoned her through a willful failure to visit for the statutory period.[82] The court reasoned that the parents' failure to visit was not willful; rather, it was caused by the animosity that had developed between them and the custodial foster parents.[83] Although the parents were not legally prevented from visiting A.M.H., they mistakenly believed that they were not allowed to return to the foster parents' residence.[84] Further, during the statutory abandonment period, the parents had continued to pursue custody of A.M.H. via legal proceedings in juvenile court.[85] Thus, the parents could not have intentionally abandoned their child if they were simultaneously petitioning for custody.[86] The court held that there can be no "willful failure to visit" constituting abandonment when parents are actively seeking judicial resolution of a custody dispute, stating that "[w]here, as here, the parents' visits with their child have resulted in enmity between the parties and where the parents redirect their efforts at maintaining a parent-child relationship to the courts[,] the evidence does not support a 'willful failure to visit' as a ground for abandonment."[87] The court declined to assess the best interests of the child, stating that such an analysis was not required since the statutory grounds for abandonment had not been met.[88]

Following its decision on abandonment, the court addressed whether the foster parents should retain custody of A.M.H. by virtue of the consent order entered by the juvenile court in June 1999.[89] The biological parents had voluntarily signed a custody order granting the foster parents custody and

79. *Blair*, 77 S.W.3d at 147-48 n.3.
80. *Id.* at 143.
81. *Id.*
82. *In re* Adoption of A.M.H., 215 S.W.3d 793, 810-11 (Tenn. 2007).
83. *Id.* at 810.
84. *See id.* at 801.
85. *Id.* at 810.
86. *See id.* at 802, 810.
87. *Id.* at 810.
88. *Id.* at 810 n.6.
89. *Id.* at 811.

guardianship of A.M.H.[90] The court acknowledged that a voluntary transfer of custody renders the parents without superior parental rights in a subsequent custody modification proceeding.[91] However, the transfer of custody will not trump the parents' superior rights if they had no "'*knowledge of the consequences of that transfer*[.]'"[92] Here, the court found that A.M.H.'s parents were unaware of the long-term consequences of the custody order.[93] The court found substantial evidence to show that "the parents were misled as to the consequences of a change in custody and uninformed about the guardianship provision . . . ."[94] The parents believed that the custody order would indirectly provide A.M.H. with health insurance and that it would establish merely a temporary transfer of custody to the foster parents.[95] The court also stressed the fact that the mother, throughout the proceedings, repeated her wish for only a temporary transfer and asked for affirmation as to the temporary nature of the transfer before she signed the papers.[96] Based upon these findings, the court invalidated the consent order transferring custody to the foster parents.[97]

The court then addressed the competing custody claims of A.M.H.'s biological and foster parents.[98] Because the biological parents benefited from superior, constitutionally protected rights, they could only be deprived of custody upon a showing that returning custody to them would result in substantial harm to A.M.H.[99] Evidence that A.M.H. lived and bonded with the foster parents for nearly seven years was not enough for the court to hold that substantial harm would result if custody were transferred back to the biological parents.[100] Similarly, the court refused to allow A.M.H.'s quality of life in China to be compared with her current "affluent surroundings" as proof that removing her from the custody of her foster parents would result in substantial harm to her.[101] Relying on prior case law, the court stated that "mere improvement in quality of life is not a compelling state interest and is insufficient to justify invasion of Constitutional rights."[102] Without a finding of

---

90. *Id.* at 798-99.

91. *Id.* at 811.

92. *Id.* (quoting Blair v. Badenhope, 77 S.W.3d 137, 147 (Tenn. 2002)).

93. *Id.* at 811-12.

94. *Id.*

95. *Id.* at 812. The foster mother testified that A.M.H.'s parents were informed that "the custody arrangement 'could go for one year or . . . for eighteen years.'" *Id.*

96. *Id.*

97. *Id.*

98. *Id.*

99. *Id.* (quoting *In re* Askew, 993 S.W.2d 1, 4 (Tenn. 1999)). Resolution of custody disputes between parents and non-parents depends on "whether there is substantial harm threatening a child's welfare if the child returns to the parents." *Id.*

100. *Id.*

101. *Id.* at 813.

102. *Id.* (quoting Hawk v. Hawk, 855 S.W.2d 573, 582 (Tenn. 1993)) (internal quotation marks and citation omitted).

substantial harm, the court granted full custody to the biological parents of A.M.H.[103] As in its analysis of grounds for abandonment, the court declined to engage in an evaluation of A.M.H.'s best interests and instead based its decision on the federal and state constitutional protections of parental rights.[104]

A. Implications for Tennessee's Termination of Parental Rights Statutes

The court's decision in *In re A.M.H.* interpreted the pertinent Tennessee parental abandonment statutes to mean that a finding of abandonment based upon willful failure to visit is improper if the biological parents were concurrently seeking judicial resolution of the matter.[105] One Tennessee court has since adopted and expanded the holding of *In re A.M.H.* to apply to other aspects of the parental termination and abandonment statutes.[106] In *In re Chelbie F.*, the Tennessee Court of Appeals relied upon the rationale of the court *In re A.M.H.* to hold that a father seeking custody did not abandon his child by willfully failing to provide financial support.[107] The court held that although the father failed to provide financial support for his child during the four months preceding the mother's petition for termination of his parental rights, he could not have statutorily abandoned his child because he was pursuing judicial resolution of a dispute over child support payments.[108] The court stated that to find otherwise would "devalue Tennessee's laws and judicial procedures" considering the decision in *In re A.M.H.*[109]

The overarching impact of the court's decision in *In re A.M.H.* is to establish that "[Tennessee] courts must protect the parent-child relationship throughout [all] judicial proceedings involving [] children."[110] Consequently, parents may seek peaceful judicial resolution of custody disputes without fearing that the child's current custodians or guardians will terminate their parental rights. Protection of parental rights throughout judicial proceedings is paramount in cases such as *In re A.M.H.* where hostility between the parties has prevented the biological parents from visiting their child at the foster parents' home.[111] The court's decision protects biological parents from foster parents who, looking to adopt, capitalize on the hostility that has developed between the parties and force the parents into abandoning their child for the statutorily prescribed abandonment period. Furthermore, it discourages foster parents from prolonging judicial proceedings in hopes that the biological parents may

---

103. *Id.*

104. *See id.* at 811-13.

105. *See id.* at 810-11.

106. *In re* Chelbie F., No. M2006-01889-COA-R3-PT, 2007 WL 1241252, at *6 (Tenn. Ct. App. Apr. 27, 2007).

107. *Id.* at *6-7.

108. *Id.* at *6.

109. *See id.* at *6-7.

110. *In re* Tiffany B., 228 S.W.3d 148, 157 n.16 (Tenn. Ct. App. 2007).

111. *In re* Adoption of A.M.H., 215 S.W.3d 793, 810 (Tenn. 2007).

statutorily abandon their child during the litigation process. As a result, the decision promotes efficiency and cooperation between the parties by encouraging them to quickly reach an amicable agreement.

However, protecting parental rights throughout the litigation process does not come without some negative consequences for the child involved. In this case, the trial on the petitions for custody, adoption, and termination of parental rights began nearly three years after the parties originally filed the petitions.[112] The ultimate resolution of the case did not come until almost four years later.[113] Thus, A.M.H. was returned to the custody of her biological parents more than seven years after she had originally been placed with the foster parents.[114] The court's decision in this case serves to strengthen the importance of parental rights as a constitutionally protected liberty; however, the court left unanswered questions regarding how to protect parental rights while minimizing the damage to children caused by the lengthy litigation process.

B. Effects on the Resolution of Custody Disputes in Tennessee

The longstanding presumption has been that a child's best interests are served by keeping the familial unit intact and awarding custody to the biological parents.[115] Accordingly, Tennessee courts historically have held that a best interests analysis is unnecessary and even constitutionally prohibited when there has been no showing that the child would suffer substantial harm if returned to her parents.[116] However, a more modern view recognizes that awarding custody to a parent when a non-parent has maintained custody for all or most of the child's life may result in permanent psychological detriment to the child.[117] Thus, many courts have chosen to balance the protection of parental rights with the best interests of the child because "the possibility of serious psychological harm may transcend all other considerations."[118] Even when parents have maintained contact with their children and possess superior parental rights, some courts have asserted that the importance of "continuity of residence" to the child trumps even constitutionally protected parental rights.[119]

---

112. *Id.* at 804.

113. *Id.* at 813.

114. *Id.* at 796-97.

115. *See* Carol A. Crocca, Annotation, *Continuity of Residence as Factor in Contest Between Parent and Nonparent for Custody of Child Who Has Been Residing with Nonparent—Modern Status,* 15 A.L.R.5th 692, 714 (1993).

116. *See In re* Adoption of Female Child, 896 S.W.2d 546, 547-48 (Tenn. 1995) (quoting Hawk v. Hawk, 855 S.W.2d 573, 577 (Tenn. 1993)); Davis v. Davis, 842 S.W.2d 588, 600 (Tenn. 1992).

117. Crocca, *supra* note 115, at 692.

118. *Id.* at 714.

119. *Id.* at 757-58. "Continuity of residence" means that the court will consider "the child's interest in retaining the security of the environment to which he or she has become accustomed, including the personal relationships he or she has established, and the detrimental effect on the child of removal from that environment." *Id.* at 712.

The Tennessee Supreme Court concluded that "the only evidence of substantial harm [to A.M.H. arose] from the delay caused by the protracted litigation and the failure of the court system to protect the parent-child relationship throughout the proceedings" and that this harm was not significant enough to justify denying the biological parents custody of their child.[120] However, custody battles and judicial proceedings are often lengthy affairs that can take years to resolve. The court never addressed how the drawn-out litigation process may affect children, how to prevent harm to children caught in the middle of custody battles, or even if courts should consider such harm when assessing custody cases. Instead, the court emphasized only the protection of parental rights,[121] resulting in public policy which favors the superiority of parental rights at the expense of all other considerations. The decision in *In re A.M.H.* prevents Tennessee from moving forward as other jurisdictions have done, recognizing that the child's interests should be considered in all custody proceedings alongside parental rights.[122]

Furthermore, the court's extensive protection of parental rights may give biological parents an incentive to capriciously relinquish custody, and when they later wish to regain custody, simply claim that they did not understand the consequences of the consent order. This standard may encourage parents to relinquish custody without taking time to fully consider the implications of such a decision and the profound impact it will have on the life of the child. The best interests of the child would suffer should she be transferred between custodians upon her parents' whim, yet this result may emerge from the unwavering protection of parental rights in Tennessee.

IV. CONCLUSION

Parents in Tennessee have superior rights to the custody of their children without interruption, but only as long as they have not relinquished their rights or engaged in conduct which would require their limitation or termination.[123] When interpreting the state's termination of parental rights statutes, Tennessee courts have strongly emphasized the superiority of parental rights as a constitutionally protected liberty.[124] Historically, the Tennessee Supreme Court has held that parental rights remain superior throughout termination of parental rights cases, and therefore, courts may not perform a best interests analysis unless there has been a showing of statutory grounds necessitating the

120. *In re* Adoption of A.M.H., 215 S.W.3d 793, 812-13 (Tenn. 2007).

121. *See id.* at 812.

122. *See supra* text accompanying notes 115-19 (discussing the modern trend toward considering the best interests of a child who is caught in a custody dispute).

123. *In re* Tiffany B., 228 S.W.3d 148, 155 (Tenn. Ct. App. 2007) (citing Blair v. Badenhope, 77 S.W.3d 137, 141 (Tenn. 2002)).

124. *See* TENN. CODE ANN. § 36-1-113 (2005); Hawk v. Hawk, 855 S.W.2d 573, 579 (Tenn. 1993).

termination.[125] Similarly, the court has consistently held that a best interests analysis need not be performed in custody disputes unless substantial harm to the child has been demonstrated.[126]

The court in *In re A.M.H.* was justified in requiring proof of the statutory elements in termination of parental rights cases before engaging in a best interests analysis. However, the same standard should not be applied to custody disputes; unlike the termination of parental rights, custody is not permanent[127] and multiple modifications to a child's custody arrangement over a period of time can be against the child's best interests. Other jurisdictions have recognized that even when substantial harm to the child may not result from a change in custody, the child's best interest is still an important factor to weigh in custody modification disputes.[128] The Tennessee Supreme Court chose to adhere to the traditional rule in which constitutionally protected parental rights supersede all other considerations.[129] However, the court should have followed the more modern trend of performing a best interests analysis in all custody disputes in order to balance the child's best interests with the rights of the biological parents.[130] This approach allows the child's well-being to be a factor in determining his or her future and recognizes that much more is at stake in a custody dispute than simply the protection of superior parental rights.

ERIN E. BYBEE

125. *See In re A.M.H.*, 215 S.W.3d at 810-11 n.6; *In re* D.L.B., 118 S.W.3d 360, 368 (Tenn. 2003); *Blair*, 77 S.W.3d at 146; *In re* Adoption of Female Child, 896 S.W.2d 546, 548 (Tenn. 1995).

126. *In re Female Child*, 896 S.W.2d at 548; *see In re D.L.B.*, 118 S.W.3d at 368 (citing TENN. CODE ANN. § 36-1-113(c) (2001)).

127. *See In re A.M.H.*, 215 S.W.3d at 811, 812-13.

128. Crocca, *supra* note 115, at 758.

129. *See In re A.M.H.*, 215 S.W.3d at 812.

130. *See* Crocca, *supra* note 115, at 714.

CRIMINAL LAW—INDICTMENT SPECIFICITY IN ALLEGING ATTEMPT CRIMES—AN INDICTMENT FOR ATTEMPTED ILLEGAL REENTRY INTO THE UNITED STATES IS NOT DEFECTIVE BECAUSE IT FAILS TO ALLEGE A SPECIFIC OVERT ACT

United States v. Resendiz-Ponce, 127 S. Ct. 782 (2007).

I. INTRODUCTION

Juan Resendiz-Ponce, a Mexican citizen, was deported from the United States in 1997 and again in 2002.[1] On June 1, 2003, he attempted reentry once again when he approached a checkpoint at the United States-Mexico border, presented photo identification of his cousin to the border agent, and announced that he was a legal United States resident traveling to Calexico, California.[2] The fraudulent photo aroused suspicions, and thereafter the defendant was questioned, taken into custody, and charged with attempted illegal reentry in violation of 8 U.S.C. § 1326(a).[3] The full text of the indictment stated the following:

> On or about June 1, 2003, JUAN RESENDIZ-PONCE, an alien, knowingly and intentionally attempted to enter the United States of America at or near San Luis in the District of Arizona, after having been previously denied admission, excluded, deported, and removed from the United States at

---

1. United States v. Resendiz-Ponce, 425 F.3d 729, 730 (9th Cir. 2005), *rev'd*, 127 S. Ct. 782 (2007). The defendant illegally entered the United States in 1988 and was first ordered deported in 1997. *Id.* He illegally entered again in 2002 and was convicted a few months later of kidnapping his common-law wife and sentenced to forty-five days in jail. *Id.* On October 15, 2002, while in jail, he divulged to an Immigration and Naturalization Service agent that he was an alien who had been previously deported and had not received the consent of the Attorney General for reentry. *Id.* He was deported that same day. *Id.*

2. United States v. Resendiz-Ponce, 127 S. Ct. 782, 786 (2007).

3. *Id.* 8 U.S.C. § 1326(a) provides that an alien will be fined or imprisoned for up to two years who

> (1) has been denied admission, excluded, deported, or removed or has departed the United States while an order of exclusion, deportation, or removal is outstanding, and thereafter (2) enters, attempts to enter, or is at any time found in, the United States, unless (A) prior to his reembarkation at a place outside the United States or his application for admission from foreign contiguous territory, the Attorney General has expressly consented to such alien's reapplying for admission; or (B) with respect to an alien previously denied admission and removed, unless such alien shall establish that he was not required to obtain such advance consent under this chapter or any prior Act . . . .

8 U.S.C. § 1326(a) (2000).

> or near Nogales, Arizona, on or about October 15, 2002, and not having obtained the express consent of the Secretary of the Department of Homeland Security to reapply for admission.
>
> In violation of Title 8, United States Code, Sections 1326(a) and enhanced by (b)(2).[4]

At trial in the United States District Court for the District of Arizona, the defendant moved to dismiss the indictment, arguing that it "'fail[ed] to allege an essential element, an overt act, or to state the essential facts of such overt act.'"[5] The district court denied the motion and the jury found the defendant guilty.[6] The defendant appealed, and the United States Court of Appeals for the Ninth Circuit reversed, holding that an indictment's omission of "an essential element of the offense is a fatal flaw not subject to mere harmless error analysis."[7] The court found that the defendant's indictment was insufficient because it failed "to allege any specific overt act that [was] a substantial step toward" the completion of the unlawful reentry.[8] On certiorari

---

4. *Resendiz-Ponce*, 425 F.3d at 731.

5. *Resendiz-Ponce*, 127 S. Ct. at 786 (alterations in original) (quoting the indictment).

6. *Id.* At trial, the defendant also moved (1) to suppress the statements he made to the Immigration and Naturalization Service agent in 2002 because he was not given proper Miranda warnings, (2) to dismiss the indictment because it did not allege an overt act, and (3) to strike the part of the indictment relating to 8 U.S.C. § 1326(b)(2) because it failed to allege that his prior deportation occurred subsequent to his prior conviction. *Resendiz-Ponce*, 425 F.3d at 731. The motions were denied. *Id.* The defendant also requested a charge instructing the jury that the Government was required to prove that he "performed the overt act of successfully reentering the United States." *Id.* The district court also denied this request. *Id.*

At sentencing, the district court concluded that the defendant was previously deported after his conviction for an aggravated felony, and as a result, his maximum sentence would be increased from two years to twenty years pursuant to 8 U.S.C. § 1326(b)(2). *Id.* The court ultimately sentenced him to sixty-three months, the median of the fifty-seven to seventy-one month range. *Id.*

7. *Resendiz-Ponce*, 425 F.3d at 732. The defendant also argued on appeal that his Miranda rights were violated, that the judge erroneously rejected his jury instructions, and that his sentence violated the Sixth Amendment. *Id.* at 730. Because the Ninth Circuit found the indictment insufficient, it did not reach the defendant's remaining claims. *Id.* The Ninth Circuit relied upon its previous decision in *United States v. Du Bo*, in which the court stated that "if properly challenged prior to trial, an indictment's complete failure to recite an essential element of the charged offense is not a minor or technical flaw subject to harmless error analysis, but a fatal flaw requiring dismissal of the indictment." United States v. Du Bo, 186 F.3d 1177, 1179 (9th Cir. 1999).

8. *Resendiz-Ponce*, 425 F.3d at 733. The Ninth Circuit panel majority explained:

> The defendant has a right to be apprised of what overt act the government will try to prove at trial, and he has a right to have a grand jury consider whether to charge that specific overt act. Physical crossing into a government inspection area is but one of a number of other acts that the government might have alleged as a substantial step toward entry into the United States. The indictment might have alleged the tendering [of] a bogus identification card; it might have alleged successful clearance of the inspection area; or it

to the United States Supreme Court, *held*, reversed and remanded.[9] An indictment for attempted illegal reentry into the United States under 8 U.S.C. § 1326(a) is not defective because it fails to allege a specific overt act. *United States v. Resendiz-Ponce*, 127 S. Ct. 782 (2007).

The United States Supreme Court granted the Government's petition for certiorari to answer the question of "whether the omission of an element of a criminal offense from a federal indictment can constitute harmless error."[10] However, because "'[i]t is not the habit of the Court to decide questions of a constitutional nature unless absolutely necessary to a decision of the case,'"[11] the Court instructed the parties to file supplemental briefs to address whether the indictment was defective because it failed to explicitly allege a specific overt act constituting an attempt.[12] Both parties agreed that a violation of 8 U.S.C. § 1326(a) requires the defendant to have committed an "overt act qualifying as a substantial step toward completion of his goal"[13] and that an indictment must specify "'each element of the crime that it charges.'"[14] The Government, however, contended that the indictment implicitly alleged that the defendant engaged in the necessary overt act[15] "simply by alleging that he 'attempted to enter the United States' unlawfully."[16] In response, the defendant asserted that the indictment was defective because it failed to explicitly allege the required overt act that he committed in seeking reentry.[17] Because the Court ultimately concluded that the indictment was sufficient, it did not reach the harmless error issue.[18]

---

might have alleged lying to an inspection officer with the purpose of being admitted. . . . A grand jury never passed on a specific overt act, and Resendiz was never given notice of what specific overt act would be proved at trial.

Id. Judge Reavley concurred, but only because he felt bound by Ninth Circuit precedent. *Id.* (Reavley, J., concurring). Otherwise, he would have found the indictment to be constitutionally sufficient because it "clearly inform[ed] the defendant of the precise offense of which he [was] accused so that he [could] prepare his defense and so that a judgment thereon [would] safeguard him from a subsequent prosecution for the same offense." *Id.*

9. *Resendiz-Ponce*, 127 S. Ct. at 790.

10. *Id.* at 785. The Government did not seek review of the Ninth Circuit's ruling that the indictment failed to allege the overt act element of attempted unlawful entry. *Id.*

11. *Id.* (quoting Ashwander v. Tenn. Valley Auth., 297 U.S. 288, 347 (1936)).

12. *See id.* at 785-86.

13. *Id.* at 787.

14. *Id.* (quoting Almendarez-Torres v. United States, 523 U.S. 224, 228 (1998)).

15. *See* Supplemental Brief for the United States at 3, United States v. Resendiz-Ponce, 127 S. Ct. 782 (2007) (No. 05-998), 2006 WL 3073300.

16. *Id.* at 8.

17. *See* Supplemental Reply Brief of Respondent at 1-2, United States v. Resendiz-Ponce, 127 S. Ct. 782 (2007) (No. 05-998), 2006 WL 3243132.

18. *Resendiz-Ponce*, 127 S. Ct. at 786.

II. DEVELOPMENT OF ATTEMPT CRIMES, INDICTMENT PLEADING, AND INDICTMENT SPECIFICITY IN ALLEGING VIOLATIONS OF 8 U.S.C. § 1326

In order to be constitutionally sufficient, an indictment must set forth each element of the crime charged.[19] In *Hamling v. United States*, the Supreme Court identified two constitutional requirements for a sufficient indictment.[20] First, the indictment must "contain[] the elements of the offense charged and fairly inform[] a defendant of the charge against which he must defend[.]"[21] Second, it must "enable[] him to plead an acquittal or conviction in bar of future prosecutions for the same offense."[22] Thus, a defendant has the right to an indictment that is written in clear and direct language establishing the crime being charged and the elements constituting that crime,[23] in sufficient detail to enable the preparation of a defense and to afford protection against future prosecutions for the same offense.[24] The test on a motion to dismiss is not whether the Government might have made the charge more certain, but whether the indictment "sufficiently apprises the defendant" of the charges against him and "contains every element of the offense intended to be charged[.]"[25]

Generally speaking, an indictment is not insufficient for simply restating the language of a federal criminal statute.[26] For example, in *Hamling*, the Court held that an indictment alleging a violation of 18 U.S.C. § 1461, which prohibits the mailing of obscene materials, was sufficient when it listed only the statutory language and not the "component parts of the constitutional definition of obscenity."[27] However, some crimes must be charged with greater specificity.[28] For instance, an indictment under 2 U.S.C. § 192, which makes it a crime for a witness summoned before a congressional committee to refuse to answer any question "pertinent to the question under inquiry,"[29] must contain more than a recitation of the statutory language.[30] The United States Supreme Court held in *Russell v. United States* that a valid indictment for violation of this statute must go beyond the words of the statute and identify "the subject which was under inquiry at the time of the defendant's alleged default or

19. *See* Almendarez-Torres v. United States, 523 U.S. 224, 228 (1998).
20. Hamling v. United States, 418 U.S. 87, 117 (1974).
21. *Id.*
22. *Id.*
23. *See* United States v. Carll, 105 U.S. 611, 612 (1881).
24. *See Hamling*, 418 U.S. at 117.
25. Cochran v. United States, 157 U.S. 286, 290 (1895).
26. *See Hamling*, 418 U.S. at 117.
27. *Id.* at 117-19.
28. *See id.* at 117-18 ("Undoubtedly the language of the statute may be used in the general description of an offence, but it must be accompanied with such a statement of the facts and circumstances as will inform the accused of the specific offence, coming under the general description, with which he is charged.").
29. 2 U.S.C. § 192 (2000).
30. *See* Russell v. United States, 369 U.S. 749, 752-55 (1962).

refusal to answer."[31] The Court noted that "[w]here guilt depends so crucially upon such a specific identification of fact, our cases have uniformly held that an indictment must do more than simply repeat the language of the criminal statute."[32]

However, the *Russell* majority also highlighted a 1872 statute which, as the Court described, "reflected the drift of the law away from the rules of technical and formalized pleading which had characterized an earlier era."[33] This statute had provided that "'no indictment . . . shall be deemed insufficient, nor shall the trial, judgment, or other proceeding thereon be affected by reason of any defect or imperfection in matter of form only, which shall not tend to the prejudice of the defendant.'"[34] Although the 1872 statute was ultimately repealed, its substance was preserved in the Federal Rules of Criminal Procedure,[35] which were promulgated in 1948 and were "designed to eliminate technicalities in criminal pleading and . . . to secure simplicity in procedure."[36] The Court's discussion of this statute in *Russell* reveals its struggle to maintain a balance between two competing desires—the need for indictments that clearly and directly state the elements of the charged crime, and the avoidance of indictments that are overwhelmed by technical and formalized pleading requirements.

At common law, the mere intent to violate a federal criminal statute is not punishable as an attempt unless it is accompanied by some significant conduct.[37] Thus, criminal attempt contains two substantive elements: the intent to commit the underlying crime and the undertaking of some action toward the commission of that crime.[38] The significant conduct element has been described as an "overt act" that constitutes a "substantial step" toward completion of the offense.[39] The substantial step must be the type of behavior

---

31. *Id.*

32. *Id.* at 764.

33. *Id.* at 762.

34. *Id.* (quoting 17 Stat. 198 (1872)). The 1872 statute became Rev. Stat. § 1025 and then 18 U.S.C. § 556 (1940). *Id.* at 762 n.11. Ultimately, the statute was repealed in 1948 because its substance was contained in Rule 52(a) of the Federal Rules of Criminal Procedure. *Id.*

35. *Id.* at 762.

36. United States v. Debrow, 346 U.S. 374, 376 (1953).

37. *See* United States v. Resendiz-Ponce, 127 S. Ct. 782, 787 (2007); Edwin R. Keedy, *Criminal Attempts at Common Law*, 102 U. PA. L. REV. 464, 468 (1954) (stating that preparation alone, without some act that is "a start towards the accomplishment of the result intended" is not enough for criminal attempt); 22 C.J.S. *Criminal Law* § 157 (2006) ("The accused must have taken a substantial step beyond mere preparation, by doing something directly moving toward, and bringing him or her nearer, the crime he or she intends to commit . . . .").

38. *See* Braxton v. United States, 500 U.S. 344, 349 (1991); United States v. Arbelaez, 812 F.2d 530, 534 (9th Cir. 1987); WAYNE R. LAFAVE & AUSTIN W. SCOTT, JR., CRIMINAL LAW 495 (2d ed. 1986); 21 AM. JUR. 2D *Criminal Law* § 175 (1998).

39. *Arbelaez*, 812 F.2d at 534; *see* MODEL PENAL CODE § 5.01(1)(c) (1985); 22 C.J.S.

that "a reasonable observer, [when] viewing it in context could conclude beyond a reasonable doubt that it was undertaken in accordance with a design to violate the statute."[40]

The prevailing common law doctrine thus dictates that the Government must allege all elements of an attempt crime in the indictment.[41] However, another view holds that because the word *attempt* implies both intent and the actual effort to consummate that intent, an indictment for an attempt crime is sufficient without explicitly alleging each individual act constituting the attempt.[42] For example, the Supreme Court of Hawaii decided a case in which the defendant was charged with attempted second degree murder of his wife,[43] and the language of the complaint was drawn directly from the language of the statute that defined criminal attempt.[44] The defendant argued that the court erroneously denied his motion to dismiss the complaint for failure to state a cause of action for attempted murder,[45] which was based upon his contention that the intent element of attempted murder was missing from the charges against him.[46] However, the Supreme Court of Hawaii decided the case more broadly, emphasizing that

> "[n]o indictment or bill of particulars is invalid or insufficient for the reason merely that it alleges *indirectly and by inference* instead of directly any matters, facts, or circumstances connected with or constituting the offense, provided that the nature and cause of the accusation can be understood by a person of common understanding."[47]

The court conceded that, by tracking the text of the statute verbatim, the language of the complaint was "awkward."[48] Nevertheless, it concluded that the complaint "set forth with reasonable clarity all of the elements of the offense" and that a "person of common understanding . . . would understand

---

Criminal Law § 157 (2006).

40. United States v. Manley, 632 F.2d 978, 987-88 (2d Cir. 1980).

41. *See* Almendarez-Torres v. United States, 523 U.S. 224, 228 (1998); *supra* notes 19-25 and accompanying text.

42. *See* People v. Miller, 42 P.2d 308, 308 (Cal. 1935) (relying on the state penal code which provided that an indictment was sufficient if the statement of the offense was "in the words of the enactment describing the offense . . . or in any words sufficient to give the defendant notice of the offense of which he is accused"); State v. Moore, 921 P.2d 122 (Haw. 1996); United States v. Toma, No. 94-CR-333, 1995 WL 65031, at *1 (N.D. Ill. Feb. 13, 1995) ("[F]or indictment purposes, use of the word 'attempt' is sufficient to incorporate the substantial step element. The word 'attempt' necessarily means taking a substantial step." (footnote omitted)).

43. *Moore*, 921 P.2d at 126-27.

44. *Id.* at 136.

45. *Id.* at 135.

46. *Id.* at 136.

47. *Id.* (emphasis added) (quoting HAW. REV. STAT. § 806-31(1993)).

48. *Id.*

. . . the nature of the accusation[.]"[49] The Supreme Court of California applied similar reasoning in its 1935 decision in *People v. Miller*.[50] In that case, a California jury convicted the defendant of attempted murder, and on appeal, he contended that the charging information was insufficient because it did not include "allegations of facts showing the overt or other acts constituting the attempt[.]"[51] The Supreme Court of California rejected the defendant's contention and concluded that the indictment was sufficient because its language gave the defendant sufficient notice of the offense.[52]

Prior to *Resendiz-Ponce*, the Ninth Circuit had already considered the elements of the crime of attempted illegal reentry in violation of 8 U.S.C. § 1326. In *United States v. Gracidas-Ulibarry*, the Ninth Circuit addressed the level of intent that the Government must "prove to convict an alien of *attempted* illegal reentry under § 1326[.]"[53] The facts in *Gracidas-Ulibarry* are similar to the facts in *Resendiz-Ponce*. The defendant, who had been deported the day before from the Calexico, California border checkpoint, was discovered riding as a passenger in a car being driven through the checkpoint at San Ysidro, California.[54] At first, the defendant claimed to be a United States citizen, but after further questioning by immigration inspectors, he admitted his Mexican citizenship and previous deportation.[55] In its analysis, the Ninth Circuit acknowledged five elements of the crime of attempted illegal reentry under § 1326, including both intent (which it characterized as a "conscious desire") and an overt act that constitutes a "substantial step" toward reentering.[56] Focusing specifically on the intent element, the Ninth Circuit held that the crime of attempted illegal reentry requires proof of a "specific intent to enter illegally."[57] Integral to the court's rationale was its conclusion that in

---

49. *Id.*

50. People v. Miller, 42 P.2d 308 (Cal. 1935).

51. *Id.*

52. *See id.*

53. United States v. Gracidas-Ulibarry, 231 F.3d 1188, 1190 (9th Cir. 2000) (en banc).

54. *Id.* at 1190-91.

55. *Id.* at 1191.

56. *Id.* at 1196. The five elements of attempted illegal reentry under § 1326(a) that the court found are as follows:

> (1) the defendant had the purpose, i.e., conscious desire, to reenter the United States without the express consent of the Attorney General; (2) the defendant committed an overt act that was a substantial step towards reentering without that consent; (3) the defendant was not a citizen of the United States; (4) the defendant had previously been lawfully denied admission, excluded, deported or removed from the United States; and (5) the Attorney General had not consented to the defendant's attempted reentry.

Id. (citing United States v. Davis, 960 F.2d 820, 826-27 (9th Cir. 1992)).

57. *Id.* at 1191-92. At issue in *Gracidas-Ulibarry* was whether the jury should have been instructed to find the defendant guilty only if it first found that he had the specific intent to reenter the United States illegally, rather than if it found that he merely had, in fact, attempted reentry. *Id.* at 1191. The court held that § 1326 does require a finding of specific intent to reenter, but that in this case, the erroneous jury instruction constituted harmless error because

enacting § 1326, Congress intended to incorporate the common law meaning of attempt, which includes both a specific intent[58] and an overt act.[59]

III. UNITED STATES V. RESENDIZ-PONCE: AN INDICTMENT FOR ATTEMPTED ILLEGAL REENTRY INTO THE UNITED STATES IS NOT DEFECTIVE BECAUSE IT FAILS TO ALLEGE A SPECIFIC OVERT ACT

In *United States v. Resendiz-Ponce*, the United States Supreme Court considered whether an indictment alleging attempted illegal reentry under § 1326 is defective because it fails to identify a particular overt act.[60] Justice Stevens delivered the opinion of the Court, joined by Chief Justice Roberts and Justices Kennedy, Souter, Thomas, Ginsburg, Breyer, and Alito.[61] Justice Scalia filed a dissenting opinion.[62] The majority of the Court held that "an indictment alleging attempted illegal reentry under § 1326(a) need not specifically allege a particular overt act or any other 'component par[t]' of the offense."[63]

The Government conceded that Resendiz-Ponce could not be guilty of attempted reentry in violation of § 1326(a) unless he committed an overt act toward the completion of his goal[64] and that an indictment must identify each element of the charged offense.[65] However, the Government contended that its indictment fulfilled these requirements; by stating that the defendant "attempted to enter the United States" unlawfully, the indictment implicitly alleged that he had committed an overt act.[66] The Government reasoned that the indictment "need not elaborate on all of the ingredients of [the] elements [of the charged offense]."[67] The Court agreed with the Government's supposition for two reasons.[68] First, the common meaning of the word *attempt* "connote[s] action rather than mere intent[.]"[69] Second, and more importantly, the word *attempt* as historically used in the law "encompasses both the overt act and intent elements."[70] As a result, the *Resendiz-Ponce* majority held that an indictment alleging a violation of § 1326(a) need not specify a particular overt act or any

---

the weight of the evidence showed that it was the defendant's conscious desire to reenter illegally. *Id.* at 1197-98.

58. *Id.* at 1193-95.

59. *See supra* notes 37-40 and accompanying text.

60. United States v. Resendiz-Ponce, 127 S. Ct. 782, 785-86 (2007).

61. *Id.* at 785.

62. *Id.*

63. *Id.* at 787-88 (alteration in original) (citing Hamling v. United States, 418 U.S. 87, 119 (1974)).

64. *See* Supplemental Brief for the United States, *supra* note 15, at 6-7.

65. *Id.* at 4.

66. *Id.* at 8.

67. *Id.* at 4.

68. *Resendiz-Ponce*, 127 S. Ct. at 787.

69. *Id.*

70. *Id.*

other "component part" of the offense; rather, the overt act is implied in the word *attempt*.[71] Additionally, the Court determined that the use of the word *attempt* in the indictment, along with more detailed information regarding the time and place of the defendant's alleged attempt to reenter, satisfied both of the constitutional requirements for an indictment identified in *Hamling v. United States*.[72]

The defendant, on the other hand, contended that the indictment was insufficient because it failed to allege any of the three overt acts that he performed during his attempted reentry: physically crossing the border, presenting a misleading identification card, and lying to the border agent.[73] The majority found that while each of those acts individually tended to prove the charged attempt crime, "none was essential to a finding of guilt[.]"[74] Rather, "all three acts were part of a single course of conduct culminating in the charged 'attempt.'"[75] Likewise, the majority analogized that it would be incorrect to claim that the defendant actually had committed three separate attempt offenses for each of the three overt acts; such reasoning would only increase the risk of successive prosecutions for the same offense.[76]

The majority offered two reasons for distinguishing *Resendiz-Ponce* from *Russell v. United States*, in which the Court held that an indictment alleging a violation of a federal criminal statute must do more than simply repeat the statutory language.[77] First, unlike the federal statute at issue in *Russell*, "guilt under 8 U.S.C. § 1326(a) does not 'depen[d] so crucially upon such a specific identification of fact.'"[78] Second, the majority relied upon the principles underlying the Federal Rules of Criminal Procedure, which the *Russell* Court had explored but ultimately found to be unpersuasive in light of the special need for particularity in charges brought under 2 U.S.C. § 192.[79] The policy

---

71. *See id.* at 787-88.

72. *Id.* at 788. *See generally supra* notes 20-22 and accompanying text (listing *Hamling*'s constitutional requirements for an indictment). In fact, the *Resendiz-Ponce* majority noted that the information provided to the defendant in the indictment concerning time and place of the attempt actually provided him better notice than would an indictment identifying specific overt acts. *Resendiz-Ponce*, 127 S. Ct. at 788. Because a defendant in Resendiz-Ponce's situation could have attempted reentry on several different occasions, the Court reasoned that an indictment specifying the time, date, and place of the attempt actually provided greater protection against the risk of multiple prosecutions for the same crime. *See id.*

73. *See Resendiz-Ponce*, 127 S. Ct. at 788; Supplemental Reply Brief of Respondent, *supra* note 17, at 8 (arguing that the indictment would have been sufficient if it had simply alleged that the defendant had presented phony identification, so that he would have had notice that the Government did not intend to prove that his substantial step was physically crossing the border or lying to border agents).

74. *Resendiz-Ponce*, 127 S. Ct. at 788.

75. *Id.*

76. *Id.* at 788 n.5

77. Russell v. United States, 369 U.S. 749, 764 (1962).

78. *Resendiz-Ponce*, 127 S. Ct. at 789 (quoting *Russell*, 369 U.S. at 764).

79. *Id. See generally supra* notes 31-36 (explaining the context of the *Russell* Court's

underlying Federal Rule of Criminal Procedure 7(c)(1), which provides that an indictment "must be a plain, concise, and definite written statement of the essential facts constituting the offense charged[,]"[80] is to "eliminate technicalities in criminal pleadings . . . [and] to secure simplicity in procedure."[81] Thus, the *Resendiz-Ponce* majority maintained that Rule 7(c)(1) eliminated the expectation of detailed allegations in an indictment which were previously required under common law pleading rules.[82]

The majority also briefly examined Federal Rule of Criminal Procedure 31(c),[83] which provides that a defendant may be found guilty of "an attempt to commit the offense charged" or "an attempt to commit an offense necessarily included in the offense charged, if the attempt is an offense in its own right."[84] The majority concluded that "[i]f a defendant indicted only for a completed offense can be convicted of attempt under Rule 31(c) without the indictment ever mentioning an overt act, it would be illogical to dismiss an indictment charging 'attempt' because it fails to allege such an act."[85]

Justice Scalia provided the lone dissent in *Resendiz-Ponce*, fundamentally disagreeing with the majority's position that the "substantial step" requirement is implicit in the word *attempt*.[86] Justice Scalia asserted that the Government was required to explicitly identify both the intent and the overt act elements of attempt in the indictment.[87] However, Justice Scalia concluded that the indictment need not have specified the particular overt act upon which the Government would rely at trial; rather, it needed only to state that the defendant "took a substantial step" toward that end.[88] The majority disagreed with this assertion based upon its reasoning that the "substantial step" element of an attempt crime is implicit in the word *attempt*.[89] Furthermore, the majority asserted that the addition of such a simple phrase to the indictment would not have given the defendant "any greater notice of the charges against him or protection against future prosecution."[90]

In his dissent, Justice Scalia characterized the majority's holding as an "exception to the standard practice" of requiring an indictment for attempt to allege both that the defendant intended to commit the crime and that he took

---

decision under 2 U.S.C. § 192 and its discussion of an 1872 statute which was eventually incorporated into the Federal Rules of Criminal Procedure).

80. FED. R. CRIM. P. 7(c)(1).

81. *Resendiz-Ponce*, 127 S. Ct. at 789 (quoting United States v. Debrow, 346 U.S. 374, 376 (1953)).

82. *Id.*

83. *Id.* at 789 n.7.

84. FED. R. CRIM. P. 31(c).

85. *Resendiz-Ponce*, 127 S. Ct. at 789 n.7.

86. *See id.* at 790 (Scalia, J., dissenting).

87. *Id.*

88. *Id.* at 792-93.

89. *Id.* at 788 n.4 (majority opinion).

90. *Id.*

steps toward that end.[91] Justice Scalia rejected the majority's reasons for its special "exception," characterizing them as "irrelevant" and "probably incorrect."[92] Justice Scalia first dismissed the majority's proposition that the word *attempt* in common parlance connotes, and thus implies, both the intent and overt act elements of an attempt crime.[93] He argued that this assertion is irrelevant because our courts have always required the elements of a crime to be explicitly set forth in the indictment, regardless of whether the name of the crime elicits them.[94] He stated that the indictment must "fully, directly, and *expressly*, without any uncertainty or ambiguity, set forth all the elements necessary to constitute the offence intended to be punished."[95] Justice Scalia also believed that the majority's argument was incorrect because it overstated the precision with which the word *attempt* is understood in common usage.[96] Specifically, Justice Scalia asserted that a reasonable grand juror could believe that the word *attempt* implies intent coupled with any minor action—rather than a "substantial step"—toward commission of the crime charged.[97]

Justice Scalia also refuted the majority's proposition that, throughout history, the legal meaning of the word *attempt* has included both the overt act and intent elements.[98] First, Justice Scalia asserted that this argument is irrelevant because the elements of many common law crimes have remained unchanged throughout history, yet our courts nevertheless require those elements to be pled with specificity in indictments.[99] Second, Justice Scalia criticized the majority for overlooking the historical inconsistency of the definition of attempt.[100] Because the definition has varied greatly throughout the past century, Justice Scalia contended that the historical meaning of attempt is not as straightforward as the majority asserted.[101] Justice Scalia reasoned that based on the majority's logic, the indictment also could have neglected to mention that the defendant "knowingly and intentionally" attempted to reenter the United States because this phrase is "understood in 'common parlance,' and has been an element of attempt for centuries."[102]

---

91. *Id.* at 790 (Scalia, J., dissenting).

92. *Id.*

93. *Id.*

94. *Id.* Justice Scalia listed burglary as an example, which in his opinion "connotes in common parlance the entry of a building with felonious intent." *Id.* He explained that, despite this common understanding, an indictment for burglary must set forth those elements. *Id.*

95. *Id.* (quoting United States v. Carll, 105 U.S. 611, 612 (1881)).

96. *See id.*

97. *Id.*

98. *Id.*

99. *Id.*

100. *Id.*

101. *Id.* A leading criminal law treatise, nearly one hundred years ago, characterized attempt as "indefinite" and lacking any "prescribed legal meaning." *Id.* (quoting FRANCIS WHARTON, A TREATISE ON CRIMINAL LAW 298 (James M. Kerr ed., Bancroft-Whitney 1912) (1846)).

102. *Id.* at 791. Justice Scalia applied the majority's reasoning to first degree murder,

Justice Scalia also disagreed with the majority's reliance on *Hamling v. United States*.[103] The *Hamling* Court had stated that "the various component parts of the constitutional definition of obscenity need not be alleged in the indictment" because obscenity has a definite legal meaning.[104] Justice Scalia recognized the absurdity of an indictment in which "every word contained within the definition of each element of a crime were *itself* an element of the crime within the meaning of the indictment requirement[.]"[105] However, he argued that *Hamling* would only be applicable to the case at hand if the indictment in *Hamling* had failed to allege the element of obscenity, just as the indictment in *Resendiz-Ponce* failed to allege the element of an overt act.[106] Justice Scalia explained that while "obscenity" is clearly one of the elements of the crime of publishing obscenity, the "various component parts of the constitutional definition of obscenity" are not.[107] Thus, while the Government still must prove the individual elements of obscenity in order to obtain a valid conviction for publishing obscenity, it need not specify them in the indictment.[108] Justice Scalia asserted that, in contrast, intent and an overt act are the two elements of attempt, and consequently, they must be identified in the indictment.[109]

Justice Scalia also found fault with the majority's assertion that "[i]f a defendant indicted only for a completed offense can be convicted of attempt . . . without the indictment ever mentioning an overt act, it would be illogical to dismiss an indictment charging 'attempt' because it fails to allege such an act."[110] Justice Scalia disagreed by explaining that a sufficient indictment for the commission of a completed offense "must persuade the grand jury that the accused's acts and state of mind fulfilled all the elements of the offense."[111] If the Government is successful, then the elements of an attempt offense will also have been fulfilled because attempt is a lesser-included offense.[112] In other words, "[a] grand-jury finding that the accused committed the crime is *necessarily* a finding that he attempted to commit the crime, and therefore the attempt need not be separately charged."[113] Justice Scalia asserted that when the Government seeks only an indictment for attempt, it must identify the

---

reasoning that "malice aforethought" could not be omitted from the indictment merely because it is understood in common usage and has always been a required element of the crime. *Id.*

103. *See id.*

104. Hamling v. United States, 418 U.S. 87, 118-19 (1974).

105. *Resendiz-Ponce*, 127 S. Ct. at 791 (Scalia, J., dissenting).

106. *See id.* at 792.

107. *Id.* at 791.

108. *Id.*

109. *See id. See generally* Almendarez-Torres v. United States, 523 U.S. 224, 228 (1998) ("An indictment must set forth each element of the crime that it charges.").

110. *Resendiz-Ponce*, 127 S. Ct. at 789 n.7 (majority opinion).

111. *Id.* at 792 (Scalia, J., dissenting).

112. *See id.*

113. *Id.*

specific elements of attempt.[114] Justice Scalia concluded by noting that he might have been willing to recognize an exception to what he considered the clear position of the law if the Government had shown that "mere recitation of the word 'attempt' in attempt indictments has been the traditional practice."[115] But because no attempt exception has been previously established by case law, Justice Scalia ultimately sided with "the general principles of our jurisprudence," which according to his interpretation, demand that attempt indictments allege both the elements of intent and an overt act.[116]

IV. IMPLICATIONS OF UNITED STATES V. RESENDIZ-PONCE

The United States Supreme Court held in *Resendiz-Ponce* that an indictment for attempted illegal reentry into the United States in violation of 8 U.S.C. § 1326(a) is not defective because it fails to allege a specific overt act.[117] The Court focused on this narrower constitutional issue rather than the question upon which it originally granted certiorari—namely, whether the omission of one or more elements of a crime from an indictment is structural error or whether it is amenable to harmless error analysis.[118] Thus, while the question upon which the Court granted certiorari was rather broad and not specific to the particular facts of *Resendiz-Ponce*, the Court's eventual holding addressed only the sufficiency of indictments under 8 U.S.C. § 1326(a).[119] An evaluation of the scope of *Resendiz-Ponce* therefore reveals that the Court did not intend to extend its holding to all attempt crimes, rather only to violations of 8 U.S.C. § 1326(a).[120] However, based on the majority's assertion that an overt act is commonly understood to be encompassed by the word *attempt*,[121] it follows that an indictment for any attempt crime that fails to expressly state an

---

114. *Id.*

115. *Id.*

116. *Id.* at 792-93. Justice Scalia noted that the question on which the Court granted certiorari was "whether a constitutionally deficient indictment is structural error . . . or rather is amenable to harmless-error analysis." *Id.* at 793. Because he wrote the lone dissenting opinion, he alone had to decide that question. *Id.* Because the Court decided a different constitutional issue in *Resendiz-Ponce*, Justice Scalia recognized that the Court undoubtedly will have to decide the issue upon which it granted certiorari at a later time. *Id.* Therefore, Justice Scalia declined to outline his views in depth, but he noted summarily that he would find the error to be structural, consistent with his opinions in *United States v. Gonzalez-Lopez*, 126 S. Ct. 2557 (2006) (majority opinion), and *Neder v. United States*, 527 U.S. 1, 26 (1999) (Scalia, J., concurring). *Resendiz-Ponce*, 127 S. Ct. at 793 (Scalia, J., dissenting). He therefore would have affirmed the judgment of the Ninth Circuit. *Id.*

117. *Id.* at 787-88 (majority opinion).

118. *See id.* at 785-86.

119. *See id.* at 787-88 ("Consequently, an indictment alleging attempted illegal reentry *under § 1326(a)* need not specifically allege a particular overt act or any other 'component par[t]' of the offense." (emphasis added) (alterations in original)).

120. *See id.*

121. *Id.* at 787.

overt act is nevertheless sufficient. The majority did not argue, after all, that an overt act is implied only for the crime of attempted illegal reentry. Instead, the majority states generally that the word *attempt* encompasses both the overt act and intent elements.[122]

Perhaps as a precaution, the majority warned that some crimes must be charged with greater specificity in the indictment than simple duplication of the language of a criminal statute can provide.[123] The Court pointed to *Russell* as an example,[124] in which the Court had held that an indictment charging a witness for refusal to answer a question before a congressional committee must go beyond a recitation of the statutory language and include the subject matter of the hearing.[125] However, the crime at issue in *Russell* was not an attempt crime; it therefore offers no additional insight into the intended boundaries of the majority's opinion in *Resendiz-Ponce*.

The majority's desire to eliminate the detailed and formalized style of common law pleadings, in keeping with the original aim of the Federal Rules of Criminal Procedure, is a goal worthy of pursuit. Justice cannot be served, after all, if those who are guilty of committing crimes escape punishment because of a technicality in the language of an indictment. But this goal of simplifying procedure threatens the more important constitutional requirement that an indictment contain the elements of the offense charged so that a defendant is fully informed of the charge against which he or she must defend. The majority in *Resendiz-Ponce* seemed determined to prevent the defendant, who was unquestionably primed to attempt illegal reentry into the United States, from escaping conviction on a technicality. But the majority allowed its zest to punish a wrongdoer to outweigh the fact that the defendant was not properly notified of the particular overt act against which he was expected to defend. Ultimately, the risk of denying a defendant his or her constitutional right of notice is a far greater offense than allowing a defendant to escape conviction for an attempt crime on a technicality.

The Court's holding in *Resendiz-Ponce* creates a dangerous slippery slope, as the bounds of its analysis are unclear. Indeed, Justice Scalia's fear that the majority's holding will "effect[] a revolution in our jurisprudence regarding the requirements of an indictment"[126] will quickly come to fruition if courts are permitted to assume that various elements of an offense are present in an indictment when they are not expressly stated. If an overt act can be inferred from the mere presence of the word *attempt*, there is likewise nothing to prevent the intent element from being implied as well. In addition, the Court is now vulnerable to challenges for other criminal offenses besides attempted illegal reentry. The only restriction the Court places on its holding is that

122. *Id.*
123. *Id.* at 789.
124. *Id.*
125. Russell v. United States, 369 U.S. 749, 764 (1962).
126. *Resendiz-Ponce*, 127 S. Ct. at 791 (Scalia, J., dissenting).

"there are some crimes that must be charged with greater specificity[,]"[127] but the Court did not clarify the crimes to which this standard would apply. By failing to more carefully limit the bounds of its holding, the Court makes itself susceptible to future challenges based upon similarly ambiguous indictments for other criminal offenses.

Not surprisingly, only one day after the release of the *Resendiz-Ponce* opinion, the Court was presented with another case in which one side sought to take advantage of definitional semantics. The Solicitor General filed a supplemental brief in a Ninth Circuit case raising the question that was avoided in *Resendiz-Ponce*—"whether the omission of an element of the offense from a federal indictment can constitute harmless error."[128] In *United States v. Omer*, the Ninth Circuit overturned the defendant's conviction for bank fraud because the indictment failed "to recite an essential element of the charged offense—materiality of falsehood."[129] The Solicitor General argued that the Court should not grant review of the case because, given its decision in *Resendiz-Ponce*, the indictment was not constitutionally deficient for failing to separately allege that the fraudulent scheme at issue was "materially false or deceptive."[130] Presumably the Solicitor General's rationale was that, just as Justice Scalia warned, if certain elements of an attempt crime can be implied from the mere presence of the word *attempt*, then the element of "misrepresentation or concealment of material fact" can be implied from the mere presence of the word *fraud*. In his statement respecting the denial of the petition for writ of certiorari in *Omer*, Justice Scalia sharply criticized the *Resendiz-Ponce* decision, characterizing it as the Court's "new some-crimes-are-self-defining jurisprudence."[131] Evaluating the potential effect of *Resendiz-Ponce*, Justice Scalia warned that "another frontier of law" had been opened by the Court, "full of opportunity and adventure for lawyers and judges."[132]

LINDSEY M. VAUGHAN

127. *Id.* at 789 (majority opinion).

128. United States v. Omer, 127 S. Ct. 1118 (2007) (mem.) (denying certiorari) (statement of J. Scalia concerning the denial of the petition for writ of certiorari).

129. United States v. Omer, 395 F.3d 1087, 1088 (9th Cir. 2005) (per curiam).

130. *Omer*, 127 S. Ct. at 1118.

131. *Id.* at 1119.

132. *Id.*

THE EVOLUTION OF ABA STANDARDS FOR CLINICAL FACULTY

PETER A. JOY* & ROBERT R. KUEHN**

I. INTRODUCTION

The value of clinical legal education courses and the faculty teaching those courses has long been contested. A focal point for this opposition has been resistance to the American Bar Association (ABA) accreditation standard that requires law schools to establish long-term employment relationships with clinical faculty and provide them with a meaningful voice in law school governance.[1] By integrating clinical faculty into law schools, the ABA aims to advance the value of clinical legal education and the professional skills and values that it promotes. In the decades since the ABA created the first clinical faculty standard, clinical legal education in the United States has developed as pedagogy and the number of clinical faculty has greatly increased. Despite these trends, a recent decision by the ABA Accreditation Committee approving short-term contracts and the denial of meaningful participation in faculty governance for clinical faculty demonstrates that the debate over the value of clinical legal education and the appropriate status for its faculty continues.[2] In

---

* Professor of Law and Director of the Criminal Justice Clinic, Washington University School of Law - St. Louis.

** Professor of Law and Associate Dean for Skills Programs, University of Alabama School of Law. The authors thank Ellen Mathews for her excellent research assistance and David Barnhizer, Bob Dinerstein, John Elson, Jeff Giddings, Steve Leleiko, Wally Mylniec, Richard Neumann, Gary Palm, Dean Rivkin, Roy Stuckey, and Karen Tokarz for their comments and suggestions. The authors also thank Mary Wolf who assisted in securing access to many of the historical documents cited in this article.

1. The current standard provides: "A law school shall afford to full-time clinical faculty members a form of security of position reasonably similar to tenure, and non-compensatory perquisites reasonably similar to those provided other full time faculty members." SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Standard 405(c) (2007) [hereinafter 2007 STANDARDS].

2. The ABA Accreditation Committee approved the Northwestern University School of Law's practice of restricting most clinical faculty to one-year employment contracts and denying those clinical faculty the participation in law school governance accorded other full-time law faculty. Letter and Decision of the Am. Bar Ass'n Accreditation Comm. from Hulett H. Askew, Consultant on Legal Educ. to the Am. Bar Ass'n, to Dr. Henry S. Bienen, President, Northwestern Univ., and David E. Van Zandt, Dean, Northwestern Univ. School of Law (Nov. 15, 2006) (on file with authors). Accreditation Committee actions are kept confidential by the ABA, but Dean David Van Zandt of Northwestern University School of Law released the decision of the Committee on a law school dean listserv. There is also a report that the Accreditation Committee approved one-year contracts for clinical faculty at St. Louis University

this debate, there is often little to no mention of the history of the accreditation standard in question, perhaps because no historical account of its evolution exists. In this article, we fill that gap in the literature by tracing the evolution of the ABA standard concerning clinical faculty status.

Part II begins with a discussion of the role of the ABA in legal education and provides a brief history of the development of clinical legal education. In Part III, we discuss the events leading up to the initial adoption in 1984 of a standard addressing clinical faculty and the reasoning that animated the ABA. In Part IV, we discuss the events leading to the strengthening of the standard in 1996 and the arguments opposing the more meaningful integration of clinical faculty into law schools. In Part V we discuss changes to the standard in 2005 and how those changes have revived the debate of the status of clinical faculty. Finally, in Part VI we discuss the current debate over clinical faculty status and the ongoing activities of various ABA groups examining the status of clinical faculty. It is our hope that by surfacing the historical debates and the evolution of the standard for clinical faculty, this article will provide the basis for reasoned, informed decisions by the ABA and the legal academy concerning the value of clinical legal education and the role of clinical faculty in law schools.

II. A BRIEF HISTORY OF THE DEVELOPMENT OF CLINICAL LEGAL EDUCATION

A. ABA's Early Role in Legal Education

The casebook method emerged at the end of the nineteenth century as the most popular way to teach in American law schools.[3] Its emphases on appellate decisions and the Socratic method signaled a shift from the applied skills training method inherent in the apprenticeship system that had been the dominant route to entry into the legal profession for more than two hundred years.[4] As academic legal education expanded rapidly starting in the 1890s, the

---

School of Law, though no one has publicly released such a decision. Paulette J. Williams, *President's Message*, CLEA NEWSLETTER (Clinical Legal Educ. Ass'n, New York, N.Y.), Feb. 2007, at 1, 2 (on file with authors).

3. *See* Charles R. McManis, *The History of First Century American Legal Education: A Revisionist Perspective*, 59 WASH. U. L.Q. 597, 598 (1982).

4. In the mid-19th Century, there were three prevailing methods in the United States for teaching law: the applied skills training approach inherent in the apprenticeship system; the general education approach of the European legal education model, adopted by some colleges and universities in the United States such as William & Mary; and "an analytical and systematized approach to law" as interconnected rational principals taught primarily through lectures at proprietary law schools. *Id.* Of these three approaches, the apprenticeship system was the most dominant until the end of the nineteenth century. *See* RICHARD L. ABEL, AMERICAN LAWYERS 42 (1989).

apprenticeship system essentially disappeared as a way to enter the legal profession.[5]

Around this same time, a small number of lawyers from almost two-thirds of the states founded the ABA, and the organization made advocating for formal legal education to better prepare students for the practice of law one of its founding objectives.[6] At its fourth annual meeting in 1881, the ABA passed resolutions promoting a three-year course of study of law, a bar examination requirement for admission to practice, and a policy that "time spent in any chartered and properly conducted law school, ought to be counted in such state as equivalent to the same time spent in an attorney's office in such state, in computing the period of study prescribed for applicants for admission to the Bar."[7] These ABA initiatives were largely aimed at tightening entry requirements into the legal profession, which was quickly growing due to the rapid rise of law schools, especially those operated as part-time enterprises.[8]

Into the first decade of the 1900s, the ABA discussed a wide range of topics affecting legal education, including prerequisites for admission to law school, the need for a three-year course of study, the contents of curriculum, and the role of practice experiences in legal education.[9] To further its efforts toward establishing bar admission requirements and to build alliances with law schools and law professors, the ABA invited delegates from select law schools to attend a meeting in 1900.[10] Thirty-five law schools sent delegates and they formed the Association of American Law Schools (AALS).[11] The ABA and the AALS asserted that their common cause was to advance law school

---

5. The reasons for this transformation are many. Apprenticeships were scarce, especially outside larger urban areas, and many existing lawyers did not welcome apprentices who were from the rising immigrant population with different ethnic, religious, or class backgrounds. ABEL, *supra* note 4, at 43. Lawyers also began to hire permanent clerks rather than take on apprentices. *Id.* Although states started to require bar exams, they often granted admission via a "diploma privilege" for law school graduates. *Id.* The combination of these factors, plus the relative low cost of attending law schools of that era, helps to explain the rapid demise of the apprenticeship system in the United States. *Id.*

6. EDSON R. SUNDERLAND, HISTORY OF THE AMERICAN BAR ASSOCIATION AND ITS WORK 5–10 (1953). The original ABA Constitution required the ABA President to appoint a Committee on Legal Education and Admissions to the Bar consisting of five members. *See Constitution*, 1 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 30–31 (1878).

7. *General Minutes*, 4 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 28 (1881).

8. ABEL, *supra* note 4, at 44. Richard Abel contends that the ABA was motivated by concerns that the number of lawyers was growing too rapidly and the status of lawyers was falling. *Id.* at 47. Abel notes that in addition to pushing for more rigorous law schools, in 1909 the ABA also sought to exclude non-citizens from entering the legal profession, a measure aimed at excluding recent immigrants from southern and eastern Europe. *Id.* at 68.

9. *See* SUNDERLAND, *supra* note 6, at 74–75.

10. Warren A. Seavey, *The Association of American Law Schools in Retrospect*, 3 J. LEGAL EDUC. 153, 157 (1950).

11. *Id.*

education.[12] These organizations primarily focused on classroom-based education that emphasized the teaching of legal doctrine and analysis.[13]

The emphasis on teaching legal doctrine and reasoning grew almost to the exclusion of experiential education. In the late 1890s and early 1900s, only a handful of law schools had established "legal dispensaries"[14] or had programs in which students worked with local legal aid offices, and those programs that did exist were largely non-credit, volunteer experiences.[15]

In a 1921 study of legal education, the Carnegie Foundation for the Advancement of Teaching identified three components necessary to prepare students for the practice of law: general education, theoretical knowledge of the law, and practical skills training.[16] The study found that legal education in the United States at the first part of the twentieth century lacked the "clinical facilities or shopwork provided by modern medical and engineering schools" and that there was no "foreign country in which education for the practice of law is so largely theoretical as it is in America."[17] The study noted that "[t]he failure of the modern American law school to make any adequate provision in its curriculum for practical training constitutes a remarkable educational anomaly."[18]

Despite this critique of legal education, law schools continued to resist practical skills training. From the 1920s to the 1940s, law faculty disagreed about the value and feasibility of teaching lawyering skills other than legal analysis.[19] During this same time period, the ABA and the AALS pushed

---

12. *See generally id.* at 154–63 (describing the birth and foundational missions of the ABA and AALS).

13. *See id.* at 155–59, 171–73.

14. *See generally* Jerome Frank, *Why Not a Clinical Lawyer-School?*, 81 U. PA. L. REV. 907, 917–18 (1933) (using the term "dispensary" to argue that law schools should offer clinical experiences much like those at work in medical education). Harvard Law School termed its in-house clinic a "legal aid bureau." *See* John S. Bradway, *The Nature of a Legal Aid Clinic*, 3 S. CAL. L. REV. 173, 174–75 (1930).

15. Law schools with volunteer legal aid bureaus or programs designed to involve law students with legal aid offices included law schools at Cincinnati, University of Denver, Harvard, University of Louisville, University of Memphis, Minnesota, Northwestern, University of Pennsylvania, Southern California, University of Tennessee, Washington University - St. Louis, Wisconsin, and Yale. Bradway, *supra* note 14, at 174; Robert MacCrate, *Educating a Changing Profession: From Clinic to Continuum*, 64 TENN. L. REV. 1099, 1102–03 (1997). A handful of law schools connected their legal aid programs to courses for credit or required participation of all third-year students. *See* Bradway, *supra* note 14, at 175–77.

16. ALFRED ZANTZINGER REED, CARNEGIE FOUND. FOR THE ADVANCEMENT OF TEACHING, TRAINING FOR THE PUBLIC PROFESSION OF THE LAW: HISTORICAL DEVELOPMENT AND PRINCIPAL CONTEMPORARY PROBLEMS OF LEGAL EDUCATION IN THE UNITED STATES WITH SOME ACCOUNTS OF CONDITIONS IN ENGLAND AND CANADA 276–78 (1921). This early Carnegie study of legal education is referred to as the Reed Report.

17. *Id.* at 281.

18. *Id.*

19. *See* ROBERT STEVENS, LAW SCHOOL: LEGAL EDUCATION IN AMERICA FROM THE 1850S

efforts to create and raise accreditation standards for law schools, yet none of the standards encouraged clinical legal education experiences.[20]

B. Academic Support for Clinical Legal Education From the 1930s–1950s

In spite of the lack of support for clinical legal education by the ABA and AALS early in the twentieth century, some of the most respected members of the legal academy were critical of law schools' exclusive reliance on what became known as the "casebook method" and doctrinal analysis.

In the 1930s and 1940s, Jerome Frank extolled the need for and virtues of clinical legal education.[21] A 1944 report of the AALS Curriculum Committee, primarily authored by Karl Llewellyn, noted that the casebook method was "failing to do the job of producing *reliable professional competence* on the by-product side *in half or more of our end-product*, our graduates."[22] In 1951, Robert Storey, then Dean of Southern Methodist University School of Law, praised the "clinical method" for exposing "the student to actual problems by confronting him with actual people who are in actual trouble."[23]

Although some law school deans and faculty saw the potential for clinical legal education to teach students a range of lawyering skills and professional values, only 35 of 126 ABA-approved law schools offered clinical experiences by the late 1950s.[24] The clinical experiences offered were typically volunteer activities for both students and faculty. Only fifteen of the thirty-five schools with clinical experiences by the late 1950s awarded limited academic credit to students for their clinical work, and only five law schools gave supervising law faculty teaching credit for their clinical courses.[25]

C. Expansion of Clinical Legal Education in the 1960s

Fueled by grant support, clinical programs began to grow significantly in the 1960s. From 1968 to 1978, the Council on Legal Education for Professional Responsibility (CLEPR), funded by the Ford Foundation, awarded grants for clinical programs to 107 ABA-approved law schools.[26] Professor

---

TO THE 1980S, at 214 (1983).

20. *See id.* at 172–80.

21. *See* Jerome Frank, *A Plea for Lawyer-Schools*, 56 YALE L.J. 1303, 1312–16 (1947); Frank, *supra* note 14, at 917.

22. Karl N. Llewellyn et. al., *Report of the Committee on Curriculum*, 1944 ASS'N OF AM. LAW SCHS. PROC. 159, 168, *quoted in* STEVENS, *supra* note 19, at 214.

23. Robert G. Storey, Foreword, *Law School Legal Aid Clinics*, 3 J. LEGAL EDUC. 533, 533 (1951).

24. *See* Joseph W. McKnight, *Report of Committee on Legal Aid Clinics*, 1959 ASS'N OF AM. LAW SCHS. PROC. 121, 121–22.

25. *Id.* at 122.

26. *See* John M. Ferren, *Prefatory Remarks*, 29 CLEV. ST. L. REV. 351, 352 (1980).

Charles Miller, who started the University of Tennessee Legal Clinic in 1947,[27] noted that the professional responsibility emphasis in CLEPR-funded programs stressed the need for a law student to assume the "lawyer role" as a necessary step to learn how to become an ethical practitioner.[28]

Funding to develop or expand clinical programs continued through U.S. Department of Education Title IX grants from 1978-97.[29] By the end of the Title IX program, there were real-client, in-house law school clinical programs in at least 147 of the 178 ABA approved law schools.[30] Today, every ABA-approved law school offers in-house clinical courses, externships, or both.[31]

The growth of clinical legal education programs during the 1970s and 1980s also was paced by the development of clinical teaching methodology. Clinical faculty of this era began to construct a "common vocabulary of discourse on educational issues"[32] and saw teaching students how to learn from experience as a primary goal of clinical legal education.[33]

D. The Debate Over the Value of Clinical Legal Education and the Status of Clinical Faculty

Despite progress in key areas, proponents of clinical legal education continued to encounter resistance in law schools. The critics' rationalization

---

27. Douglas A. Blaze, *Déjà Vu All Over Again: Reflections on Fifty Years of Clinical Education*, 64 TENN. L. REV. 939, 939 (1997).

28. Charles H. Miller, *Living Professional Responsibility—Clinical Approach*, *in* CLINICAL EDUCATION FOR THE LAW STUDENT: LEGAL EDUCATION IN A SERVICE SETTING 99, 99 (1973).

29. Margaret Martin Barry, Jon C. Dubin & Peter A. Joy, *Clinical Education for This Millennium: The Third Wave*, 7 CLINICAL L. REV. 1, 19–20 (2000).

30. *Id.*; *see* Am. Bar Ass'n, Section of Legal Educ. and Admissions to the Bar, ABA-Approved Law Schools by Year (last visited Mar. 24, 2008), http://www.abanet.org/legaled/approvedlawschools/year.html.

31. *See generally* LAW SCHOOL ADMISSIONS COUNCIL & AM. BAR ASS'N, 2008 ABA-LSAC OFFICIAL GUIDE TO ABA-APPROVED LAW SCHOOLS (Wendy Margolis et al. eds., 2007) (listing clinical course offerings for all ABA-approved law schools). ABA accreditation Standard 302(b)(1) requires every law school to offer substantial opportunities for "live-client or other real-life practice experiences, appropriately supervised and designed to encourage reflection by students on their experiences and on the values and responsibilities of the legal profession, and the development of one's ability to assess his or her performance and level of competence." 2007 STANDARDS, *supra* note 1, at Standard 302(b)(1). "The offering of live-client or real-life experiences may be accomplished through clinics or field placements. A law school need not offer these experiences to every student nor must a law school accommodate every student requesting enrollment in any particular live-client or other real-life practice experience." *Id.* at Interpretation 302-5.

32. Gary Bellow, *On Teaching the Teachers: Some Preliminary Reflections on Clinical Education as Methodology*, *in* CLINICAL EDUCATION FOR THE LAW STUDENT: LEGAL EDUCATION IN A SERVICE SETTING, *supra* note 28, at 374, 375.

33. *See* Barry, Dubin & Joy, *supra* note 29, at 17.

was that law graduates would learn lawyering skills and values when they entered practice,[34] an argument that some still make today. This attitude belied the reality that most law graduates were not receiving postgraduate mentoring and that "even in the best settings and with the best of tutors . . . certain commercial or institutional forces" interfered with the learning process.[35] As a result, CLEPR and its supporters viewed law schools as the best venue for future lawyers to learn lawyering skills and the professional responsibilities of the legal profession.[36] An early CLEPR newsletter explained: "In the law school, removed from the necessity to earn a fee, the law student has [the] best and possibly [the] only opportunity to learn about managing a proper commitment to a client and his cause."[37]

At a CLEPR workshop in 1971, participants discussed the challenges of starting and running clinical programs.[38] They concluded that in consideration of the indifference or resistance of many traditional faculty members, a good clinical program required that a clinical faculty member have "security at the law school and . . . prestige among his colleagues."[39] Without the security of a continuing employment relationship with the law school and a voice in law school governance, clinical faculty and the courses they taught were marginalized.

By the mid-1970s, CLEPR sought to enhance the status of faculty teaching clinical courses by a series of grants to raise clinical faculty salaries "to parity with classroom teachers" as a step to "eliminate one of the most serious handicaps in recruitment and retention of qualified clinical supervisors . . . ."[40]

---

34. William Pincus, *Educational Values in Clinical Experience for Law Students*, CLEPR NEWSLETTER (Council on Legal Educ. for Prof. Resp., Inc., New York, N.Y.), Sept. 1969, at 1–2.

35. *Id.* at 2.

36. *See id.*

37. *Id.* at 3.

38. *See CLEPR Holds Workshops on Life and Times of the Clinical Law Professor*, CLEPR NEWSLETTER (Council on Legal Educ. for Prof. Resp., Inc., New York, N.Y.), Nov. 1971, at 1.

39. *Id.* at 3.

40. *Parity between Clinical and Academic Salaries Supported by New CLEPR Grants to Two Law Schools*, CLEPR NEWSLETTER (Council on Legal Educ. for Prof. Resp., Inc., New York, N.Y.), Jan. 1977, at 1. In January 1977, CLEPR awarded a grant to Northwestern University, which was matched by law school funds, to increase the salaries of five-tenure track clinical faculty to establish salary parity with non-clinical faculty. *Id.* at 1. Northwestern pledged to maintain the salary parity in future budgets, and CLEPR noted Northwestern's "pioneer role in establishing new promotion and tenure criteria which take into account the special demands of clinical teaching." *Id.* at 1–2. It is ironic that today Northwestern has relegated most clinical faculty to short-term contracts and pushed the ABA Accreditation Committee for a ruling approving of this inequitable treatment of clinical faculty compared to non-clinical faculty. *See supra* note 2. The other school to receive a salary parity grant in January 1977 was the University of Tennessee, which received funds to increase the salaries of twelve attorney/instructors. *Id.* at 2. In May through July 1977, CLEPR awarded additional

At a series of conferences held by CLEPR in 1978, ineligibility for tenure emerged as the most fundamental difference between clinical and non-clinical faculty.[41] The participants saw the questions of status and working conditions "inextricably tied to the law school's basic perception of and commitment to clinical education."[42] As long as law schools granted tenure to academic teachers, the participants at the conference "agreed that if clinicians are to be truly equal members of the law school community . . . they should be considered for and granted tenure on the basis of demonstrated excellence . . . ."[43]

As clinical programs became more prevalent in the 1970s, the status of faculty members teaching clinical courses became a matter of some debate, not just among clinical faculty, but also within the legal academy and ABA. By the end of the decade, many outside the legal academy were calling for law schools to establish long-term employment commitments to clinical faculty and to integrate clinical faculty into the governance of the law school as a means to further the development of clinical courses. The following section discusses the development of ABA Accreditation Standard 405(e), the first standard directed toward the status of clinical teachers.

III. ADOPTION OF THE INITIAL ACCREDITATION STANDARD 405(E)

Long before the ABA first adopted an accreditation standard on the status of law school faculty teaching clinical courses, leaders of the legal profession and reports on legal education repeatedly expressed concerns over what they considered the unfair treatment of clinical faculty and its negative effects on the development of clinical legal education.

---

salary parity grants to Hofstra (for six full-time clinical supervisors), University of New Mexico (for five clinical faculty), New York University (for twelve clinical faculty), Rutgers University-Newark (for four clinical faculty), and Yale University (for four clinical positions). *New CLEPR Grants Give Priority to Parity for Clinicians and to a Study for Clinic Guidelines*, CLEPR NEWSLETTER (Council on Legal Educ. for Prof. Resp., Inc., New York, N.Y.), Sept. 1977, at 1.

41. Laura Sager, *Career Perspectives for Clinical Teachers (A First Report)*, CLEPR NEWSLETTER (Council on Legal Educ. for Prof. Resp., Inc., New York, N.Y.), Apr. 1978, at 2.

42. *Id.* at 4. A CLEPR report on the second and third conferences noted that "the question of tenure for clinicians is a difficult and controversial one" because "[t]he tenure system itself is now under attack from many quarters and may ultimately be abandoned by the universities and law schools." Laura Sager, *More on Career Perspectives for Clinical Teachers*, CLEPR NEWSLETTER (Council on Legal Educ. for Prof. Resp., Inc., New York, N.Y.), Oct. 1978, at 2.

43. *Id.* at 2–3.

A. Reports on Legal Education Favored Improved Status for Clinical Faculty

The first ABA report to identify the importance of clinical faculty status in legal education was the 1979 report "Lawyer Competency: The Role of Law Schools" (known as the "Cramton Report").[44] The Cramton Report identified a series of institutional factors inhibiting improved law school training for the legal profession and recommended that law schools place greater value on skill development and on the faculty teaching lawyering skills, arguing:

> Law school policies and practice of faculty appointment, promotion, and tenure should pay greater rewards for commitment to teaching, including teaching by techniques that foster skills development. Experimentation with and creation of new teaching methods and materials that focus on the improvement of such fundamental lawyer skills as legal writing, oral communication, interviewing and counseling, or trial advocacy should be valued no less highly than research on legal doctrine.[45]

In 1980, another ABA study on legal education, the Foulis Report, observed that "the status of clinicians in the academic setting has not been satisfactorily resolved" and recommended "that appropriate weight be assigned to the effective teaching of legal skills."[46]

While these two independent reports on legal education were reaching similar conclusions that the status of clinical faculty should be improved to secure the development of clinical legal education, a joint committee of the ABA and the AALS was developing law school guidelines for clinical legal education that were based on similar conclusions.[47] This study resulted in the

---

44. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, REPORT AND RECOMMENDATIONS OF THE TASK FORCE ON LAWYER COMPETENCY: THE ROLE OF THE LAW SCHOOLS (1979). The report is named after the chair of the twelve-person task force, Dean Roger C. Cramton of Cornell Law School. The task force included three judges, one university president, two law school deans, a law professor, and five attorneys. Nine members of the task force were present or former members of the ABA Section of Legal Education and Admissions to the Bar. *Id.* at vii. Three members of Committee were later Chairs of the Council of the ABA Section on Legal Education and Admissions to the Bar: Willard L. Boyd, 1980–81; Gordon D. Schaber, 1981–82; and Robert B. McKay, 1983–84.

45. *Id.* at 26.

46. AM. BAR ASS'N, LAW SCHOOLS AND PROFESSIONAL EDUCATION: REPORT AND RECOMMENDATIONS OF THE SPECIAL COMMITTEE FOR A STUDY OF LEGAL EDUCATION OF THE AMERICAN BAR ASSOCIATION 9, 105 (1980). This report is referred to as the "Foulis Report" after Ronald J. Foulis, the chair at the time the report was issued. The report was the final product of a seven-year study of legal education. *Id.* at vii, 1.

47. *See generally* AM. BAR ASS'N & ASS'N OF AM. LAW SCHS., CLINICAL LEGAL EDUCATION: REPORT OF THE ASSOCIATION OF AMERICAN LAW SCHOOLS/AMERICAN BAR ASSOCIATION COMMITTEE ON GUIDELINES FOR CLINICAL LEGAL EDUCATION (1980) [hereinafter CLINICAL LEGAL EDUCATION GUIDELINES] (providing new guidelines that emphasized the

1980 joint ABA and AALS report "Clinical Legal Education," which included "guidance to law school faculties wishing to initiate clinical training programs or to evaluate existing programs."[48] The committee was independent and highly respected, and its conclusions were far reaching. The ABA and the AALS selected the members of the committee and no member was on the clinical faculty of a law school.[49] The committee chair was former law school dean Robert McKay, and the remaining members were a university president, two law school deans, two tenured non-clinical law school faculty members, and one member of the public.[50] The ABA and AALS issued guidelines providing law school deans and faculties with useful suggestions for the elements of sound clinical legal education programs and the reasoning for each guideline.[51]

The guidelines concerning faculty status for clinical faculty provided: "One or more of the faculty who have principal responsibility for the clinical legal studies curriculum should have the same underlying employment relationship as the faculty teaching in the traditional curriculum."[52] Further, the guidelines noted that in addition to clinical faculty with equal status, some "individual schools may wish to have some principal clinical teaching responsibilities fulfilled by individuals not eligible for tenure" due to "budgetary considerations" and "the experimental nature of clinical legal studies [at this time]," but "full-time positions not eligible for tenure should be long-term employment" if the non-tenure track clinical faculty were to develop expertise in clinical teaching, develop components of the curriculum, or supervise the training of other faculty who were also teaching clinical studies.[53]

---

importance of clinical legal education).

48. *Id.* at iii.

49. *See id.* at 3. Two non-voting staff members for the Committee were clinical faculty: Steven H. Leleiko, Assistant Dean and Associate Clinical Professor at New York University School of Law, served as Project Director; and David Barnhizer, Professor of Law at Cleveland-Marshall College of Law and Chair of AALS Section on Clinical Legal Education, served as Special Consultant. *Id.* at 4.

50. *Id.* at i, 3. At the time he served as committee chair, McKay, former dean of New York University School of Law, was Director of the Justice Program of the Aspen Institute. *Id.* at 3. The university president was Willard L. Boyd, University of Iowa. *Id.* at 3. The law school deans were David J. McCarthy, Jr., Georgetown University Law Center, and Gordon D. Schaber, McGeorge School of Law. Henry W. McGee, Jr., University of California School of Law - Los Angeles, and Norman Penney, Cornell University School of Law, were the non-clinical faculty members. *Id.* The public member was Thomas B. Stoel, Jr., an attorney with the Natural Resources Defense Council. *Id.*

51. *See id.* at 6.

52. *Id.* at 33 (noting that "[a]t most schools eligibility for tenure is the basic employment relationship").

53. *Id.* The guidelines stated:

> *Long-Term Employment Positions:* In addition to the foregoing faculty, individual schools may wish to have some principal clinical teaching responsibilities fulfilled by individuals not eligible for tenure. Reasons for having such nontenure-eligible positions

The guidelines explained that addressing the status issue was necessary because "the importance of clinical legal studies to the law school curriculum requires the application of tenure status to individuals principally teaching in the clinical legal studies curriculum."[54] The guidelines recommended that at least one faculty member principally teaching in the clinic should have the same status as other faculty to satisfy the educational needs of the clinical program.[55] The guidelines anticipated that one reason law schools might want some clinical faculty on long-term contractual relationships rather than tenure was due to "the experimental and innovative nature of clinical legal studies, [and] schools considering or participating in its early development may not wish to commit themselves to a large number of tenure-track relationships."[56]

In drafting and explaining these guidelines, which were not designed as accreditation standards but rather to provide guidance for creating sound educational programs,[57] the drafters studied the state of clinical legal education in the late 1970s. The drafters expressed their belief that status equivalency between clinical and non-clinical faculty was important for the development of clinical legal education and that only the experimental nature of clinical legal education at many law schools justified the unequal security of position for

---

include, but are not limited to, budgetary considerations, the experimental nature of the clinical legal studies curriculum, and the professional responsibility to live clients. Such full-time positions not eligible for tenure should be long-term employment if the occupants are required to: (1) possess or develop an expertise in the theoretical and empirical literature related to the issues covered in the clinical studies curriculum and engage in the related teaching; (2) develop or teach classroom components of the clinical legal studies curriculum; and (3) supervise the training or teaching of other faculty or professional staff engaged in teaching the clinical legal studies curriculum.

Decisions Regarding the Status of Supervising Attorneys: Decisions regarding the status of full-time supervising attorneys should be made no later than the third year of their employment. During that year the law school should decide on one of the following as to each supervising attorney: (1) termination; (2) long-term appointment with change of status to an appropriately titled position (e.g., assistant clinical professor); or (3) placement on the tenure track.

Id. at 33–34.

54. *Id.* at 113.

55. *Id.*

56. *Id.* at 114. The reasoning was further explained in this way:

The Committee concluded that law schools must balance their concern for committing tenure-track slots to individuals in a field which is still young, comparatively underdeveloped, and experimental with the need to develop within the faculty individuals who have the expertise and experience necessary to successfully teach in the clinical legal studies curriculum. The Committee felt, therefore, that to help accomplish this, law schools wishing to limit the number of tenure track positions in the clinical legal studies curriculum could establish long-term employment positions.

Id. at 115.

57. *Id.* at 6.

some clinical faculty.[58] Even then, the guidelines stated that law schools should provide clinical teachers with long-term contractual relationships.[59]

In addition to the issue of security of position, the guidelines contemplated that the integration of clinical legal education into the curriculum required law schools "to avoid any isolation of clinical legal studies" and to provide clinical faculty "with the opportunity to participate in and contribute to such decision-making processes."[60] The guidelines stated that clinical courses and faculty should be treated on par with other law school courses and faculty, and "the failure to consider clinical legal studies in the context of the overall curriculum leads to a second-class status for clinical legal studies."[61] The guidelines stressed that the full integration of clinical studies into legal education required the full integration of clinical teachers into law faculties.[62] Not long after the clinical guidelines and two influential ABA reports favoring improved status for clinical faculty were issued, the ABA began to address the status of clinical faculty through accreditation standards.

B. The ABA's First Standard on the Status of Clinical Faculty

Starting in the late 1970s, ABA site inspection teams began "reporting to the accreditation committee that many schools were not providing their clinicians an opportunity to achieve tenure or any other form of job security."[63] Prior to the 1980s, ABA law school accreditation standards included a general standard on the competence of all members of the faculty but nothing that specifically addressed clinical faculty: "The law school shall establish and maintain conditions adequate to attract and retain a competent faculty."[64] At the time, Standard 405(d) provided that each law school "shall have an

---

58. *Id.* at 114.

59. *Id.*

60. *Id.* at 55.

61. *Id.* at 59. The guidelines further stated: "[C]linical legal studies should be considered in relation to the law school's overall educational objectives. To view clinical legal studies as part of an integrated law school curriculum requires an institutional perspective in which . . . individuals responsible for traditional and clinical studies are viewed and treated as members of the law school community . . . ." *Id.* at 17.

62. *Id.* at 59. The guidelines stated:

> Individuals teaching clinical legal studies are part of the law school community. They are entitled to the respect and collegiality traditionally accorded members of the law faculty. The importance of this to the development of good relations within the faculty is recognized in law school analyses of the role of the clinical teacher in the law school. The Committee intended to emphasize the importance of integrating those who teach clinical legal studies into traditional collegial activities.

Id.

63. Roy Stuckey, A Short History of Standard 405(e), at 1 (Apr. 1994) (unpublished manuscript) (on file with authors).

64. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Standard 405 (1983) [hereinafter 1983 STANDARDS].

established and announced policy with respect to academic freedom and tenure of which Annex I herein is an example but is not obligatory."[65] The ABA became concerned over site inspection reports indicating that some law schools did not consider clinical faculty covered by the academic freedom and tenure standard.[66]

In July 1980, the Council of the ABA Section of Legal Education and Admissions to the Bar (Council) acted on these reports that schools were not providing tenure opportunities for clinical faculty and adopted Interpretation 2 of Standard 405(d):

> Individuals in the "academic personnel" category whose full time is devoted to clinical instruction and related activities in the J.D. program constitute members of the "faculty" for purposes of Standard 405, and denial to them of the opportunity to allow tenure appears to be in violation of Standard 405(d).[67]

This Interpretation was suspended shortly thereafter "following a negative reaction from some law schools, and [the Council] created a subcommittee of the accreditation committee, chaired by Gordon Shaber, to consider how the problem should be resolved."[68]

In 1982, the ABA's Accreditation Committee and Clinical Legal Education Committee proposed to the Council that it adopt and submit to the House of Delegates a new Standard 405(e) and Interpretations.[69] The proposed Standard from the Accreditation Committee provided that "[f]ull-time clinical faculty members shall be entitled to an employment relationship substantially equivalent to that required for other members of the faculty under Standard

---

65. *See* Memorandum 7980-13 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of Approved Law Schools (Sept. 26, 1979) (on file with authors).

66. Memorandum D8384-51 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools 4 (May 22, 1984) [hereinafter Memorandum D8384-51] (on file with authors).

67. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Interpretation 2 of Standard 405(d) (1981) [hereinafter 1981 STANDARDS]. During this time period, the ABA House of Delegates had delegated to the Council the power to interpret accreditation standards. Dean Rivkin & Roy Stuckey, *Update on 405(e)*, CLINICAL LEGAL EDUCATION NEWSLETTER (Ass'n of Am. Law Schs., Wash., D.C.), June 1984, at 2–3.

68. Stuckey, *supra* note 63, at 1. The ABA Standards contained the Interpretation in 1981. *See* 1981 STANDARDS, *supra* note 67, Interpretation 2 of Standard 405(d). But by 1983 the Interpretation was no longer included in the published Standards. *See* 1983 STANDARDS, *supra* note 64, Interpretation 1 of Standard 405(d). The ABA did not publish a 1982 version of the Standards. Telephone Interview with Maxine A. Klein, Executive Assistant to the Consultant on Legal Educ. to the Am. Bar Ass'n (July 17, 2007).

69. Memorandum from Frederick R. Franklin, Staff Director, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Members of the Clinical Legal Educ. Comm. (April 28, 1982) (on file with authors); Memorandum D8384-51, *supra* note 66.

405."[70] The Interpretation explained that the employment relationship could be satisfied in one of three ways: (1) the same tenure track as the other members of the faculty; (2) a separate tenure track; or (3) "an approach that provides features substantially equivalent to tenure."[71] The Council considered the proposed Standard and Interpretation at its May 1982 meeting but chose not act on them.[72]

At its July 1982 meeting, the Accreditation Committee recommended that the Interpretation be revised to define employment relationships substantially equivalent to tenure as "[e]mployment contracts, such as successive renewable, long-term contacts that provide features substantially equivalent to tenure. The approach chosen shall also include terms and conditions of employment substantially equivalent to those offered to non-clinical, full-time members of the faculty."[73] The next month, the Council referred the new proposed Standard and Interpretation to the Standards Review Committee for consideration and recommendation.[74] The Committee then sought comments and held public hearings on the proposed language.[75]

In response, then President of the AALS, Berkeley Law School Dean Sanford H. Kadish, reported in November 1982 that the AALS "Executive Committee has been studying the proposal [for Standard 405(e)] for several months . . . . We see the issues as having considerable importance for the law

---

70. Memorandum from Frederick R. Franklin, *supra* note 69, at 1.

71. The Interpretation of proposed Standard 405(e) stated:

> Full-time clinical faculty members are entitled to an employment relationship substantially equivalent to that enjoyed by other members of the full-time faculty. This Standard may be satisfied by: (1) the inclusion of full-time clinical faculty on the same tenure track as the other members of the faculty; (2) a separate tenure track; or (3) an approach that provides features substantially equivalent to tenure. The law school bears the burden of establishing that its approach is substantially equivalent. This Standard is not meant to preclude employment of full-time clinical teachers on fixed, short-term employment relationships, for example, in situation where a law school receives a short-term grant to fund a clinic in a specific subject matter.

Id.

72. *See* Memorandum D8283-17 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Dec. 8, 1982) (on file with authors).

73. *Id.* at 2–3. The revised Interpretation to Standard 405(e) provided:

> Full-time clinical faculty members are entitled to an employment relationship substantially equivalent to that enjoyed by other members of the full-time faculty. This Standard may be satisfied by: (1) the inclusion of full-time clinical faculty on the same tenure track as the other members of the <u>full-time</u> faculty; (2) a separate tenure track; or (3) <u>Employment contracts, such as successive renewable, long-term contacts that provide features substantially equivalent to tenure. The approach chosen shall also include terms and conditions of employment substantially equivalent to those offered to non-clinical, full-time members of the faculty</u>.

Id. (underscores in original).

74. *Id.* at 3.

75. *Id.*

schools of the country and we intend to participate actively in their consideration and resolution."[76] In addition, *The Chronicle of Higher Education* reported that at the 1983 AALS Annual Meeting "[t]he academic status of the clinical faculty members—and what to do about it—commanded the attention of members of the law-school association for much of their meeting."[77] Professor Elliott Milstein, then Director of the Clinical Program at American University, argued:

> . . . law schools treat clinicians with something approaching disdain . . . . [T]he law schools withhold the symbols and perquisites of the profession from us. They deny us promotions and titles. They deny us voting rights and salaries of other faculty members. This leads to the myth that teaching lawyering skills is beneath the law schools.[78]

Professor Clinton Bamberger, who was then the Co-Director of Clinical Education at the University of Maryland, argued that the effort to propose alternatives to tenure for clinical faculty was "an effort to hold clinical faculty 'outside,' so the changes in the method of law-school teaching that we have encouraged will not be successful."[79]

Opposing the measure, the new President of the AALS Professor David H. Vernon of the University of Iowa characterized the ABA proposal as "premature," arguing that "the proposed standard is an invasion of traditional law-school territory. It is an expression of lack of confidence in the law schools. It implies that we are unfit to govern ourselves."[80] Vernon's opposition did not address the merits of clinical education or the necessity of giving job security as a means of both advancing the acceptance of clinical legal education and ensuring academic freedom for clinical faculty. Rather, Vernon cast the proposed accreditation requirement as an intrusion into law school self-governance and sought to reframe the debate to focus on law school autonomy versus the accreditation process and standards.

At its business meeting during the AALS 1983 Annual Meeting, the Clinical Education Section of the AALS passed the following resolution and forwarded it to the AALS Executive Committee:

> That the question of status of clinicians at the nation's law schools is an appropriate matter for an ABA Accreditation Standard; and that such a Standard should provide for the preservation and enhancement of high quality programs of clinical legal education by assuring clinicians academic freedom,

---

76. Sanford H. Kadish, *President's Message*, AALS NEWSLETTER (Ass'n of Am. Law Schs., Wash., D.C.), Nov. 1982, at 2.

77. Beverly T. Watkins, *Teachers of Clinical Law Seek Recognition, Better Treatment*, CHRON. HIGHER EDUC., Jan. 19, 1983, at 14.

78. *Id.*

79. *Id.*

80. *Id.*

> appropriate job security and equality of treatment with non-clinical law school faculty.[81]

In February 1983, the ABA Standards Review Committee met to consider the comments from several public meetings.[82] A memorandum to deans of ABA-approved law schools from James White, Consultant on Legal Education to the ABA, stated:

> [I]n light of the comments and views which were expressed by constituencies, the Committee did not formulate a recommendation at this time, but determined to continue to study the matter and request further assistance from law schools in developing a recommendation concerning the proposed amendment to the Standards relating to the status of clinical law teachers.[83]

At a July 1983 meeting, the Council deferred consideration of proposed Standard 405(e) until its December 1983 meeting, upon the recommendation of the Standards Review Committee.[84] The Committee planned to continue studying the issue and to present a report and recommendation to the Council at its December 1983 meeting.[85] At the Council meeting in July, Professors Dean Rivkin and Joe Harbaugh, both Council members and clinical teachers, expressed concerns over the delays in considering this issue.[86]

In the midst of the prolonged debate over the adoption of a clinical faculty accreditation standard, yet another independent ABA report expressed support for greater recognition of the contributions of clinical faculty. The ABA Task Force on Professional Competence, known as the Friday Report, stated:

> Consistent with the Foulis Report, we believe that the contributions of clinical teachers should receive greater and more appropriate weight than is now often the case. We believe that the distinctive role and workload of the clinical teacher should be recognized as a desirable and acceptable substitute

---

81. Memorandum from Kandis Scott, Chairperson, Ass'n of Am. Law Schs. Clinical Educ. Section, and Rodney Jones, Chairperson of Faculty Status Comm., Ass'n of Am. Law Schs. Clinical Educ. Section, to Members of the Ass'n of Am. Law Schs. Clinical Educ. Section (Jan. 24, 1983) (on file with authors).

82. Memorandum D8283-26 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Feb. 21, 1983) (on file with authors).

83. *Id.* In his memorandum, White requested more comments and "information from law schools which have adopted, or are considering adopting, policies or practices regarding the appointment of clinical faculty." *Id.*

84. Memorandum D8384-6 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Aug. 12, 1983) (on file with authors) (referring to an attached excerpt of draft minutes of the Council's July 1983 meeting).

85. *Id.*

86. *Id.*

> for the traditional scholarship of a law faculty member in tenure and promotion criteria.[87]

The Friday Report recommended that "[c]linical teachers should receive greater support for successful teaching in clinical settings than is now often the case"[88] and that the ABA should adopt a policy of including clinical law faculty on law school accreditation inspection teams.[89]

The Standards Review Committee met again in November 1983 and considered the proposals and comments concerning Standard 405(e).[90] The Committee unanimously recommended a version of Standard 405(e) that law schools "shall afford to full-time faculty members whose primary responsibilities are in its professional skills program, a form of security of position reasonably similar to tenure and perquisites reasonably similar to those provided full-time faculty members."[91] The accompanying Interpretation explained that security of position reasonably similar to tenure could be a "separate tenure track" or "[a] program of renewable long-term contracts . . . that shall thereafter be renewable."[92]

In addition, the Standards Review Committee proposed two additional Interpretations. First, law schools "should develop criteria for retention,

---

87. TASK FORCE ON PROF'L COMPETENCE, AM. BAR ASS'N, FINAL REPORT AND RECOMMENDATIONS OF THE TASK FORCE ON PROFESSIONAL COMPETENCE 12 (1983). This ABA report is sometimes referred to as the Friday Report, after its chair Herschel H. Friday.

88. *Id.* at 29.

89. *Id.*

90. Memorandum D8384-51, *supra* note 66, at 5.

91. *Id.* at 1. Proposed Standard 405(e) stated:

> The law school shall afford to full-time faculty members whose primary responsibilities are in its professional skills program, a form of security of position reasonably similar to tenure and perquisites reasonably similar to those provided other full-time faculty members by Standards 401, 402(b), 403 and 405. The law school shall require these faculty members to meet standards and obligations reasonably similar to those required of full-time faculty members by Standards 401, 402(b), 403 and 405.

Id.

92. *Id.* at 2. Interpretation A of proposed Standard 405(e) stated:

> A form of security of position reasonably similar to tenure includes a separate tenure track or a renewable long-term contract. Under a separate tenure track, a full-time faculty member, after a probationary period reasonably similar to that for other full-time faculty, may be granted tenure as a faculty member in a professional skills program. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the professional skills program.
>
> A program of renewable long-term contracts should provide that, after a probationary period reasonably similar to that for other full-time faculty, the services of the faculty member in a professional skills program may be either terminated or continued by the granting of a long-term contract that shall thereafter be renewable. During the renewal period the contract may be terminated for good cause, including termination or material modification of the professional skills program.

Id.

promotion and security of employment of full-time faculty members in its professional skills program."[93] Second, proposed "Standard 405(e) does not preclude fixed, short-term appointments in a professional skills program such as full-time visiting faculty members and full-time supervising attorneys."[94]

The ABA Standards Review Committee sent its report and recommendation to the Council along with additional materials consisting of a detailed analysis of the proposed Standard 405(e) prepared by Council member Rivkin, a copy of standards and procedures governing the status of clinical teachers at the Georgetown Law Center, and a background paper on the status of clinical faculty endorsed by the clinical faculty group at New York University Law School.[95] These documents provided examples of law schools establishing successful systems for integrating clinical faculty into law schools consistent with the proposed Standard 405(e).

At its December meeting, the Council decided to give notice of its intent to adopt Standard 405(e) and Interpretations substantially as proposed, including the "shall" language relative to "security of position reasonably similar to tenure."[96] As part of the process, the Council scheduled additional public hearings and solicited comments on the proposed Standard.[97] Among the comments submitted was a letter written by Dean Paul D. Carrington and signed by two other law school deans voicing opposition to the proposed clinical security of position standard.[98] The three deans argued that the

---

93. *Id.* at 3. Interpretation B of proposed Standard 405(e) stated:

> In determining if full-time faculty members in a professional skills program meet standards and obligations reasonably similar to those provided for other full-time faculty, competence in the areas of teaching and scholarly research and writing should be judged in terms of the responsibilities of the faculty member in the professional skills program. Each school should develop criteria for retention, promotion and security of employment of full-time faculty members in its professional skills program.

Id.

94. *Id.*

95. Memorandum C8483-16 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to the Council of the Section of Legal Educ. and Admissions to the Bar (Nov. 22, 1983) (on file with authors).

96. Memorandum D8384-27 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools, at 1 (Dec. 7, 1983) [hereinafter Memorandum D8384-27] (on file with authors); Memorandum from Roy Stuckey, Professor, Univ. of S.C. Sch. of Law, to Clinical Colleagues (Dec. 5, 1983) (on file with authors).

97. Memorandum D8384-27, *supra* note 96, at 7.

98. Letter from Paul D. Carrington, Dean, Duke Univ. Sch. of Law, et al., to Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (April 27, 1984) (on file with authors). The two other deans were Gerhad Casper, University of Chicago Law School, and Terrance Sandalow, University of Michigan Law School. Ironically, just a few years earlier, Dean Sandalow, as one of two people on an American Association of University Professors (AAUP) subcommittee investigating the increasing use of non-tenure-track teaching staff, argued that only with "very limited exceptions" should universities make academic appointments with anything other than tenure. Judith J. Thompson & Terrance Sandalow, *On Full-Time Non-*

accrediting process should be "lean" and should not intrude on the "autonomy and sense of professional responsibility of the institution being regulated."[99] They also argued:

> The proposed standard does nothing to encourage those law schools without clinical programs to develop them; it affects only law schools with a commitment to clinical legal education . . . . More important, we think it likely that the proposed standard would deter schools from starting new clinical programs or expanding ones they already have.[100]

They concluded that they thought "it unlikely that this standard can improve clinical legal education or legal education generally, and we see a substantial danger that it will make it worse."[101]

The leadership of the AALS also continued to oppose the proposed standard. In May 1984, the AALS Executive Committee, which included Dean Paul Carrington who had already registered his personal opposition to the proposed Standard 405(e), issued a statement arguing that law schools should have the freedom to establish a variety of employment approaches for clinical faculty and echoing the argument raised by Carrington and other deans that the proposed Standard "may well impede instead of support the development of clinical legal education."[102] After considering all of the comments, the Standards Review Committee rejected the position of the AALS leadership and recommended that the Council adopt the "shall" language for full-time faculty members whose primary responsibilities are in professional skills programs.[103]

The debate over the proposed Standard nevertheless continued. At the May 1984 Council meeting, Professor Joseph Julin, former Dean at the University of Florida School of Law and the President of the AALS, "gave a lengthy and passionate speech in opposition to adoption of the standard."[104] He argued that a study was needed to see if the Standard was necessary and that opponents

---

Tenure-Track Appointments, AAUP BULLETIN, Sept. 1978, at 267, 273. Sandalow argued that "administrators and faculty members who support institutional arrangements of the kind we have been surveying [namely, full-time non-tenure-track appointments] should recognize clearly that they are supporting practices which are inequitable, harmful to morale, and a threat to academic freedom." *Id.*

99. Letter from Paul D. Carrington to Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, *supra* note 98, at 1.

100. *Id.*

101. *Id.*

102. Statement of the Executive Comm. of the Ass'n of Am. Law Schs. on Proposed Standard 405(e), Am. Bar Ass'n Standards for Approval of Law Schools 3–4 (May 17, 1984) (on file with authors).

103. *See* Memorandum D8384-51, *supra* note 66, at 5. The Consultant's May 1984 memo traces the genesis of the 1984 adoption of Standard 405(e) back to the January 1980 Report of the AALS/ABA Joint Committee on Clinical Legal Education Guidelines. *Id.* at 3–6.

104. Memorandum from Dean Hill Rivkin, Professor, Univ. of Tenn., to Colleagues 2 (May 23, 1984) [hereinafter Rivkin Memorandum] (on file with authors).

would fight the proposal at the ABA House of Delegates.[105] Arguing in favor of the proposed Standard, Robert McKay, former Dean of New York University School of Law and Chair of the Section of Legal Education and Admissions to the Bar, stated that "equity, fairness, and educational necessity underpin this issue."[106] Norman Redlich, Dean of New York University School of Law and a member of the Council, stated the issue of status for clinical faculty was the "most important that he has faced in the accreditation of law schools," and Judge Henry Ramsey, another Council member, argued "that it was grossly unfair to discriminate against law teachers on the basis of what they teach."[107]

After three years of review and public hearings, and in light of several independent ABA reports recommending the necessity of improving the status of clinical faculty to advance clinical legal education, the Council adopted Standard 405(e) with the "shall" language on "security of position reasonably similar to tenure" by a unanimous vote.[108] The report accompanying the proposed Standard and Interpretations for consideration by the ABA House of Delegates stated the following as the reason for requested action: "The employment status of clinicians and other professional skills teachers has been debated for years and thoroughly studied by the Legal Education Section since 1981. Two rounds of hearings (for a total of four hearings) have been held. The matter is ripe for decision."[109]

Although the Council unanimously recommended the adoption of Standard 405(e) before sending it to the ABA House of Delegates, some within the

---

105. Statement of the Executive Comm. of the Ass'n of Am. Law Schs., *supra* note 102, at 3. The ABA Consultant on Legal Education later explained some of the opposition to Standard 405(e): "Many of the opponents of the Standard argued that improvements were occurring and would continue at an appropriate rate, with or without the issue being addressed directly by an accreditation standard." Memorandum D9091-25 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Nov. 12, 1990) (on file with authors).

106. Rivkin & Stuckey, *supra* note 67, at 4.

107. *Id.*

108. Memorandum D8384-51, *supra* note 66, at 1; Rivkin & Stuckey, *supra* note 67, at 4. In recommending that the House of Delegates adopt Standard 405(e), the Council also adopted a resolution stating that "[f]ull compliance with this Standard shall be required with the commencement of the 1986–87 academic year. In the intervening two years, each approved law school shall develop a plan, in conformity with this Standard." Memorandum D8384-51, *supra* note 66, at 1–2.

109. General Information Form submitted by Robert B. McKay, Chairman, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (Summer 1984) (on file with authors). The report of the Section of Legal Education and Admissions to Bar to the House of Delegates noted: "Equal treatment of clinical teachers and other skills teachers has been advocated by numerous ABA committees and task forces . . . ." Robert B. McKay, *Report of the Section of Legal Education and Admissions to the Bar*, 109 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 894, 895 (1984) (referencing the 1979 Cramton Report, 1980 Foulis Report, and 1983 Friday Report).

leadership of the AALS still opposed the Standard. In an effort to counteract the opposition, the law school deans on the Council—Richard Huber of Boston College Law School, Norman Redlich of New York University School of Law, and Gordon Schaber of McGeorge School of Law—and three other members sent a five-page letter to the deans of all law schools explaining why they favored Standard 405(e) with "shall" language.[110] Their letter noted that in the prior two years there had been a series of public hearings, yet no law school dean, including those now urging defeat of Standard 405(e), had appeared in opposition to the proposed Standard.[111]

The letter also stated that tenure or its equivalent was necessary to ensure both the quality of legal education and the academic freedom of clinical faculty:

> Few have ever questioned the relationship of tenure status to quality of legal education when applied to traditional academic faculty. Tenure, or some equivalent status, provides the assurance of academic freedom, which has long been regarded as essential for a quality faculty. This is no less true for teachers in a professional skills training program. The assurance of academic freedom affects quality in at least two ways: (a) it permits teachers to perform their academic responsibilities, in the classroom and in scholarship, without fear of reprisal; and (b) it helps to recruit high-quality faculty since potential teachers of distinction are more likely to be attracted to academic life if they can be assured of permanent status on a law school faculty.[112]

With regard to the argument that Standard 405(e) was "an example of over-regulation," the Council members noted that "it has never been suggested that a requirement of a tenure system for full-time faculty was an instance of over-regulation."[113] They also stated that tenure was not required, but security of position and "benefits and obligations . . . reasonably equivalent to those of other faculty members" were required.[114]

---

110. Letter from Richard Huber, Dean, Boston Coll. Law Sch., et. al., to Deans of ABA-Approved Law Schools (June 18, 1984) (on file with authors).

111. *Id.* at 1.

112. *Id.* at 2.

113. *Id.*

114. *Id.* at 3. They explained:

> We believe that Standard 405(e) is important not only as an assurance of high-quality professional skills teaching, but also as a matter of elemental fairness and decency. There should be no second class citizens among the full-time members of an academic faculty. Persons who are dedicating their professional careers to teaching our students the essential ingredients of lawyering skills should not be forced to tolerate a status which the rest of us would find wholly unacceptable. They should be carefully evaluated by whatever standards the faculty establishes, but once they meet those standards, they have as much right to full membership in the academic community as does anyone else. In academic life, such full membership means tenure, or the substantial equivalent thereof.

Id. at 4–5.

In June 1984, the AALS Executive Committee held a special meeting to reaffirm its opposition to the proposed Standard and promised that there would be a contested vote in the ABA House of Delegates.[115] The AALS opposition prompted second thoughts by the ABA. The Council conducted a mail ballot of its members in July and voted to change the language in Standard 405(e) from "shall" to "should."[116] In explaining the reason for the retreat, McKay wrote that "all of us would have preferred the 'shall' language; but there was at least some risk that we would lose the entire proposal, since many law schools (including some of the most influential, although often inactive in the section) would vigorously oppose a mandatory standard."[117]

In addressing the ABA House of Delegates prior to its vote on Standard 405(e) in August 1984, McKay explained that he originally supported the "shall" language but that he was

> now fully persuaded that we should not move that fast because a number of American law schools want still to be persuaded that the time has now come. . . . But we believe that the good sense of it would persuade schools[,] even though they are told only should and not shall[,] to adopt the kind of tenure qualifications that we believe are important. It is [a] question of fairness, equity and equality that there should be such a recognition.[118]

Speaking on behalf of the Law Student Division of the ABA, a delegate urged reinserting the "shall" language in order to attract and retain better clinical faculty and to promote clinical legal education in law schools.[119] Another

---

115. Letter from Roy Stuckey, Professor, Univ. of S.C. Sch. of Law, et al., to Colleagues 1–2 (June 29, 1984) (on file with authors); *see also* Memorandum from Joseph R. Julin, President, Ass'n of Am. Law Schs., to Deans of Member Schools and Members of the Am. Bar Ass'n House of Representatives 1 (June 29, 1984) (arguing that the ABA standards should not be amended such that a mandatory relationship would exist between a law school and its clinical faculty).

116. Memorandum from Robert B. McKay, Chairman, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Am. Bar Ass'n Board of Governors (July 26, 1984) (on file with authors).

117. Letter from Robert B. McKay, Chairman, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Roy T. Stuckey, Professor, Univ. of S.C. Sch. of Law (July 30, 1984) (on file with authors).

118. Transcript of Am. Bar Ass'n House of Delegates Debate on Standard 405(e), at 4–5 (Aug. 7, 1984) (unedited transcript of tape 4, attached to Memorandum from Fred Franklin, Staff Director, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Marilyn V. Yarbrough, Council Member, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, and Roy T. Stuckey, Professor, Univ. of S.C. Sch. of Law (Nov. 26, 1984)) [hereinafter Transcript of Am. Bar Ass'n House of Delegates 1984 Debate] (on file with authors).

119. Transcript of Am. Bar Ass'n House of Delegates 1984 Debate, *supra* note 118, at 6–7. The delegate stated:

> It is the position of the Law Student Division that there is nothing to be gained from a legal education which affords inferior status to clinical law teachers. Clinical training is an essential component to legal education. . . . There is no room for second class faculty in

delegate, identifying herself as not an educator but a "small firm practitioner with a heavy courtroom practice," stated that clinical law professors should not be "relegated to second class status as teachers when they provide such a valuable service to the actual practicing bar across the country."[120]

Responding to these and other comments, the President of the AALS urged the adoption of the "should" proposal by pledging that the AALS would "encourage and assist our member schools to develop and adopt appointment and governance policies which ensure and enhance the quality of the profession[al] skills education. Our responsibility to the public permits us to do no less."[121]

The House of Delegates adopted the revised version of Standard 405(e) at its annual meeting in August 1984 with the "should" language.[122] In addition to adopting Standard 405(e), the House of Delegates adopted the three Interpretations proposed by the Council in May.[123] The 1984 version of

---

our law schools. Law schools need to attract better teachers to clinical education programs. By affording similar status and some sort of job security to these individuals not only will law schools attract these kinds of teachers but they will also be able to keep them. . . . An aspirational goal utilizing the language of *should* does recognize the problem but does not solve it. A standard of accreditation utilizing the language of *shall* not only recognizes the problem but affords a remedy.

Id. (emphasis added).

120. *Id.* at 9–10. The delegate stated:

I don't understand this discrimination against clinical law professors. I fail to grasp why they are relegated to a second class status as teachers when they provide such a valuable service to the actual practicing bar across the country. . . . But what I ask you as a practical matter is to recognize without a doubt and with no uncertainty that clinical law professors ought to be given the status they deserve and to support the shall standard.

Id.

121. *Id.* at 12.

122. Memorandum D8485-6 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Aug. 10, 1984) [hereinafter Memorandum D8485-6] (on file with authors). Standard 405(e) provided:

The law school should afford to full-time faculty members whose primary responsibilities are in its professional skills program a form of security of position reasonably similar to tenure and perquisites reasonably similar to those provided other full-time faculty members by Standards 401, 402(b), 403 and 405. The law school should require these faculty members to meet standards and obligations reasonably similar to those required of full-time faculty members by Standards 401, 402(b), 403 and 405.

Id.; *see* SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Standard 405(e) (1986) [hereinafter 1986 STANDARDS]; *see also* Grant H. Morris & John H. Minan, *Confronting the Question of Clinical Faculty Status*, 21 SAN DIEGO L. REV. 793, 793–94 (1984) (discussing the implementation of Standard 405 (e)).

123. The Interpretations stated:

A. Interpretation

A form of security of position reasonably similar to tenure includes a separate tenure track or a renewable long-term contract. Under a separate tenure track, a full-time faculty member, after a probationary period reasonably similar to that for other full-time faculty,

Standard 405(e) remained the applicable accreditation provision on the status of clinical faculty until continued lack of improvement in the status of clinical faculty led the ABA to revisit the "should" versus "shall" issue in 1996.

IV. EVENTS LEADING TO THE ADOPTION OF STANDARD 405(C) IN 1996

After the adoption of Standard 405(e) in 1984, ABA committees and reports continued to express concern about the treatment of clinical law faculty. Evidence indicated that law schools were slow to adopt the appointment and governance policies for clinical faculty that the AALS had pledged to support in order to "ensure and enhance the quality of the professional skills education," which is what those arguing in favor of the "should" language claimed would occur.[124] Some law schools also were terminating clinical faculty with little or no notice, and many law schools did not permit clinical faculty to participate meaningfully in faculty governance.[125]

---

may be granted tenure as a faculty member in a professional skills program. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the professional skills program.

A program of renewable long-term contracts should provide that, after a probationary period reasonably similar to that for other full-time faculty, the services of the faculty member in a professional skills program may be either terminated or continued by the granting of a long-term contract that shall thereafter be renewable. During the initial long term contract or any renewal period, the contract may be terminated only for good cause, including termination or material modification of the professional skills program.

B. Interpretation

In determining if the members of the full-time faculty of a professional skills program meet standards and obligations reasonably similar to those provided for other full-time faculty, competence in the areas of teaching and scholarly research and writing should be judged in terms of the responsibilities of faculty members in the professional skills program. Each school should develop criteria for retention, promotion and security of employment of full-time faculty members in its professional skills program.

C. Interpretation

Standard 405(e) does not preclude a limited number of fixed, short-term appointments in a professional skills program predominantly staffed by full-time faculty members within the meaning of this Standard, or in an experimental program of limited duration.

See 1986 STANDARDS, *supra* note 122, at Standard 405(e) and Interpretations; Memorandum D8485-6, *supra* note 122, at 2–3; *see also* Stephen F. Befort, *Musings on a Clinic Report: A Selective Agenda for Clinical Legal Education in the 1990s*, 75 MINN. L. REV. 619, 629 (1991) (discussing how Standard 405(e) was adopted by the ABA to help remedy the "second-class treatment" of clinicians); Marjorie Anne McDiarmid, *What's Going on Down There in the Basement: In-House Clinics Expand Their Beachhead*, 35 N.Y.L. SCH. L. REV. 239, 274–75 (1990) (stating that the ABA House of Delegates added Standard 405(e) in 1984).

124. Transcript of Am. Bar Ass'n House of Delegates 1984 Debate, *supra* note 118, at 12.

125. *See infra* notes 127–33 and accompanying text.

A. Studies on the Effects of the "Should" Language

The first report to discuss the impact of the "should" language was a report of the ABA Skills Training Committee in 1986, and its conclusions were mixed. At its April 1986 meeting, the Skills Training Committee expressed its "sense that the 1984 amendments are having the intended effect of improving the overall quality of professional skills training programs in law schools."[126] However, the Skills Training Committee expressed concern about the manner in which law schools were terminating clinical professors and recommended that the Council adopt a statement concerning early notification of professional skills faculty about non-retention decisions.[127] Responding to this concern, the Council issued a special memorandum in August 1986 calling on law schools to provide sufficient notice of non-retention of professional skills faculty to allow them the opportunity to seek other positions.[128]

After the adoption of the "should" language, the Council realized that many law schools were still denying professional skills faculty opportunities to participate in law school governance. At its June 1988 meeting, the Council approved a Standards Review Committee recommendation to circulate for comment a proposed Interpretation suggesting that law schools should permit faculty teaching in the professional skills programs to participate in law school governance.[129] In December 1988, after receiving comments, the Council adopted an Interpretation of Standards 205, 403 and 405(e).[130] This new Interpretation on governance rights for clinical faculty provided:

---

126. Memorandum from Roy Stuckey, Chair, Skills Training Comm., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Members of the Council, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar 1 (Apr. 24, 1986) (on file with authors).

127. *Id.* at 1–2.

128. Memorandum from Kathleen S. Grove, Office of the Consultant on Legal Educ., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Members of the Skills Training Comm., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (Aug. 14, 1986) (on file with authors); Memorandum D8586-6 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Aug. 15, 1986) (on file with authors). The statement read as follows:

> The Council is informed that, during the process generated by the August, 1984 amendment of Standard 405(e) of the ABA Standards for Approval of Law Schools, certain law schools may have replaced or otherwise terminated the employment of professional skills teachers who were hired prior to the adoption of amended Standard 405(e) with little notice.
>
> The Council encourages any school that decides not to continue in service a professional skills teacher hired prior to the adoption of amended Standard 405(e) to provide sufficient notice to the teacher to allow a fair opportunity to seek another position.

OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 1986–1987 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 38 (1987).

129. Memorandum D8788-70 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (June 10, 1988) (on file with authors).

130. OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 1988–1989 ANNUAL

> A law school should afford to full-time faculty members whose primary responsibilities are in its professional skills program an opportunity to participate in law school governance in a manner reasonably similar to other full-time faculty members. This Interpretation does not apply to those persons referred to in Interpretation 3 of Standard 405(e) [that is, those with fixed, short-term appointments or in an experimental program of limited duration].[131]

The action of the Council was later explained as necessary "to make it clear that the 'perquisites' and 'obligations' language in S405(c) [then as S405(e)] includes participation in governance by full-time professional skills teachers."[132] Although the Council thought that participation in faculty governance was apparent from the language of the Standard, the Council explained that it adopted the new Interpretation because some law schools did not believe that the Standard covered governance.[133]

In July 1992, yet another ABA report called on law schools to provide appropriate status to clinical faculty. The influential report "Legal Education and Professional Development—An Educational Continuum," known as the MacCrate Report, observed that while status for clinical faculty was improving and the number of full-time professional skills faculty was increasing, the "progress has not been uniform, and at some institutions, it has come slowly and without the commitment that is necessary to develop and maintain skills

---

REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 37 (1989); Memorandum D8889-33 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools, at 1 (Dec. 15, 1988) [hereinafter Memorandum D8889-33] (on file with authors).

131. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Interpretations to Standard 405 (1990); Memorandum D8889-33, *supra* note 130, at 1.

132. Memorandum from Roy Stuckey, Professor, Univ. of S.C. Sch. of Law, to Members of the Council, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar 1 (May 17, 1996) (on file with authors). Professor Stuckey was a member of the Council from 1988–1994, a member of the Skills Training Committee from 1984–1996, and a member of the Standards Review Committee from 1991–1995.

133. The Council's action was explained as follows:

> In December of 1988, the Council adopted this Interpretation to make it clear that the "perquisites" and "obligations" language in S405(c) [then as S405(e)] includes participation in governance by full-time professional skills teachers. There was no uncertainty among members of the Council about this. The only question was whether an Interpretation was needed or whether it was sufficiently apparent from the language of the Standard. After hearing evidence that not every school understood that S405(c) includes governance, Rosalie Wahl [Justice, Minnesota Supreme Court, and Chair of the Council 1987–88] brought the discussion to an end by commenting that "if that is what we mean, we should not hesitate to be clear about it." I do not believe that there was a dissenting vote.

Id.

instruction of a quality commensurate with the school's overall educational aspirations."[134]

Data supporting the MacCrate Report's concern was assembled in a 1991 study for the ABA Office of the Consultant on Legal Education.[135] It reported the results of seven years of questionnaires to law schools on "the status of professional skills teachers."[136] The purpose of the study was "to assess both the quality and quantity of changes being made" on the status of clinical faculty.[137]

The study found that the percentage of full-time professional skills faculty holding "tenure eligible slots" actually dropped by over five percent during the seven-year period from 1984 to 1991.[138] An interim report theorized that the drop could be because "as the number of professional skills teachers has expanded, the number of slots eligible for job security under Standard 405(e) has remained relatively static. A disproportionate number of new teachers is being put in temporary slots with little or no job security."[139] Additionally, the percentage of professional skills teachers holding positions that did not meet the requirements of Standard 405(e) declined only slightly from 1985 to 1990, leading the study's author to explain that "S405(e) may have had no impact on the job security of the people it was primarily intended to assist."[140] The ABA study concluded: "In sum, the data produced by this project does not demonstrate that ABA Accreditation Standard 405(e) has improved the status of full-time teachers of professional skills, nor does the data indicate trends which would suggest a probability of significant future progress."[141] At a minimum, the seven-year study refuted the claim of opponents of an ABA

---

134. SECTION OF LEGAL EDUC. & ADMISSIONS TO THE BAR, AM. BAR ASS'N, REPORT OF THE TASK FORCE ON LAW SCHOOLS AND THE PROFESSION: NARROWING THE GAP (1992).

135. ROY STUCKEY, OFFICE OF THE CONSULTANT ON LEGAL EDUC. FOR THE AM. BAR ASS'N, FINAL REPORT: RESULTS OF SURVEYS AND QUESTIONNAIRES REGARDING THE STATUS OF PROFESSIONAL SKILLS TEACHERS 1984–1991 (1991) [hereinafter STATUS OF SKILLS TEACHERS FINAL REPORT].

136. *Id.* at 1.

137. *Id.* The memorandum accompanying each questionnaire explained:

> Throughout the process leading up to the adoption of ABA Accreditation Standard 405(e) in August, 1984 many legal educators and bar leaders felt that the status of full-time teachers of professional skills should be improved. Many opponents of the Standard argued that improvements were occurring and would continue at an appropriate rate, with or without the issue being addressed directly by an accreditation standard.
>
> The Council wishes to obtain an overview of this process so it will be able to assess both the quality and quantity of changes being made.

Id.

138. *Id.* at 3.

139. ROY STUCKEY, OFFICE OF THE CONSULTANT ON LEGAL EDUC. FOR THE AM. BAR ASS'N, PRELIMINARY REPORT NO. 3: RESULTS OF SURVEYS AND QUESTIONNAIRES REGARDING THE STATUS OF PROFESSIONAL SKILLS TEACHERS, at ii (1989).

140. STATUS OF SKILLS TEACHERS FINAL REPORT, *supra* note 135, at 4.

141. *Id.* at 5.

standard for the job security of clinical faculty that improvements in status would occur without a mandatory accreditation standard.

Around the same time as the release of the MacCrate Report on the status of professional skills teachers, the ABA's Skills Training Committee raised concerns over the enforcement of 405(e).[142] The Skills Training Committee's concern echoed a 1987 AALS Section on Clinical Legal Education survey of law school clinical programs that found that a significant percentage of schools were disregarding 405(e).[143]

A 1991 AALS Section on Clinical Legal Education study, "Report of the Committee on the Future of the In-House Clinic," also found that Standard 405(e) was not significantly affecting the status of clinical teachers.[144] The report's survey of law schools with in-house clinics found that at a majority of law schools Standard 405(e) was having no perceptible effect, with forty percent of those schools indicating either that the Standard was "not a factor at their school or that their faculty was disregarding" it.[145] Those reporting some effects from Standard 405(e) indicated that the Standard was more likely to help than to harm their security and they felt that the future effects of the Standard largely would be helpful.[146] This series of reports and studies pointed to the failure of the "should" language and the need for the ABA to address the status of clinical faculty once more.

B. ABA Changes "Should" to "Shall" in Standard 405(c)

In June 1994, the ABA's Council, upon the recommendation of the Standards Review Committee, declared its intention to amend Standard 405 again.[147] Significantly, Millard Rudd suggested that it was time to change the Standard.[148] Rudd, as Executive Director of the AALS, had led that organization's fight in 1984 to make the security of position requirement in Standard 405 "should" rather than "shall."[149] His argument at the time was that

---

142. OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 1992–1993 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 70 (1993).

143. McDiarmid, *supra* note 123, at 276–77 (reporting the results of a questionnaire sent to law schools by the AALS Section on Clinical Legal Education in 1987).

144. *Report of the Committee on the Future of the In-House Clinic*, 42 J. LEGAL EDUC. 508, 556 (1992).

145. *Id.* at 542–43, 556.

146. *Id.* at 543.

147. OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 1994–1995 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 41 (1995) [hereinafter CONSULTANT'S 1994–1995 REPORT]; Joseph W. Bellacosa, *Report No. 2 of the Section of Legal Education and Admissions to the Bar*, 120 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 351, 352 (1995).

148. Roy Stuckey, *New ABA Accreditation Standards: An Insider's View*, CLEA NEWSLETTER (Clinical Legal Educ. Ass'n, New York, N.Y.), Sept. 1996, at 10, 13.

149. *Id.*

clinical legal education was relatively new to many law schools and that law schools needed more time to adjust to providing security of position for clinical faculty.[150] By 1994, Rudd, now a member of the Standards Review Committee, was convinced that law schools had sufficient opportunity to adjust and that it was time for security of position for clinical faculty to be mandatory.[151]

After two public hearings, the Council decided not to recommend any substantive changes of Standard 405(e) for approval by the ABA's House of Delegates at its February 1995 meeting.[152] Instead, the Council simply moved the status standard in 405(e) to 405(c); no change was made in the security of position language of old section (e) or its relevant Interpretations.[153]

Although the ABA did not address the status of clinical faculty in 1995, the issue remained unsettled. The ABA was in the midst of a major recodification of the Standards and Interpretations, and this set the stage for further debate over the status of clinical faculty. The Standards Review Committee held four public hearings on proposed changes and received hundreds of written comments dealing with "almost every conceivable position on every subject covered by the Standards . . . advocated pro or con," including the argument for a more laissez faire, deregulatory approach to clinical faculty standards.[154] In contrast to those seeking to roll back standards for clinical faculty, the Clinical Legal Education Association (CLEA) argued that it was time to strengthen the standards. CLEA contended that in 1984, the year when "shall" was changed to "should" after the Council had unanimously recommended the "shall" language, "the change was premised on the idea that the more permissive

150. *Id.*

151. *Id.*

152. CONSULTANT'S 1994–1995 REPORT, *supra* note 147, at 41–42; Bellacosa, *supra* note 147, at 351–53; *General Minutes*, 120 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 24 (showing the approval of the Council's recommendation); *Actions of the House of Delegates*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Spring 1995, at 15.

153. CONSULTANT'S 1994–1995 REPORT, *supra* note 147, at 42; SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS 36–37, 43–44 (1995); Bellacosa, *supra* note 147, at 352. The February 1995 changes to Section 405 also labeled the existing introductory sentence new subsection (a), deleted the previous language in subsection (a) on compensation, moved the standards in (b) and (c) on research, travel, and secretarial support to Interpretation 3, and moved the requirement to have a policy on academic freedom and tenure from subsection (e) to (b). *See* CONSULTANT'S 1994–1995 REPORT, *supra* note 147, at 42–43; *see also* Memorandum D9495-34 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Dec. 15, 1994) (on file with authors) (providing the proposed amendments); Memorandum D9495-40 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (Feb. 21, 1995) (on file with authors) (providing the amendments as adopted).

154. JAMES P. WHITE, CONSULTANT ON LEGAL EDUC. TO THE AM. BAR ASS'N, 1995–1996 ANNUAL REPORT 25 (1996) [hereinafter CONSULTANT'S 1995–1996 REPORT].

language would be temporary. Sufficient time has passed to demonstrate the need for the adoption of mandatory language."[155]

The Council rejected the call to deregulate the status of clinical teachers and instead decided that the "should" language was not having its desired effect at all law schools. At its meeting in June 1996, the Council voted to amend Standard 405(c) by replacing the words "professional skills" with "clinical" and changing the word "should" to "shall," and the ABA House of Delegates adopted these changes at its Annual Meeting in August 1996.[156] The ABA explained "that full-time clinical faculty members must be afforded a form of security of position reasonably similar to tenure, and noncompensatory perquisites reasonably similar to other full-time faculty members."[157] The new Standard also explained that it did "not preclude a limited number of fixed, short-term appointments in a clinical program predominantly staffed by full-time faculty members, or in an experimental program of limited duration."[158] The ABA made similar changes to the Interpretations to 405(c) and added a sentence to Interpretation 405-7 to address further the security of position of clinical teachers: "A law school should develop criteria for retention, promotion, and security of employment of full-time clinical faculty."[159]

---

155. Letter from Mark J. Heyrman, Secretary/Treasurer, Clinical Legal Educ. Ass'n, to James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, enclosure at 3 (Mar. 20, 1996) (on file with authors).

156. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS 43 (1996) [hereinafter 1996 STANDARDS]; Erica Moeser, *Report No. 1 of the Section of Legal Education and Admissions to the Bar*, 121 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 375–76 (1996); *see also id.* at 28 (showing the approval of the Council's recommendation).

157. *Recodification of Standards Nears Completion*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Winter 1996, at 1, 14 [hereinafter *Recodification of Standards*].

158. 1996 STANDARDS, *supra* note 156, at 43; Moeser, *supra* note 156, at 375–76.

159. 1996 STANDARDS, *supra* note 156, at 44; Moeser, *supra* note 156, at 377. The changes to the Interpretations included:

Interpretation 405-6:

> A form of security of position reasonably similar to tenure includes a separate tenure track or a renewable long-term contract. Under a separate tenure track, a full-time *clinical* faculty member, after a probationary period reasonably similar to that for other full-time faculty, may be granted tenure ~~as a faculty member in a professional skills program~~. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the ~~professional skills~~ *clinical* program.
>
> A program of renewable long-term contracts should provide that, after a probationary period reasonably similar to that for other full-time faculty, the services of a faculty member in a ~~professional skills~~ *clinical* program may be either terminated or continued by the granting of a long-term contract that shall thereafter be renewable. During the initial long-term contract or any renewal period, the contract may be terminated for good cause, including termination or material modification of the professional skills program.

The ABA also amended Standard 405(c) by adding the phrase "non-compensatory" before the language requiring that clinical faculty members be afforded "perquisites reasonably similar to those provided other full-time faculty members."[160] This change reflected the agreement with the U.S. Department of Justice in a 1996 anti-trust consent decree not to adopt or enforce any Standard or Interpretation that had the purpose or effect of imposing requirements as to base salary, stipends, fringe benefits, or other compensation paid to law school employees.[161] These 1996 changes remained the applicable ABA Standard and Interpretations on the status of clinical faculty until the ABA amended the Interpretations in the summer of 2005.

V. THE 2005 CHANGES TO STANDARD 405(C) INTERPRETATIONS

Resistance to the "shall" language in Standard 405 requiring reasonably similar treatment of clinical faculty with other faculty continued after the 1996 amendments, though no longer by the AALS. In the years following the amendments, several efforts were made, each unsuccessful, to persuade the ABA to abandon its support for greater integration of clinical faculty and courses in law schools.

A. The Association of Law Deans of America's Resistance to Status for Clinical Faculty

In 1999, the Standards Review Committee considered a dramatic restructuring of Standard 405 to eliminate all references to tenure, for both

---

Interpretation 405-7:

In determining if the members of the full-time *clinical* faculty ~~of a professional skills program~~ meet standards and obligations reasonably similar to those provided for other full-time faculty, competence in the areas of teaching and scholarly research and writing should be judged in terms of the responsibilities of *clinical* faculty ~~members in the professional skills program~~. *A law school should develop criteria for retention, promotion, and security of employment of full-time clinical faculty.*

Interpretation 405-8:

A law school ~~should~~ *shall* afford to full-time *clinical* faculty members ~~whose primary responsibilities are in its professional skills program~~ an opportunity to participate in law school governance in a manner reasonably similar to other full-time faculty members. This Interpretation does not apply to those persons referred to in *the last sentence of Standard 405(c).*

Id. at 376–77 (italics, underscores, and strike-outs in original).

160. 1996 STANDARDS, *supra* note 156, at 43.

161. *See* Moeser, *supra* note 156, at 349–50; *Informational Report from the ABA Board of Governors to the House of Delegates*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Spring 1996, at 4; *see also* CONSULTANT'S 1995–1996 REPORT, *supra* note 154, at app. g (containing the final judgment and consent decree from *United States v. American Bar Association*, 934 F. Supp. 435 (D.D.C. 1996)).

clinical and nonclinical faculty.[162] The Association of Law Deans of America (ALDA) urged the changes as early as 1996, arguing that Standard 405(c) should be deleted because requiring a form of security of position for clinical faculty reasonably similar to tenure was inconsistent with the lack of any ABA "requirement that a law school have a tenure system at all."[163] ALDA also proposed the elimination of Standard 302(d), which requires law schools to offer some form of live-client or other real-life practice experience.[164] ALDA's opposition to live-client or other real-life practice experience represented resistance to any accreditation requirement that law schools offer their students clinical legal education and was consistent with ALDA's opposition to security of position and participation in law school governance by clinical faculty.

After holding public hearings and receiving comments, the Standards Review Committee proposed removing all mention of tenure from Standard 405.[165] Instead, it recommended that law schools be required to adopt such policies for security of position and academic freedom as are necessary to attract and retain a competent faculty and noted that these policies "may vary with the duties and responsibilities of different faculty members."[166] Standards

---

162. *Validation of Standards Chapters 3 and 4—Preliminary Proposals and Request for Comments*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Winter 1999, at 1, 17–18 [hereinafter *Validation of Standards*].

163. *Final Commentary on Changes in Chapters Three and Four of the Standards for Approval of Law Schools, 1998–1999*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Summer 1999, at 8, 10, 15 [hereinafter *Final Commentary*]; *see* Am. Law Deans Ass'n, Statement of the American Law Deans Association on Proposed Modification of the Standards for the Approval of Law Schools of the American Bar Association 9–10 (attachment to Letter from Ronald A. Cass, President, Am. Law Deans Ass'n, to James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n (Apr. 21, 1997)) (on file with authors). ALDA did not object to the retention of Standard 405(b), which requires law schools to have an established policy with respect to tenure, but only objected to any requirement of security of position reasonably similar to tenure for clinical faculty. *See Final Commentary*, *supra*, at 15.

164. *Final Commentary*, *supra* note 163, at 8–9.

165. *See Validation of Standards*, *supra* note 162, at 16.

166. Memorandum D9899-78 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools (July 21, 1999) [hereinafter Memorandum D9899-78] (on file with authors). The entire text of the proposed revisions to Standard 405 provided:

> (a) A law school shall establish and maintain conditions adequate to attract and retain a competent faculty.
>
> (b) A law school shall have established and announced policies designed to afford full-time faculty members, including clinical and legal writing faculty, whatever security of position and other rights and privileges of faculty membership as may be necessary to (i) attract and retain a competent faculty, (ii) provide students with a program of legal education that satisfies the requirements of Chapter 3 of these Standards, and (iii) safeguard academic freedom. The forms and terms of security of position and other rights and privileges of faculty membership may vary with the duties and responsibilities of different faculty members.

Review proposed the restructuring to move away from the concept of tenure and instead focus "on the programmatic objectives that 'security of position and other rights and privileges of faculty membership' are designed to achieve . . . ."[167] The Committee described such objectives as "ensuring that there is a faculty competent to fulfill the educational missions" of the Standards and "preserving academic freedom."[168] The Committee, however, did not endorse the ALDA proposal to eliminate the live-client or other real-life practice experience requirement in Standard 302(d).[169]

At one of the public hearings in May 1999, Carl Monk, Executive Director of the AALS, testified that the AALS Executive Committee voted to oppose all proposed changes to Standard 405.[170] The AALS Executive Committee opposed removing the tenure policy requirement in Standard 405 because "such a change to such a major core traditional value of the academy should not be made without very broad consultation that goes beyond these series of hearings with all types of law faculty and others in the higher education community."[171] In support of the AALS position, Monk recounted an example of a dean discussing a major dispute on her campus and her belief that "faculty were much more willing to speak up without fear who in fact had tenure."[172] David Short, Dean of Northern Kentucky University College of Law, spoke in support of Monk's comments and noted that the elimination of tenure would weaken law schools within their universities.[173]

The Council considered and adopted the Standard Review Committee's recommendation to preserve live-client and other real-life practice experiences in Standard 302(d), but it did not send the Committee's recommendation on restructuring Standard 405 out for public comment.[174] The 1999-2000 Annual

---

Id.

167. *Final Commentary*, *supra* note 163, at 15.

168. *Id.* Even though the proposed revisions eliminated references to tenure, proposed Interpretation 405-2 still stated that "[a]ttraction and retention of competent clinical faculty members presumptively requires a form of security of position, appropriate opportunities to participate in law school governance, and other rights and privileges of faculty membership that are reasonably similar to that provided to full-time non-clinical faculty members." Memorandum D9899-78, *supra* note 166, attachment at 20; *Validation of Standards*, *supra* note 162, at 18.

169. Memorandum D9899-78, *supra* note 166, attachment at 2.

170. Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, Standards Review Comm. Hearing on Recodification of Standards, in S.F., Cal., at 7–8 (May 19, 1999) [hereinafter 1999 Standards Review Committee Hearing] (transcript on file with authors) (statement of Mr. Monk).

171. *Id.* at 8.

172. *Id.*

173. *Id.* at 8–9 (statement of Mr. Short).

174. Office of the Consultant on Legal Educ., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, Commentary on the Proposed Changes to Chapters Five, Six and Seven of the Standards for the Approval of Law Schools 1999–2000 (attachment to Memorandum D9900-26 from James P. White, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of

Report of the Consultant on Legal Education explained that the Council rejected the call to eliminate language concerning job security "[b]ecause of its belief in the important role of tenure in protecting academic freedom."[175] In explaining its rejection of ALDA's call to repeal the requirement for live-client experiences, the Council noted the benefits of such real-life practice experiences and the fact that a law school need not offer the experience to all students.[176]

In 2001, the ABA House of Delegates did adopt changes to an Interpretation to Standard 405(c).[177] The changes to Interpretation 405-6 clarified that once a faculty member had clinical tenure or a renewable long-term contract, the clinical faculty member could only be terminated for good cause, which includes termination or material modification of the "entire" clinical program.[178]

ABA-Approved Law Schools (Dec. 22, 1999)) [hereinafter Commentary on Proposed Changes 1999–2000] (on file with authors).

175. OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 1999–2000 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 31 (2000) [hereinafter CONSULTANT'S 1999–2000 REPORT]; *see also Validation of Standards*, *supra* note 162, at 17–18 (indicating that in order to keep a professional environment, law schools must have a policy promoting academic freedom); Commentary on Proposed Changes 1999–2000, *supra* note 174, at 2 ("The council voted not to place the Standards Review Committee's revised recommendation on Standard 405 out for comment because of its belief that the standard's current tenure requirement is an important protection of academic freedom.").

176. *Validation of Standards*, *supra* note 162, at 10; Moeser, *supra* note 156, at 365.

177. *Report No. 2 of the Section of Legal Education and Admissions to the Bar*, 126 ANNUAL REPORT OF THE AMERICAN BAR ASSOCIATION 725–26 (2003); *see also id.* at 50 (showing the approval of the recommendation).

178. *Id.* at 725–26. The new Interpretation adopted by the House of Delegates stated:

> Interpretation 405-6: A form of security of position reasonably similar to tenure includes a separate tenure track or a renewable long-term contract. Under a separate tenure track[,] a full-time clinical faculty member, after a probationary period reasonably similar to that for other full-time faculty, may be granted tenure. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the *entire* clinical program.
>
> A program of renewable long-term contracts should provide that, after a probationary period reasonably similar to that for other full-time faculty, the services of a faculty member in a clinical program may be either terminated or continued by the granting of a long-term contract that shall thereafter be renewable. During the initial long-term contract or any renewal period, the contract may be terminated for good cause, including termination or material modification of the ~~professional skills program~~ *entire clinical program.*

Id. (italics, underscores, and strike-outs in original); *see* SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Interpretation 405-6 (2001). The ABA explained:

> This change clarifies Interpretation 405-6. The Council concluded that the legislative history made clear that Standard 405(c) intended to provide clinic-wide job security for a person who has security of employment under Standard 405(c). A law school may not

Early in 2003, the Council and the Accreditation Committee asked the Standards Review Committee "to consider the meaning of 'renewable' in Interpretation 405-6."[179] The request noted that:

> There is a question and no agreement about whether "renewable" means "presumptively renewable," so that a person holding such a contract could rely on long-term and continuing employment so long as the person's work performance was satisfactory, or "capable of being renewed," meaning that the contract is not subject to a term limit or cap on the length of time that the person could be in such a position. The history of Standard 405(c) suggests that this question was not resolved at the time the Standard was adopted.[180]

In September 2003, the Standards Review Committee again recommended, as it had done in 1999, that any reference to tenure be deleted from Standard 405(b) and instead that the definition of academic freedom be expanded, that the minimum protection that must be provided for academic freedom be set forth, and that the Standard should explain who is entitled to the protection of academic freedom.[181] As to clinical faculty status, the Committee went beyond its 1999 proposal by drafting a new Interpretation to "[r]equire that if a school has a system of tenure, full-time clinical faculty must be provided the type of 'similar treatment' that is now provided by 405(c) and Interpretations 405-6, -7 and -8."[182] If a school did not have a system of tenure, the proposed Interpretation provided that "clinical faculty shall be afforded reasonably

---

limit Standard 405(c) protection to the continuation of a particular clinical program. Standard 405(c) continues to provide that it does not preclude a law school from having a limited number of fixed, short-term appointments in a clinical program predominantly staffed by full-time faculty members or in an experimental program of limited duration.

Comment and review on this change was mixed. A number of law school deans objected on the grounds that the change provides less flexibility for law schools. Other deans and law school faculty, primarily those who teach in clinical programs, supported the change on the grounds that it was consistent with the intent of Standard 405(c) and was good policy.

OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 2000–2001 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 24–25 (2001); *see Proposed Revisions to Standards, Interpretations, Rules of Procedure and Bylaws Being Circulated for Comment*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Summer/Fall 2000, at 8, 10–11.

179. Office of the Consultant on Legal Educ., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, Commentary on the Changes to the Standards for the Approval of Law Schools and the Work of the Standards Review Committee 2002–2003, at 14 (Aug. 2003) [hereinafter Commentary on Changes 2002–2003] (on file with authors).

180. *Id.* At the time, the Accreditation Committee was using the latter interpretation of the term. *Id.*

181. Standards Review Comm., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, Revisions to Chapters 3 and 4: Tentative Decisions/Drafting Directions 11 (Sept. 19–20, 2003) (on file with authors).

182. *Id.*

similar treatment to that afforded other full-time faculty."[183] After meeting again in November 2003, the Standards Review Committee drafted and forwarded to the Council proposed changes consistent with the Committee's initial recommendations.[184] In February 2004, the Council approved proposed changes to Standards 401-404 for public comment but did not include any recommended change to Standard 405 or its Interpretations, thus rejecting the call to delete any reference to tenure from Standard 405 for a second time.[185]

The Standards Review Committee continued its other work on Standard 405 and in June 2004 asked the Council to delay sending proposed changes to Chapter 4 to the House of Delegates "until the Committee had the opportunity to consider recommending other revisions to Standard 405."[186] In November 2004, the Committee recommended additional changes to an Interpretation of Standard 405 to specify that "long-term contracts" must be at least five years in length and renewable to satisfy the "'reasonably similar to tenure'" requirement for employment relationships with clinical faculty.[187]

At its December 2004 meeting, the Council approved for notice and comment revisions to Interpretations to Standard 405.[188] With regard to security of position for clinical faculty, proposed Interpretation 405-6 stated that "'long-term contract' means at least a five-year renewable contract."[189]

---

183. *Id.*

184. Memorandum from Michael J. Davis, Chairperson, Standards Review Comm., Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar, to Council of the Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (Jan. 15, 2004) (on file with authors).

185. Memorandum from John A. Sebert, Consultant on Legal Educ. to the Am. Bar Ass'n, and Michael J. Davis, Chairperson, Standards Review Comm., to Deans of ABA-Approved Law Schools et al. 1, 3 (Feb. 20, 2004) (on file with authors).

186. *Commentary on Revisions to Standards for Approval of Law Schools 2004–05*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Fall 2005, at 59 [hereinafter *Commentary on Revisions to Standards 2004–2005*]; OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 2003–2004 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 54 (2004) [hereinafter CONSULTANT'S 2003–2004 REPORT].

187. OFFICE OF THE CONSULTANT ON LEGAL EDUC., AM. BAR ASS'N, 2004–2005 ANNUAL REPORT OF THE CONSULTANT ON LEGAL EDUCATION TO THE AMERICAN BAR ASSOCIATION 56 (2005) [hereinafter CONSULTANT'S 2004–2005 REPORT]; *Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 74.

188. *Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 59; CONSULTANT'S 2004–2005 REPORT, *supra* note 187, at 56.

189. *Proposed Revision of Chapter 4 of the Standards*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Feb. 2005, at 12 [hereinafter *Proposed Revision of Chapter 4*]. Proposed Interpretation 405-6 stated:

> A form of security of position reasonably similar to tenure includes a separate tenure track or *a program of renewable long-term contracts* ~~a renewable long-term contract~~. Under a separate tenure track, a full-time clinical faculty member, after a probationary period reasonably similar to that for other full-time faculty, may be granted tenure. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the entire clinical program.

Proposed Interpretation 405-8 defined participation in faculty governance to be "participation in faculty meetings, committees and other aspects of law school governance in a manner reasonably similar to other full-time faculty, including voting on non-personnel matters."[190]

A memorandum accompanying the proposed changes from the Consultant on Legal Education and Chair of the Standards Review Committee explained: "The proposed revision to Interpretation 405-6 clarifies the circumstances under which a program of long-term contracts will be considered to provide full-time clinical faculty a 'form of security of position reasonably similar to tenure' as required by Standard 405(c)."[191] The memorandum summarized the history of the debate and explained that the Council was acting because the Accreditation Committee had been approving schools with three-year contracts and no presumption of renewal and that such contracts were "inconsistent with the plain meaning of that Standard [405(c)]."[192]

> A program of renewable long-term contracts *shall* ~~should~~ provide that, after a probationary period reasonably similar to that for other full-time faculty, *during which the clinical faculty member may be employed on short-term contracts*, the services of a faculty member in a clinical program may be either terminated or continued by the granting of a long-term *renewable* contract ~~that shall thereafter be renewable~~. *For the purposes of this Interpretation, "long-term contract" means at least a five-year renewable contract.* During the initial long-term contract or any renewal period, the contract may be terminated for good cause, including termination or material modification of the entire clinical program.

Id. at 14 (italics, underscores, and strike-outs in original); *see also* Memorandum from John A. Sebert, Consultant on Legal Educ., and J. Martin Burke, Chair, Standards Review Comm., to Deans of ABA-Approved Law Schools et al. 8 (Dec. 10, 2004) (on file with authors) (showing the changes made to Interpretation 405-6).

190. *Proposed Revision of Chapter 4*, *supra* note 189, at 13. Proposed Interpretation 405-8 stated:

> A law school shall afford to full-time clinical faculty members *participation* [in] ~~an opportunity to participate in faculty meetings, committees, and other aspects of~~ law school governance in a manner reasonably similar to other full-time faculty members, *including voting on non-personnel matters*. This interpretation does not apply to those persons referred to in the last sentence of Standard 405(c).

Id. at 15 (italics, underscores, and strike-outs in original).

191. Memorandum from Sebert & Burke to Deans of ABA-Approved Law Schools, *supra* note 189, at 4; *see also Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 12 (including the same explanation for the proposed revisions by the Council).

192. Memorandum from Sebert & Burke to Deans of ABA-Approved Law Schools, *supra* note 189, at 4. The memorandum explained:

> There has been considerable debate regarding the role of the Standards in establishing conditions and terms of employment. Considering, however, that the Standards continue to establish conditions and terms of employment, it was the prevailing view that the practice developed by the Accreditation Committee—that a three-year renewable contract carrying no presumption regarding renewal is a "form of security of position reasonably similar to tenure" within the meaning of Standard 405(c)—is inconsistent with the plain meaning of that Standard. The proposed change . . . makes clear that a "program of

B. The 2005 Changes to the Clinical Status Interpretation

From January through May 2005, the ABA held a number of public hearings on the proposals to change the Interpretations to Standard 405(c).[193] ALDA opposed the changes, arguing:

> The specific terms of employment at most law schools are already sufficient to secure excellent clinical faculty. There is presently variety in what schools offer to clinical faculty [in terms of security of position and participation in law school governance] . . . . This variety has generally been healthy and there is no reason to stifle it with new restrictions on schools.[194]

Professor John Elson, a former member of the Accreditation Committee, submitted a letter criticizing the position of ALDA, stating:

> [ALDA's] basic justification that 405c is unnecessary to secure excellent clinical faculty ignores the historical circumstances that led to 405c's adoption. In adopting this Standard, the ABA realized that clinical teaching had come to play a critical role in the preparation of law students for practice and that clinical teachers could not become an effective presence in legal education unless a significant number of them were assured some security in their jobs and a significant role in law school governance. . . . [ALDA president Dean Saul Levmore's] counter-factual hypothesis that the free market would be sufficient to attract and retain clinical faculty of the quality and experience needed to provide excellent clinical supervision is not only without factual support and is contradicted by the pre-Standard 405c state of legal education, but it is also contradicted by the prevailing incentive structure in legal education, which rewards faculty excellence in scholarship

---

> renewable long-term contracts" will only be "reasonably similar to tenure" if, following a probationary period during which a full-time clinical faculty could be employed on short-term contracts, the employment of the faculty member is either terminated or continued by a granting of a renewable contract at least five years in length. The five-year term reflects the pattern for post-tenure review that is evolving at many schools. By providing greater security of position than the Accreditation Committee's practice, the proposed revision is designed to achieve the goal of Standard 405(c), i.e., to ensure that law schools can attract and retain quality full-time clinical faculty and thereby strengthen the clinical component of the law school curriculum. . . . A proposal that renewable long-term contract carries with it a presumption of renewal was considered and ultimately rejected.

Id.; *see also Proposed Revision of Chapter 4*, *supra* note 189, at 12 (including the same explanation for the proposed revisions by the Council).

193. *Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 59; CONSULTANT'S 2004–2005 REPORT, *supra* note 187, at 61 (reprinting the ABA Section of Legal Education and Admission to the Bar's "Commentary on Revisions to Standards for Approval of Law Schools 2004–05").

194. Letter from Saul Levmore, Dean, Univ. of Chi. Law Sch., to Stephen Yandle, Deputy Consultant, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (Apr. 28, 2005) (on file with authors).

far more than it does faculty excellence in preparing students for their role as practitioners.[195]

After considering all public comments, the Standards Review Committee recommended to the Council that it "adopt without change the proposed revisions to Interpretation 405-6."[196] The Committee explained that its proposed revisions did not enlarge the security of position for clinical faculty but instead "provide much-needed specific guidance to law schools and the Accreditation Committee regarding the proper interpretation of the language of Standard 405(c)."[197] The Committee noted that long-term contracts not only ensure that law schools can attract and retain quality clinical faculty but also "play a significant role in ensuring the academic freedom of full-time clinical faculty."[198]

At its June 2005 meeting, the Council reviewed the recommendations from the Standards Review Committee.[199] The Council also decided to revisit the issue of whether long-term contracts must be presumptively renewed (which had previously been considered and rejected by Standards Review) and added the following language to Interpretation 405-6: "For the purposes of this Interpretation, 'long-term contract' means at least a five-year contract that is presumptively renewable or other arrangement sufficient to ensure academic freedom."[200]

The House of Delegates concurred with the proposed changes at its Annual Meeting in August 2005.[201] The resulting Standard 405(c) and Interpretations,

---

195. Letter from John S. Elson, Professor, Northwestern Univ. Sch. of Law, to Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (May 3, 2005) (on file with authors) (providing his response to Dean Levmore's call, on behalf of ALDA, for abolition or modification of Standard 405(c)).

196. Memorandum from J. Martin Burke, Chairperson, Standards Review Comm., to Council of the Section of Legal Educ. and Admissions to the Bar 1, 1 (May 22, 2005).

197. *Id.* at 1–2.

198. *Id.* at 2. Standard 405(b) requires that a law school shall have an established and announced policy on academic freedom and tenure and references as an example the 1940 Statement of Principles on Academic Freedom and Tenure of the American Association of University Professors (AAUP). 2007 STANDARDS, *supra* note 1, at Standard 405(b) & app.1. The AAUP Statement explains that tenure is a means to promote freedom in teaching and research and to provide sufficient economic security to make the profession attractive. *Id.* at app.1. It further states that "[f]reedom and economic security, hence, tenure, are indispensable to the success of an institution in fulfilling its obligations to its students and to society." *Id.*

199. *Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 59.

200. *Approved Changes to the Standards Approval of Law Schools and Associated Interpretations*, SYLLABUS (Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar), Fall 2005, at 73–74 [hereinafter *2005 Approved Changes*]; *see also Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 63 (providing commentary on the adoption of the change).

201. *Commentary on Revisions to Standards 2004–2005*, *supra* note 186, at 59; CONSULTANT'S 2004–2005 REPORT, *supra* note 187, at 16–17.

which are in effect at the time this article is being written, retain the language that "[a] law school shall afford to full-time clinical faculty members a form of security of position reasonably similar to tenure" but note that where a school chooses a system of long-term contracts "'long-term contract' means at least a five-year contract that is presumptively renewable or other arrangement sufficient to ensure academic freedom."[202]

---

202. *2005 Approved Changes*, *supra* note 200, at 73–74. Current Standard 405(c) and the relevant Interpretations state:

Standard 405(c):

A law school shall afford to full-time clinical faculty members a form of security of position reasonably similar to tenure, and non-compensatory perquisites reasonably similar to those provided other full-time faculty members. A law school may require these faculty members to meet standards and obligations reasonably similar to those required of other full-time faculty members. However, this Standard does not preclude a limited number of fixed, short-term appointments in a clinical program predominantly staffed by full-time faculty members, or in an experimental program of limited duration.

Interpretation 405-6:

A form of security of position reasonably similar to tenure includes a separate tenure track or a program of renewable long-term contracts. Under a separate tenure track, a full-time clinical faculty member, after a probationary period reasonably similar to that for other full-time faculty, may be granted tenure. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the entire clinical program.

A program of renewable long-term contracts *shall* ~~should~~ provide that, after a probationary period reasonably similar to that for other full-time faculty, *during which the clinical faculty member may be employed on short-term contracts*, the services of a faculty member in a clinical program may be either terminated or continued by the granting of a long-term *renewable* contract ~~that shall thereafter be renewable~~ that shall thereafter be renewable. *For the purposes of this Interpretation, "long-term contract" means at least a five-year contract that is presumptively renewable or other arrangement sufficient to ensure academic freedom*. During the initial long-term contract or any renewal period, the contract may be terminated for good cause, including termination or material modification of the entire clinical program.

Interpretation 405-7:

In determining if the members of the full-time clinical faculty meet standards and obligations reasonably similar to those provided for other full-time faculty, competence in the areas of teaching and scholarly research and writing should be judged in terms of the responsibilities of clinical faculty. A law school should develop criteria for retention, promotion, and security of employment of full-time faculty.

Interpretation 405-8:

A law school shall afford to full-time clinical faculty members *participation* . . . in *faculty meetings, committees, and other aspects of* law school governance in a manner reasonably similar to other full-time faculty members. This interpretation does not apply to those persons referred to in the last sentence of Standard 405(c).

In commentary on the revisions, the Consultant on Legal Education noted that the initial proposal on Interpretation 405-6 drew a large number of comments, "sparked considerable debate," and produced an effort "to provide clarity and transparency that reconciled the language of Standard 405(c), requiring that law schools afford to full-time clinical faculty members a form of security of position reasonably similar to tenure, with the special constraints on providing such security as articulated by a number of law school deans."[203] The Consultant's commentary stated:

> The new definition of long-term contract—a contract of at least five years that is presumptively renewable—might be viewed as identifying a clear "safe harbor" that is consistent with the black letter law of Standard 405(c). The alternative avenue might be viewed in part as responsive to concerns of deans that flexibility must be preserved to allow schools to demonstrate that they meet the spirit and intent of the Standard by a route other than a five-year presumptively renewable contract and in part as responsive to expressions by clinical faculty of the importance of protecting academic freedom of clinical faculty.[204]

Unfortunately, the Council's failure to solicit input from the Standards Review Committee or the public before inserting the phrase "or other arrangement sufficient to ensure academic freedom" into Interpretation 405-6 has resulted in even less "clarity and transparency" about the appropriate means to provide security of position for clinical faculty.

VI. AFTERMATH OF THE 2005 CHANGES TO THE INTERPRETATIONS

The changes to the accreditation Interpretations in 2005 still have not settled the clinical faculty status issue. Since that time, there has been continued resistance to treating clinical faculty reasonably similar to non-clinical faculty, apparent difficulty in applying the new Interpretations, and additional debate over status for clinical faculty.

A. The Challenges Before the U.S. Department of Education

Facing difficulty in persuading the ABA to drop efforts to improve the status of clinical faculty and fully integrate clinical legal education in law

---

Interpretation 405-9:

> Subsection (d) of this Standard does not preclude the use of short-term or non-renewable ~~or non-renewal~~ contracts for legal writing teachers, *nor does it preclude law schools from offering fellowship programs designed to produce candidates for full-time teaching by offering individuals supervised teaching experience.*

Id. (italics, underscores, and strike-outs in original); *see also* 2007 STANDARDS, *supra* note 1, at 34 (including the same version of Standard 405(c)).

203. CONSULTANT'S 2004–2005 REPORT *supra* note 187, at 61.

204. *Id.*

schools, ALDA has challenged the certification of the ABA as the accrediting agency for legal education.[205] In March 2006, the ALDA's board of directors argued in a letter to the U.S. Department of Education, which was considering renewal of the ABA's status as the accrediting agency, that all Standards that presuppose a system of tenure or a tenure-like alternative should be revised or rescinded.[206] The ALDA's board focused in particular on security of position for clinical faculty: "Standard 405(c) is an unnecessary intrusion into the economic relationship amongst the law schools and those who run their clinical programs."[207] In June 2007, the Secretary of Education "re-recognized" the ABA as the accrediting agency but limited the recognition period to eighteen months because of concerns over Standard 211 and diversity issues.[208]

B. The Accreditation Committee's New Approach to Clinical Faculty: Short-Term Contracts and No Meaningful Participation in Faculty Governance

In another significant development in the evolution of clinical faculty standards, during the same period that the ALDA board of directors was opposing Standard 405(c), the Accreditation Committee reviewed the results of an ABA site visit and found that Northwestern University School of Law was not in compliance with Standard 405(c).[209] The Accreditation Committee's 2004 decision noted three bases for Northwestern's noncompliance:

> (1) full-time clinical faculty members are not afforded a form of security of position similar to tenure; (2) long-term contracts that are renewable are not granted after a probationary period reasonably similar to that for all other full-time faculty; and (3) full-time clinical faculty members are not afforded

---

205. Public Comment, Am. Law Deans Ass'n, Application of the American Bar Association for Reaffirmation of Recognition by the Secretary of Education as a Nationally Recognized Accrediting Agency in the Field of Legal Education (Mar. 2006) [hereinafter ALDA Public Comment] (on file with authors).

206. *Id.* at 1–2; Leigh Jones, *ABA's Tenure Power is Disputed*, NAT'L L.J., Apr. 3, 2006, at 1, 12 (noting that "[a]lthough ALDA asserts that the ABA should not have a say in the terms and conditions for employment of all law school professionals, the groups' comment focuses on tenure for clinicians and library professionals."). A news item reports that the general membership of ALDA decided in January 2007 not to endorse the position of ALDA's board of directors toward ABA accreditation standards. *Threat to Tenure at Law Schools*, INSIDE HIGHER ED, May 4, 2007, http://www.insidehighered.com/layout/set/print/news/2007/05/04/abatenure (last visited Apr. 2, 2008). The same article reports that "at least one member [of the ALDA board of directors] does not recall a formal vote [by the board on the ABA accreditation issues], and another said he doesn't believe it was unanimous." *Id.*

207. ALDA Public Comment, *supra* note 205, at 4.

208. Memorandum from Hulett H. Askew, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools et al. 5–6 (Aug. 21, 2007) (on file with authors).

209. Decision of the Am. Bar Ass'n Accreditation Comm., *supra* note 2, at 1–2.

> an opportunity to participate in law school governance in a manner reasonably similar to other full-time faculty members.[210]

Northwestern did not appeal the Accreditation Committee's initial decision, and the school was required to report back to the Committee by March 2005 on the steps taken to achieve compliance.[211] In a March 2005 letter, Northwestern reported that it had been able to attract "dedicated and active clinical faculty" under its current system and had not made any changes in response to the Accreditation Committee's action.[212] At its April 2005 meeting, the Accreditation Committee found that the law school had still failed to demonstrate compliance with Standard 405(c) and ordered it to show cause as to why it should not be placed on probation or removed from the list of ABA-approved law schools.[213]

The law school sent letters to the Accreditation Committee in September and October of 2006 arguing that although only seven of its thirty-eight clinical faculty had tenure or contracts of more than one year, the remaining thirty-one clinical faculty on one-year contracts had a form of security of position that was reasonably similar to tenure because the university had an academic freedom policy that the law school followed.[214] The law school argued that it complied with Standard 405(c) because of the August 2005 revision to Interpretation 405-6, which stated that "'long-term contract' means at least a five-year contract that is presumptively renewable or other arrangement sufficient to ensure academic freedom."[215]

The Accreditation Committee agreed. In reaching its decision, the Committee read the provision "other arrangement sufficient to ensure academic freedom" as a completely separate avenue for ensuring security of position reasonably similar to tenure.[216] The Accreditation Committee's action equates "other arrangement sufficient to ensure academic freedom" with "long-term contract," which the same sentence in Interpretation 405-6 defines first as a "five-year contract that is presumptively renewable." CLEA explained the problem with the Committee's reading:

> Northwestern's short contracts have been read by the Committee to be long-term contracts. Essentially, under the Committee's ruling, a law school can have one-day, at will contracts that have academic freedom protections; however, this is not consistent with the "form of security of position

210. *Id.*
211. *Id.* at 2.
212. *Id.*
213. *Id.*
214. *Id.* at 3–4.
215. *Id.*
216. *Id.* at 4.

reasonably similar to tenure" in both Standard 405(c) and Interpretation 405-6.[217]

The Accreditation Committee's decision also briefly addressed the issue of participation in faculty governance required by Interpretation 405-8. In response to a request from the Committee, Northwestern explained that the vast majority of clinical faculty do not have any vote in faculty governance since participation in governance is accorded only to tenured or tenure-track faculty, although other clinical faculty do serve on faculty committees other than those dealing with appointment and tenure of faculty.[218] Without explanation, the Accreditation Committee ultimately concluded that Northwestern had demonstrated compliance with Interpretation 405-8 even though the overwhelming majority of clinical faculty has no vote in faculty governance.[219] This lack of explanation makes it impossible to understand how the Committee interpreted and applied the requirement that clinical faculty members at Northwestern shall be afforded participation in law school governance "in a manner reasonably similar to other full-time faculty members."

The Accreditation Committee's approval of Northwestern's treatment of clinical faculty has rekindled the more than twenty-five year long debate over whether clinical faculty should be treated reasonably similar to other full-time law faculty. Indeed, the Accreditation Committee wrote to the Standards Review Committee in February of 2007 noting "that there continues to be much debate about just what is required to comply with Standard 405(c) with respect to security of position reasonably similar to tenure" and that Interpretation 405-6's language including "or other arrangement sufficient to ensure academic freedom" creates uncertainty.[220] The Accreditation Committee, and later the Council, requested the Standards Review Committee to review the matter and clarify Interpretation 405-6.[221] In addition, the Council voted in August 2007

---

217. Letter from Paulette J. Williams, President, Clinical Legal Educ. Ass'n, to William R. Rakes, Chair, Am. Bar Ass'n Section of Legal Educ. and Admissions to the Bar (Mar. 5, 2007) (on file with authors).

218. Decision of the Am. Bar Ass'n Accreditation Comm., *supra* note 2, at 2. Interpretation 405-8 states: "A law school shall afford to full-time clinical faculty members participation in faculty participation in faculty meetings, committees, and other aspects of law school governance in a manner reasonably similar to other full-time faculty members." 2007 STANDARDS, *supra* note 1, at Interpretation 405-8.

219. Decision of the Am. Bar Ass'n Accreditation Committee, *supra* note 2, at 3.

220. Am. Bar Ass'n Standards Review Comm., Draft Revisions to Standards for Approval of Law Schools and Explanation of Amended Interpretation 405-6 (attached to E-mail from Hulett Askew, Consultant on Legal Educ. to the Am. Bar Ass'n, to Michael Pinard, President, Clinical Legal Educ. Ass'n (Feb. 15, 2008)) [hereinafter Draft Revisions and Explanation of Amended Interpretation 405-6] (on file with authors).

221. *Id*; Memorandum from Richard Morgan, Chair, Standards Review Comm., & Hulett Askew, Consultant on Legal Educ. to the Am. Bar Ass'n, to Deans of ABA-Approved Law Schools et al. (Aug. 21, 2007) (on file with authors) (identifying the Committee's agenda for academic year 2007–2008); E-mail from Mark Aaronson to Clinical Legal Educ. Ass'n Board of

to form a special committee to look at the issue of security of position and governance rights for clinicians.[222]

After considering the ambiguous language added by the Council in 2005, the Standards Review Committee unanimously approved and forwarded to the Council a revised version of Interpretation 405-6 that provided:

> A form of security of position reasonably similar to tenure includes a separate tenure track or a program of renewable long-term contracts sufficient to ensure academic freedom. Under a separate tenure track, a full-time clinical faculty member, after a probationary period reasonably similar to that for other full-time faculty, may be granted tenure. After tenure is granted, the faculty member may be terminated only for good cause, including termination or material modification of the entire clinical program.
>
> A program of renewable long-term contracts shall provide that, after a probationary period reasonably similar to that for other full-time faculty, during which the clinical faculty member may be employed on short-term contracts, the services of a faculty member in a clinical program may be either terminated or continued by the granting of a long-term renewable contract. For the purposes of this Interpretation, "long-term contract" means a contract for a term of at least a five-years ~~contract~~ that is presumptively renewable or includes other provisions ~~arrangement~~ sufficient to ensure academic freedom. During the initial long-term contract or any renewal period, the contract may be terminated for good cause, including termination or material modification of the entire clinical program.[223]

The accompanying explanation stated that the proposed amendment makes clear that "a one year contract plus a policy on academic freedom is not sufficient under this Standard [405(c)]."[224]

After considering the proposed amendment in February 2008, the Council decided to postpone any action until after the report from the special committee on security of position in the summer of 2008.[225] This postponement means that the Accreditation Committee will likely continue to experience difficulty in applying Standard 405(c) and Interpretation 405-6.

Directors (May 17, 2007) (on file with authors) (reporting the actions of the Standards Review Committee at its May 16, 2007 meeting).

222. Paulette J. Williams, *President's Message*, CLEA NEWSLETTER (Clinical Legal Educ. Ass'n, New York, N.Y.), Sept. 2007, at 1 (on file with authors).

223. Draft Revisions and Explanation of Amended Interpretation 405-6, *supra* note 220 (underscores in original).

224. *Id.*

225. E-mail from Michael Pinard, President, Clinical Legal Educ. Ass'n, to lawclinic@lists.washlaw.edu (Feb. 14, 2008) (on file with authors) (reporting on ABA actions concerning Interpretation 405-6); E-mail from Dan Freeling, Deputy Consultant on Legal Educ. to the Am. Bar Ass'n, to Peter Joy (Feb. 15, 2008) (on file with authors) (confirming reports of ABA actions concerning proposed amendments to Interpretation 405-6).

C. *The ABA Accreditation Task Force's Statement on Eliminating "Security of Position" from the Standards*

The continued debate over the status of clinical faculty has not been confined to ALDA's efforts to pressure the ABA through the Department of Education, nor by the Accreditation Committee's uncertain approach in applying the Standard. Recently, a special ABA Accreditation Task Force was charged with looking at accreditation from a policy perspective and over a quarter of the report focused on the "security of position" issue, principally as contained in Standard 405(c).[226] The report noted that although tenure or a form of position reasonably similar to tenure is not explicitly required in standards of other accrediting bodies, clinical law faculty may be distinguishable "because of documented history of repeated attempts at outside interference with litigation and other forms of advocacy by law school clinics."[227]

The Task Force was unable to reach a consensus on a recommendation concerning "security of position," but a majority signed on to the following assessment of the issue:

> Even if the existing system is imperfect, it is far from self-evident that adequate alternative mechanisms can be fashioned. The removal of all "security of position" provisions from the Standards would have implications that go far beyond simply allowing law schools to determine for themselves whether to have a tenure system for doctrinal faculty or one that affords "a form of security of position reasonably similar to tenure" for clinical faculty. If the current provisions are deleted, and no other provisions for "security of position" are promulgated, a law school could choose to staff all or a major part of its programs with faculty members who serve as at-will employees or in some similar capacity.... It seems highly doubtful that such arrangements would promote the goals of a sound program of legal education, academic freedom, and a well-qualified faculty. In the absence of any specific standard, however, that would have to be determined on a case-by-case basis. If that inquiry were taken seriously, the likely result would be an accreditation process far more intrusive, costly, and labor-intensive than that which currently exists. On the other hand, if that inquiry were not taken seriously, there would be little point in having an accreditation process at all.[228]

The Task Force's assessment of the issue recognizes the difficulty in seeking to undo, without a demonstrated need for change and a suitable alternative, a Standard that has evolved by a consensus on the Council over more than two decades.

---

226. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, REPORT OF THE ACCREDITATION POLICY TASK FORCE 1, 17 (May 29, 2007) (on file with authors).

227. *Id.* at 22.

228. *Id.*

The Council agreed with the Task Force's assessment and has appointed a Special Committee on Security of Position, which is scheduled to issue an interim report by May 2008. The Special Committee's charge is to explore whether "security of position" language could be eliminated and some other language be inserted in either the Standards or Interpretations that would protect "academic freedom, attraction and retention of well-qualified faculty, and 'ensur[e] that law school governance decisions that can affect curriculum will have the benefit of the comments of sectors of the law school faculty whose knowledge and perspective otherwise might be unrepresented.'"[229] In its deliberations over new possible language, the Committee is to consider whether the proposed provisions will "serve the interests underlying the existing 'security of position' provisions as effectively, more effectively, or less effectively than the existing provisions[.]"[230] Clearly, the history of clinical faculty standards did not end with the 2005 revisions to Standard 405(c), as more proposed changes may be forthcoming.

VII. CONCLUSION

The history of the ABA Standard addressing the status of clinical faculty demonstrates that the value of clinical legal education and the faculty teaching those courses has long been contested. The historical record indicates that Standard 405(c), originally labeled 405(e), was first and foremost premised on the need, recognized by prominent members of the legal profession and numerous ABA committees and reports, that to further the development of clinical legal education it was necessary to integrate faculty teaching clinical courses into the law school through long-term employment relationships and participation in law school governance. In addition, prominent deans on the ABA Council in 1984 maintained that tenure or a reasonable equivalent was essential to securing academic freedom,[231] the AALS added its voice to support tenure as a means of guaranteeing academic freedom in 1999,[232] and the Council reaffirmed this position in 1999 and again as recently as 2004.[233]

In the intervening years since the adoption of Standard 405(c), clinical legal education has become more integrated into the typical law school curriculum. Clinical programs are featured prominently in most law school admissions materials, websites, magazines, and brochures. Commentators writing about the history of legal education in the United States note that, of all of the

229. Memorandum from Chief Justice Ruth McGregor, Chairperson, Council on Legal Educ. and Admissions to the Bar, to Special Committee Appointees and Interested Legal Education Organizations (Oct. 8, 2007), *available at* http://www.abanet.org/legaled/committees/SpecialCommitteeAppointment.doc (last visited Mar. 24, 2008).

230. *Id.*

231. *See supra* notes 110–18 and accompanying text.

232. *See generally* 1999 Standards Review Committee Hearing, *supra* note 170 (providing the decision of the accreditation committee).

233. *See supra* notes 175–76 & 181–85 and accompanying text.

curricular developments since the introduction of the casebook method, clinical legal education is the most significant.[234] In the face of this progress of and recognition for clinical legal education, one might expect that the faculty teaching those courses would be fully integrated into today's law schools and that as an accreditation matter, the status of clinical faculty would be well settled.

History demonstrates, however, that no other accreditation issue has been as contentious as the ABA's efforts to secure reasonably similar treatment of clinical faculty with their classroom faculty counterparts. Integration of clinical faculty into the governance life of law schools as a means of encouraging the development of clinical legal education has faced continuous opposition. At first, those opposed to giving security of position and a voice in law school governance to clinical faculty based their arguments on the then newness and experimental nature of clinical programs. Later, the argument against making clinical courses and the faculty teaching those courses integral parts of the law school shifted to an attack on accreditation standards that were perceived to infringe on law school autonomy. Despite the repeated findings of ABA commissions and numerous Standards Review Committee and Council recommendations that law schools fully integrate clinical faculty, their status within the legal academy remains uncertain.

The well-publicized action by the Accreditation Committee concerning its recent application of Standard 405(c) and its Interpretations, particularly Interpretation 405-6, to Northwestern University School of Law initially prompted this investigation into the history and development of Standard 405. The Accreditation Committee action approving one-year contracts for clinical faculty and no meaningful participation in faculty governance appears to negate more than two decades of work by the ABA's Standards Review Committee and Council on this Standard, as well as the salutary effects of this Standard on the development of clinical legal education. In one sense it can be said that clinical legal education has "come of age." But relative to other law school courses that maturation has been institutionally frozen at a point of permanent adolescence in those law schools that deny equal treatment of clinical courses and faculty.

The last intensive review of Standard 405 and its Interpretations in 2005 responded to requests by the Council for the Standards Review Committee to consider the meaning of "renewable" in Interpretation 405-6's reference to a system of long-term contracts for clinical faculty. It sought to resolve the lack

234. *See, e.g.*, PHILIP G. SCHRAG & MICHAEL MELTSNER, REFLECTIONS ON CLINICAL LEGAL EDUCATION 5 (1998) (stating that clinical legal education is "so often called the most significant change in how law was taught since the invention of the case method that it now sounds trite"); STEVENS, *supra* note 19, at 211 (stating that "[o]f all the renewed interest in skills, the particular interest in the skills embraced in the concept of clinical legal education was to prove the most important").

of agreement about whether "renewable" means "presumptively renewable" or "capable of being renewed."[235]

In the years prior to 2005, the Accreditation Committee had drifted toward approving three-year contracts as long-term contracts.[236] The specific incorporation by the Standards Review Committee of five-year contracts as a definition for long-term contracts marked an effort to provide "greater security of position than the Accreditation Committee's practice" and was "designed to achieve the goal of Standard 405(c), i.e., to ensure that law schools can attract and retain quality full-time clinical faculty and thereby strengthen the clinical component of the law school curriculum."[237] The action by the Accreditation Committee approving one-year contracts is in direct conflict with the Council's longstanding aim of providing greater security of position to clinical faculty and belies any meaningful understanding of the phrase "long-term." The recent proposal by the Standards Review Committee to amend Interpretation 405-6 to clarify that one-year contracts do not provide security of position sufficient to protect academic freedom and integrate clinical faculty into law schools is a rejection of the Accreditation Committee's action and affirms the plain language of Standard 405(c).

Not only is the Accreditation Committee's recent decision at odds with the history of Standard 405, but it reinforces a marginalization of both clinical courses and faculty teaching those courses in legal education. A recent Carnegie Foundation for the Advancement of Teaching report on legal education argues that the failure to fully incorporate clinical faculty and clinical courses into the law school sends a message to law students that such courses are not valued.[238] The report notes that such courses are usually taught by "a faculty that is not typically tenured and that has lower academic status. In many of the schools we visited, students commented that faculty view courses directly oriented to practice as of secondary intellectual value and importance."[239] The Carnegie finding substantiates the more than twenty-five year concern that unless faculty teaching clinical courses have security of position and participation in faculty governance reasonably similar to that of other full-time law faculty, clinical legal education will never be a truly valued part of law school education and will never achieve its full potential in teaching law students the skills and values of the legal profession.

---

235. *See supra* notes 191–92 and accompanying text.

236. *See supra* note 192 and accompanying text.

237. Memorandum from Sebert & Burke to Deans of ABA-Approved Law Schools, *supra* note 189, at 4; *Proposed Revision of Chapter 4*, *supra* note 189, at 12.

238. WILLIAM M. SULLIVAN ET AL., EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW 87–88 (2007). This observation is consistent with the observation of other commentators that "although clinical legal education is a permanent feature in legal education, too often clinical teaching and clinical programs remain at the periphery of law school curricula." Barry, Dubin & Joy, *supra* note 29, at 32.

239. SULLIVAN ET AL., *supra* note 238, at 88.

The history of the Standards for clinical faculty demonstrates that although some in legal education have been resistant, the ABA has long supported the full integration of clinical courses and the faculty teaching those courses into law schools. The history shows an unbroken movement by the ABA toward a system that provides a long-term relationship between the clinical faculty member and the law school so that the clinical faculty member has job security and the ability to participate in faculty governance comparable to other full-time law faculty teaching doctrinal courses. As a majority of the members on the recent ABA Accreditation Task Force concluded, any change to Standard 405(c) should not occur without a demonstrated need for change and a suitable alternative.

A TALE OF TWO CASE METHODS

BENJAMIN H. BARTON*

INTRODUCTION

Legal education is approaching a crossroads of sorts. The Carnegie Foundation for the Advancement of Teaching just released a quite critical review of American law schools.[1] Roy Stuckey recently published *Best Practices for Legal Education*, which also highlights the many gaps in law school coverage.[2] Both of these studies note how little law school does to prepare students for the actual practice of law.[3]

Given these criticisms, now is an apt time to consider how other professional schools prepare students to practice their professions.[4] Law faculties tend to be rather insular; even insofar as they consider the work and scholarship of other disciplines, they tend to focus on the scholarship of academic disciplines such as economics, literary theory, or evolutionary biology, rather than on the educational approaches of other professional schools.[5] This is a shame, because law schools can learn much about pedagogy from other professional schools.[6]

---

* Associate Professor of Law, University of Tennessee College of Law. B.A., Haverford College, 1991; J.D., University of Michigan, 1996. The author gives special thanks to Indya Kincannon, Mae Quinn, Paula Williams, Doug Blaze, Patience Crowder, Deirdre O'Connor, Dean Rivkin, Penny White, Stephen Rosenbaum, Kristin Henning, Cynthia Adcock, and the participants in the University of Tennessee College of Law's celebration of the Clinic's sixtieth anniversary; the University of Tennessee College of Law for generous research support; and the Honorable Diana Gribbon Motz.

1. *See generally* WILLIAM M. SULLIVAN ET AL., EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW (2007) (discussing the weaknesses of legal education and training).

2. *See generally* ROY STUCKEY & OTHERS, BEST PRACTICES FOR LEGAL EDUCATION (2007) (discussing shortcomings of and methods for improving legal education).

3. *See* STUCKEY, *supra* note 2, at 11–13; SULLIVAN, *supra* note 1, at 30–34.

4. A cynic might note that the time to discuss law school teaching has been apt for awhile. Professors Peter Joy and Bob Kuhn present a lengthy history of one small front of this battle. *See generally* Peter A. Joy & Robert R. Kuehn, *The Evolution of ABA Standards for Clinical Faculty*, 75 TENN. L. REV. 183 (2008) (explaining the history of the ABA's accreditation standards regarding tenure and participation in law school governance for clinical faculty). The Carnegie Foundation's earlier report on this topic is another great companion piece. *See generally* ALFRED ZANTZINGER REED, TRAINING FOR THE PUBLIC PROFESSION OF THE LAW (1921) (echoing concerns about students' preparation for the practice of law).

5. *See generally, e.g.*, LAW & EVOLUTIONARY BIOLOGY (Lawrence A. Frolik et al. eds., 1999) (discussing the role of evolutionary biology scholarship in the law, but not the pedagogy of graduate programs in biology); SANFORD LEVINSON & STEVEN MAILLOUX, INTERPRETING LAW AND LITERATURE: A HERMENEUTIC READER (1988) (same with respect to literary theory); RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW (7th ed. 2007) (same with respect to

This Essay argues that adopting the business school case method would substantially improve law school teaching. Part I presents brief descriptions of the business school and law school versions of the case method. Part II argues that the business school case method would constitute a large improvement to law school instruction. Part III asserts that one aspect of legal education, the legal clinic, admirably serves many of the goals of the business school case method. Finally, Part IV contends that the root cause of the pedagogical differences between business schools and law schools are a result of market forces rather than any underlying differences in the professions.

I. TWO CASE METHODS

The history of both case methods begins with Christopher Langdell becoming the Dean of Harvard Law School in 1870.[7] He immediately instituted the case method as the law school's educational model.[8] The law school case method in 1870 was remarkably similar to that employed today: Students were required to read leading cases by themselves in an effort to distill the fundamental principles of law.[9] In class, students were led in discussions of these cases via the Socratic method, a series of professorial questions meant to teach the students how to identify and focus on the rules of law found within the cases.[10]

Langdell saw the study of law as a scientific endeavor,[11] contrary to the then-prevailing model of lawyer training, which was practice-oriented and based on the apprenticeship model.[12] From Harvard, Langdell's model spread nationally; by the early twentieth century, the case method was the dominant model in law schools and formal education had largely replaced apprenticeships.[13]

A couple of features of the law school case method are worth noting at the outset. First, the law school case method sought to be a form of scientific or academic training, not practical career training.[14] Second, case method

---

economic analysis).

6. *See, e.g.*, STUCKEY, *supra* note 2, at 47–48 (suggesting that a problem solving approach such as that used in business schools should be incorporated into legal education).

7. *See* Russell L. Weaver, *Langdell's Legacy: Living with the Case Method*, 36 VILL. L. REV. 517, 520 (1991).

8. *Id.*

9. *Id.* at 526–28.

10. *See* Carrie Menkel-Meadow, *Taking Law and _____ Really Seriously: Before, During and After "The Law,"* 60 VAND. L. REV. 555, 561 (2007).

11. Weaver, *supra* note 7, at 527.

12. *See, e.g.*, Jeffrey D. Jackson, *Socrates and Langdell in Legal Writing: Is the Socratic Method a Proper Tool for Legal Writing Courses?*, 43 CAL. W. L. REV. 267, 269–70 (2007); Weaver, *supra* note 7, at 522, 527.

13. *See* Edward Rubin, *What's Wrong with Langdell's Method, and What to Do About It*, 60 VAND. L. REV. 609, 612–13 (2007); Weaver, *supra* note 7, at 541–43.

14. *See* Weaver, *supra* note 7, at 527–29, 531.

professors are predominantly academics selected for their scholarly abilities and potential, not necessarily practitioners chosen for their real-world experience.[15] Third, the casebooks consist of selected appellate-level cases with little additional textual material.[16] Lastly, the class grades are determined by a written, final exam based upon hypotheticals.[17]

The history of the business school case method likewise starts at Harvard. Harvard Business School (HBS) was founded in 1908; from the outset, it taught using the case method.[18] The business school case method was partly inspired by the success of the Harvard Law School case method.[19] Similarly to its law school counterpart, the business school case method spread from Harvard to other schools and has now become the primary national model for graduate business education.[20]

Despite originating from the law school version, the business school case method is quite distinct. The single biggest difference is the meaning of the word "case." In law school, the case method concerns itself with written, appellate court decisions.[21] Business school case files, on the other hand, are by and large drawn from real-life scenarios,[22] and the students are asked to place themselves in the role of a manager in charge of the scenario.[23] A typical business case file is between ten and thirty pages.[24] The case file presents a mass of data—which are sometimes conflicting—to the student, who must make a managerial decision based on the information.[25] Given this method of

---

15. *See, e.g.*, Menkel-Meadow, *supra* note 10, at 562 (describing "the modern professoriate" as "a breed of professionally separate academic lawyers who were not expected to practice or know much about the real world of practice, but were (and are) chosen for their academic excellence as students and their presumed intellectual acuity, if not for their teaching ability or legal professional achievements, and who developed into 'outsider' critics of the law as written and practiced"); Rubin, *supra* note 13, at 614 ("All of the earthly rewards that a faculty member can obtain . . . depend on scholarly production, not on teaching.").

16. *See* Menkel-Meadow, *supra* note 10, at 561–62.

17. *Id.* at 562.

18. *See* MELVIN T. COPELAND, AND MARK AN ERA: THE STORY OF THE HARVARD BUSINESS SCHOOL 27–28 (1958). At the outset, the case method was also known as the "problem method." *Id.* at 27.

19. *Id.* at 28.

20. *See* ROBERT AARON GORDON & JAMES EDWIN HOWELL, HIGHER EDUCATION FOR BUSINESS 368 (1959). By 1959, the case method was already the dominant model of graduate business education. *See id.* (noting the already rapidly spreading use of the case method in business schools at the time of publication).

21. Menkel-Meadow, *supra* note 10, at 561–62.

22. *See* Harvard Business School, Making a Case: The Birth of an HBS Case Study, http://www.hbs.edu/corporate/enterprise/case.html (last visited Feb. 19, 2008) (describing the development of HBS case files).

23. DAVID W. EWING, INSIDE THE HARVARD BUSINESS SCHOOL 20 (1990).

24. *Id.*

25. *See id.* at 21–23 (describing a typical case file). *See generally* Samuel E. Bodily & Robert F. Bruner, Enron, 1986–2001: Supplement for the Instructor (2002),

instruction, management theory is rarely taught or lectured to the students directly; instead, students study management by actually practicing their skills in making managerial decisions.[26]

Business school classes usually handle no more than one case file per day, which the students have worked on in teams prior to class.[27] The case files are then discussed and dissected during class.[28] In fact, at HBS, class participation is fifty percent of a student's grade.[29] Other business schools place a greater emphasis on written work and team-created projects.[30] Throughout each of these variations on the business school case method, there is an effort to grade students on the quality of the work itself as much as on their understanding of any underlying theories.[31]

II. THE ADVANTAGES OF THE BUSINESS SCHOOL CASE METHOD

Despite the fact that the business school case method is a younger cousin to the law school version, I believe the adoption of some of the features of the business school case method would immeasurably improve law school pedagogy. Many of these criticisms cluster around the idea that law schools could and should do more to prepare students for the actual practice of law than they currently do. If you disagree with that premise, I do not expect to convince you.

A. Role and Process

The single biggest advantage to the business school case method is that it teaches the students management skills by placing them in the role of a manager and asking them to perform the same tasks that managers perform in practice. The classes thus are naturally more focused on the process of being a

---

http://ssrn.com/abstract=302155 (describing the Enron case file).

26. *See* EWING, *supra* note 23, at 24–25 (quoting Professor Charles Gragg as saying, "The outstanding virtue of the case system is that it is suited to inspiring activity, under realistic conditions, on the part of students; it takes them out of the role of passive absorbers and makes them partners in the joint processes of learning and of furthering learning.").

27. *See* EWING, *supra* note 23, at 31.

28. *See id.* at 31–32.

29. *See* Harvard Business School, How Does the Case Method Work?, http://www.hbs.edu/case/case-work.html (last visited Feb. 19, 2008).

30. *See, e.g.*, Frances Fabian, Management Policy Course Syllabus (Fall 2007), http://www.belkcollege.uncc.edu/fhfabian/MBA%20courses/MBAD6194Thurs.htm (stating that sixty percent of a student's grade is based on written work and team presentations); James A. Fitzsimmons, Service Management Syllabus 7 (Spring 2006), http://www.mccombs.utexas.edu/dept/irom/courses/syllabi_spg2006/index.asp (follow "MSC 386 1 Service Management (Fitzsimmons)" hyperlink) (stating that team assignments made up forty percent of a student's grade and thirty-five percent was based on individual written work).

31. *See generally* MICHAEL MASONER, AN AUDIT OF THE CASE STUDY METHOD 1–8 (1988) (describing the diversity of teaching applications of the case method).

manager. There is obviously a certain artificiality to a classroom setting, but outside of that limitation, the business school case method attempts to teach the necessary skills by simulating the actual work.[32]

By comparison, the law school case method generally does not focus on any professional role in particular. The students are tasked with distilling the rule of law from cases, but they generally are not asked to put themselves in the role of one of the players in any actual case.[33] Insofar as any particular role is discussed or considered, it is generally the role of the judge or justices in the case, not the lawyers.[34] Professors regularly ask law students to consider other ways a case might have been decided or what might be wrong with a certain decision.[35] Professors rarely ask a student how a lawyer in a particular case could have advocated differently or more effectively.[36]

On the contrary, reading appellate cases discourages students from thinking about the actual lawyering that underpins the cases because the cases, notes, questions, and answers in the casebooks mostly deal with the legal rules from each case, rather than any practical lawyering issues.[37] Further, appellate courts almost never discuss the strategic choices made by the lawyers on appeal, let alone the choices at trial. To the extent that students do think about the lawyering, it is usually the lawyering on appeal. In actual practice, though, the great bulk of litigation work occurs before trial and the fact remains that the vast majority of disputes never see trial, much less an appellate proceeding.[38] This means that law school classes focus on a small, and in some ways anachronistic, set of materials that reflect a small subset of practice skills that are rarely, if ever, necessary in actual litigation practice.

Moreover, this assumes that law schools aspire to teach litigation skills. There is much lawyering (such as transactional work, tax and intellectual

---

32. *See* C. ROLAND CHRISTENSEN WITH ABBY J. HANSEN, TEACHING AND THE CASE METHOD 24 (1987) ("To the extent that one can learn business practice in a classroom—and the limits are substantial—it achieves its goal efficiently.").

33. *See* ANTHONY T. KRONMAN, THE LOST LAWYER: FAILING IDEALS OF THE LEGAL PROFESSION 120–21 (1993).

34. *See id.*

35. *See* Susan Sturm & Lani Guinier, *The Law School Matrix: Reforming Legal Education in a Culture of Competition and Conformity*, 60 VAND. L. REV. 515, 526 (2007) ("The professor structures interactions with students by invoking the style of an appellate judge who questions lawyers to ferret out the weaknesses in their positions and validate winning arguments.").

36. *See id.* at 529 (noting that the law school case method creates an "overly . . . formalistic idea of law and a litigation model of lawyering" while "distorting . . . and marginalizing [the] value" of other lawyering approaches).

37. *Id.*

38. Marc Galanter, *The Vanishing Trial: An Examination of Trials and Related Matters in Federal and State Courts*, 1 J. EMPIRICAL LEGAL STUD. 459, 459 (2004) ("The portion of federal civil cases resolved by trial fell from 11.5 percent in 1962 to 1.8 percent in 2002, continuing a long historic decline. More startling was the 60 percent decline in the absolute number of trials since the mid 1980s.").

property compliance, and estate planning) that occurs outside the litigation context. The law school case method does little to prepare students on that front.[39] The law school case method *does* teach students how to read legal cases and to discern legal rules. The hope is that it also teaches students how to apply these legal rules in alternate factual situations.

Would it not be possible to teach the skill of application in a case method that more closely approximates the business school version? Could we not give students a set of facts, a portfolio of related law, and a necessary legal task, and then grade them on their actual work? While law school students would still be required to glean a legal rule from cases and apply it, this method would better prepare students to become practicing lawyers. Furthermore, students could learn and apply a broad range of practical skills, such as document drafting, negotiation, and pretrial litigation. In fact, the list of skills that could be covered in tandem with the substantive material is virtually endless, bounded only by the limitations of the classroom setting and the professor's willingness to create and grade the new case files.

Lastly, the alleged core strength of the law school case method is that it teaches students to "think like a lawyer." The business school case method, however, teaches students both how to learn and how to make difficult, real-life decisions.[40] Specifically, HBS "emphasizes learning over teaching," meaning that the students learn from preparing, discussing, and acting out real-life situations.[41] Further, business schools force their students to form the habit of actually making decisions and performing difficult managerial tasks.[42] Thus, while business students learn the familiar law school lessons of gray areas and indeterminacy,[43] they also acquire the additional—and more valuable—lesson that action must be taken despite imperfect information.[44]

B. Rational Grading and Team Building

Business school grading also seems more rational than law school grading because it better approximates the experiences Master of Business Administration (MBA) students will have when they graduate. MBA students are graded on the strength of their actual work, not on a single exam taken at the end of the semester.[45] Business schools also give students regular feedback

---

39. *See* Carol R. Goforth, *Use of Simulations and Client-Based Exercises in the Basic Course*, 34 GA. L. REV. 851, 853 (2000) (arguing that the traditional curriculum and the law school case method shortchange transactional work).

40. *See* EWING, *supra* note 23, at 13–14.

41. *Id.*

42. *Id.* at 25–26.

43. *See* Benjamin Barton, The Emperor of Ocean Park*: The Quintessence of Legal Academia*, 92 CAL. L. REV. 585, 593–95 (2004) (book review) (discussing the "Siren Song of Indeterminacy," the "[r]ecognition of [which] is crucial to understanding and practicing law").

44. *See* EWING, *supra* note 23, at 25–26.

45. *See, e.g.*, sources cited *supra* note 30.

on their work because written projects and class participation are graded throughout the course.[46]

By comparison, the grades in traditional law school classes are based on a single written exam at the end of the semester.[47] The typical questions on these exams require the students to "issue spot": Students receive a lengthy factual scenario and must apply the law studied that semester.[48] Traditionally, these exams are closed-book; although strict adherence to this format has softened in second- and third-year classes, it remains the dominant model, especially in the all-important first-year classes.[49] Nevertheless, this grading format has been widely criticized for failing to offer regular feedback and even inflicting psychological distress on first-year students.[50]

Also worrisome is the fact that traditional exams do not test skills that are particularly relevant in legal practice. I have often queried other professors as to whether they had ever experienced the following in practice:[51] A potential client comes in with a crazy set of facts that hits all sorts of different and disparate legal issues, and without looking at notes or doing any research, the lawyer must outline all the potential client's different possible claims in three hours. It is hard to guess when this might ever happen in practice.[52] Yet it

---

46. *See, e.g.*, sources cited *supra* note 30.

47. *See* Menkel-Meadow, *supra* note 10, at 562.

48. Linda R. Crane, *Grading Law School Examinations: Making a Case for Objective Exams to Cure What Ails "Objectified" Exams*, 34 NEW ENG. L. REV. 785, 785–86 (2000).

49. *See* Christian C. Day, Essay, *Law Schools Can Solve the "Bar Pass Problem"—"Do the Work!,"* 40 CAL. W. L. REV. 321, 349 (2004) (advocating a return to "traditional closed-book exams").

50. *See, e.g.*, Crane, *supra* note 48, at 786 ("During the typical law school examination, students are asked . . . to perform well on a single test that is worth 100% of their grade, and upon which their entire class standing and future careers rest."); Robert C. Downs & Nancy Levit, *If It Can't Be Lake Woebegone . . . A Nationwide Survey of Law School Grading and Grade Normalization Practices*, 65 UMKC L. REV. 819, 822–24 (1997) (criticizing the law school practice of the "single, end-of-semester" exam); Carol M. Parker, *A Liberal Education in Law: Engaging the Legal Imagination Through Research and Writing Beyond the Curriculum*, 1 J. ASS'N LEGAL WRITING DIRS. 130, 138 n.46 (2002) ("[T]he traditional one-exam evaluation of first-year doctrinal courses may imbue students with a sense of powerlessness.").

51. This assumes that the other professor has practiced law prior to teaching, which is not always true of American law professors. *See* Rodney J. Uphoff et al., *Preparing the New Law Graduate to Practice Law: A View from the Trenches*, 65 U. CIN. L. REV. 381, 397 (1997) ("Because a significant number of law professors never practiced law, or did so briefly in a large firm with minimal client contact, few law professors are familiar with or interested in the interpersonal aspects of lawyering.").

52. In a perverse way, it is fun to think of scenarios where such things might happen. Certainly thinking quickly on one's feet is important in oral argument before a trial or appellate court. My favorite hypothetical situation is as follows: A client comes in at two o'clock in the afternoon on the last day before the statute of limitations runs on a valuable litigation claim, and the lawyer has three hours to write out a complaint. Needless to say, this is not a practical everyday experience.

happens every fall and spring semester on most law school exams and is standard procedure for evaluating law students.

The great bulk of legal work allows time for research and careful thought on the issue at hand. To reinforce this notion, in the University of Tennessee's Advocacy Clinic, we spend time every semester teaching the students the regular use of the invaluable phrase: "I don't know the answer to that question off the top of my head, but I can research it and get back to you." Yet inside the classroom, the Socratic method and semester exams encourage students to answer every question as quickly and thoroughly as possible, frequently regardless of whether they truly know the answers. As a result, this emphasis on speed and snap judgments may do more than merely test only one legal skill; it may actually lead future lawyers to fly by the seat of their pants. In fact, it is not an exaggeration to say that eighty percent or more of a first-year law student's grade point average is based on testing an activity that is rarely, if ever, required in practice. While rapid legal analysis may be a useful ability for lawyers, the form of rapid legal analysis tested in law schools is quite unrealistic.

Again, the question is whether professors could test that skill—along with a wealth of other legal skills, such as careful legal research and writing—in a more realistic setting. The answer is a resounding yes. Adopting the business school case method would allow testing of important additional skills in settings that much more closely track actual practice.

Even assuming that an issue-spotting exam tested the most important practical legal skill, it would still be an odd fit for first-year classes. Traditional first-year classes spend very little time on issue spotting. Though some class time is reserved for answering hypotheticals, the first time many law students face written questions requiring issue spotting and legal analysis is during their first-semester exams. It is difficult to argue that a rational examination should rely so heavily upon skills and activities that are rarely covered in class, yet the traditional law school exam does exactly that.

The business school case method avoids this irrationality. In many classes, the majority of a student's grade is earned over the course of the semester.[53] The final exam is just a reprise of what was required in class: The students receive another case file to work on.[54] The students work in teams and at least a portion of the grades come from team projects.[55] Team learning is rational for business schools because few, if any, business managers work completely solo. Moreover, team projects teach MBA students the invaluable professional lesson that "[t]he group's best effort is almost always better than the individual's best effort."[56]

---

53. *See, e.g.*, sources cited *supra* note 30.

54. *See, e.g.*, sources cited *supra* note 30.

55. *See, e.g.*, sources cited *supra* note 30.

56. JEFFREY L. CRUIKSHANK, SHAPING THE WAVES: A HISTORY OF ENTREPRENEURSHIP AT HARVARD BUSINESS SCHOOL 349 (2005).

Similarly, requiring law students to work in teams as part of their grades would provide a much more realistic reflection of the current practice of law. While some lawyers do work as solo practitioners, the great majority of lawyers work in firms or multi-person government offices.[57] Therefore, teamwork is an essential, but largely untaught, skill.

Finally, some states are beginning to require the Multistate Performance Test ("MPT"), a practice-oriented case file exercise "designed to assess case planning, problem solving, factual investigation and other skills that are important to the competent practice of law."[58] In fact, the MPT exercise is quite similar to the method I propose in this Essay. Insofar as these skills are necessary for the bar exam—a precursor to practice—law schools should certainly teach them.

C. Flexibility and Openness to Other Disciplines

Ever since the case method became the dominant law school teaching model, there has been a growing dialogue about integrating more practical skills education into the law school curriculum.[59] More recently, there has been much discussion about internationalizing law schools, or at least exposing law school classes to some of the realities of globalization.[60] There has also been increasing momentum for interdisciplinary scholarship in law schools. For example, there have been long-standing efforts to create a context for professional responsibility and to teach the subject throughout the curriculum.[61] Nonetheless, law school classes generally remain focused on reading and analyzing cases without regard to other academic or professional disciplines.[62]

Law schools should consider why these various reforms tend to flounder. Because the first-year classes are kept basically unchanged, every law school innovation has a "grafted on" feeling, and the current structure of law school sends a clear message to students about what the faculty consider most valuable. A large reason that the various efforts at reforming legal education

---

57. *See* Andrew M. Perlman, *A Career Choice Critique of Legal Ethics Theory*, 31 SETON HALL L. REV. 829, 831 (2001).

58. *See, e.g.*, New York Bar Examiners, Multistate Performance Test (MPT), http://www.nybarexam.org/MPT.htm (last visited Nov. 5, 2007).

59. *See* Alberto Bernabe-Riefkohl, *Tomorrow's Law Schools: Globalization and Legal Education*, 32 SAN DIEGO L. REV. 137, 138–39 (1995) (referencing the ever-static nature of legal education and the need for practical change).

60. *See, e.g., id.* at 152–56 (theorizing about the future effects of globalization on legal education).

61. *See generally* DEBORAH L. RHODE, PROFESSIONAL RESPONSIBILITY: ETHICS BY THE PERVASIVE METHOD (2d ed. 1998) (applying professional responsibility standards to various subjects within the legal curriculum).

62. *See* Rob Atkinson, *Growing Greener Grass: Looking from Legal Ethics to Business Ethics, and Back*, 1 U. ST. THOMAS L.J. 951, 979 (2004) (noting similarities between business and law school case methods, except that "the case method in business schools has tended to foster, rather than forestall, interdisciplinarity").

have failed to make headway, especially in the first year of law school, is the inflexibility of the law school case method. The law school case method requires the use of leading appellate opinions as the source materials.[63] Because appellate opinions rarely consider international law, the practice of law, legal ethics, or other academic disciplines, law school professors are left at a loss as to how to integrate these other areas into their classes. Cases alone do not lend themselves to new or different approaches to legal education. Furthermore, teaching law students to "think like lawyers" frequently requires the students to discard their previous life experiences and expertise.[64]

By contrast, the business school case method is remarkably flexible. The source materials are drawn from real-life scenarios and students are asked to make the decisions that actual managers were forced to make.[65] Business school cases are limited only by the constraints of writing out the case itself; professors might find it difficult to realistically reduce the complexity of a critical business decision to a ten to thirty page written case file.[66] Still, the case files are uniquely flexible teaching tools because they cover a wide variety of circumstances and require students to perform any number of managerial tasks.[67]

Moreover, the flexibility of the business school case file encourages innovation and an interdisciplinary approach on several levels.[68] First, the professor must analyze each case from every possible perspective in order to guide students in developing the best response.[69] Second, the class is student driven; the perspectives of a variety of students coming from diverse backgrounds are critical components of a successful class.[70] In *Teaching and the Case Method*, the authors argue that each business school class "provides [the] opportunity for new intellectual adventure" that "also meets [the] faculty's

---

63. Menkel-Meadow, *supra* note 10, at 561–62.

64. *Cf.* Jerry J. Phillips, *Thinking*, 72 TENN. L. REV. 697, 735 (2005) (noting that a colleague told his first-year law students, "Your brains are mush right now; we are going to make you think like a lawyer!").

65. EWING, *supra* note 23, at 20.

66. *Cf. id.* (explaining that case files are typically taught and written by the professors).

67. *See* CHRISTENSEN, *supra* note 32, at 26 (discussing the "continuity and change" of the business school case method and how it provides its students with a vast range of experiences).

68. *See* Edward J. Conry & Caryn L. Beck-Dudley, *Meta-Jurisprudence: A Paradigm for Legal Studies*, 33 AM. BUS. L.J. 691, 731 (1996) ("Business schools focus on broad, real-world tasks . . . and every business task is evaluated by psychology, law, economics, finance, sociology, ethics, history, and/or mathematics.").

69. *See* EWING, *supra* note 23, at 27–28 (discussing the importance of a business professor's role in "constantly probing, questioning, hypothesizing, challenging, and rephrasing [students'] comments in an effort to help them analyze the case and understand its implications").

70. *See* Mimi Wolverton & Larry Edward Penley, *What It Takes to Be Strategically Innovative*, *in* ELITE MBA PROGRAMS AT PUBLIC UNIVERSITIES: HOW A DOZEN INNOVATIVE SCHOOLS ARE REDEFINING BUSINESS EDUCATION 17, 30–31 (Mimi Wolverton & Larry Edward Penley eds., 2004).

teaching and research needs. . . . [and] links instructors to the world of practice."[71]

The business school case method lends itself more naturally to the consideration of globalization and international issues. The business school case method has been adopted all over the world, and American business schools have been much faster to partner with foreign business schools.[72] The business school case method also is better situated for discussions of both professional practice and ethics. Reading an appellate opinion does little to encourage law students to consider how they would handle a particular case, ethical dilemma, or legal task, whereas the business school case method puts each of these questions front and center.[73]

As a historical matter, the enduring separation between Langdell's law school case method and practical training is quite interesting.[74] From the outset, law schools made every effort to be academic institutions rather than professional institutions.[75] At Harvard Law School there was a conscious break with the practical side of legal education, a schism that has yet to heal.[76] In fact, many law professors do not even see the schism as a valid shortcoming of legal education.[77]

In contrast, business schools send the message that the most important part of business school is to "learn by doing" in preparation for becoming a business manager.[78] Adopting the business school case method would allow for easier integration of ethics and skills into law school teaching and would "unite the clans" by allowing the practice of law once again to become central to law school.

D. Connecting Teaching and Scholarship

A growing body of empirical scholarship is beginning to address the connection, if any, between teaching and scholarship.[79] I have drafted a study

---

71. *See* CHRISTENSEN, *supra* note 32, at 24–25.

72. *See generally* ROBERT R. LOCKE, MANAGEMENT AND HIGHER EDUCATION SINCE 1940: THE INFLUENCE OF AMERICA AND JAPAN ON WEST GERMANY, GREAT BRITAIN, AND FRANCE 159–211 (1989) (analyzing the global influence of the American business school model).

73. *Cf.* CHRISTENSEN, *supra* note 32, at 3 ("Lectures about judgment typically have limited impact.").

74. *Cf.* Bernabe-Riefkohl, *supra* note 59, at 138–39 (referencing the lack of practical training within the law school case method).

75. *See* Weaver, *supra* note 7, at 529–31.

76. *See id.* at 529–31, 544.

77. *See id.* at 544 (examining the perpetual use of the case method by professors who themselves were taught using the case method and, thus, feel that "the case method is the proper way to teach law").

78. *See* CHRISTENSEN, *supra* note 32, at 23.

79. *See, e.g.*, Kenneth A. Feldman, *Research Productivity and Scholarly Accomplishment of College Teachers as Related to Their Instructional Effectiveness: A Review and Exploration*, 26 RES. HIGHER EDUC. 227, 275 (1987); John Hattie & H.W. Marsh, *The Relationship Between*

that shows little or no correlation between teaching evaluations and scholarly productivity among American law schools' faculty.[80] This finding is consonant with the great bulk of studies from other disciplines.[81]

The lack of a connection between scholarly productivity and teaching evaluations in American law schools has been quite puzzling for many legal academics, who assume that active knowledge of a field and a curious mind—two hallmarks of scholarly productivity—should translate well in the classroom.[82] One possible reason for the disconnect is how distinct teaching is from scholarly writing in today's law schools.

It is noteworthy, therefore, that proponents of the business school case method say that it "is intellectually stimulating for the faculty"[83] and helps "meet[] a faculty's teaching and research needs."[84] Some of the most influential recent works of business school scholarship look very similar to case studies.[85] Further, the enormously popular *Harvard Business Review* includes one case study in each issue along with other scholarly articles.[86] The *Harvard Law Review*, on the other hand, hardly ever features teaching materials or notes.

E. Harvard Business School Versus Yale Law School

Both American law schools and business schools have a single, unquestioned lead institution.[87] Among business schools, that leader is Harvard;[88] for law schools, it is Yale.[89] These schools are ranked first by organizations such as *U.S. News & World Report*, are among the most selective in student acceptance rates, and thus serve an important function as the "lead dog" in their respective areas.[90]

Research and Teaching: A Meta-Analysis, 66 REV. EDUC. RES. 507, 529 (1996).

80. Benjamin Barton, Is There a Correlation Between Scholarly Productivity, Scholarly Influence and Teaching Effectiveness in American Law Schools? An Empirical Study 2 (July 1, 2006) (unpublished paper presented at the First Annual Conference on Empirical Legal Studies), http://ssrn.com/abstract=913421.

81. *Id.*; *cf.* Feldman, *supra* note 79, at 275; Hattie & Marsh, *supra* note 79, at 529.

82. *See* Hattie & Marsh, *supra* note 79, at 511.

83. CHRISTENSEN, *supra* note 32, at 24.

84. *Id.* at 25.

85. For example, two well-known books by a former Stanford Business School professor, JAMES C. COLLINS, GOOD TO GREAT: WHY SOME COMPANIES MAKE THE LEAP . . . AND OTHERS DON'T (2001) and JAMES C. COLLINS & JERRY I. PORRAS, BUILT TO LAST: SUCCESSFUL HABITS OF VISIONARY COMPANIES (1994), are essentially large-scale case studies reduced to unifying themes.

86. The University of Chicago Library, Case Studies, http://www.lib.uchicago.edu/e/busecon/guides/casestudy.html (last visited Nov. 2, 2007).

87. *See* U.S. NEWS & WORLD REP., AMERICA'S BEST GRADUATE SCHOOLS: 2008 EDITION 20, 44 (2007).

88. *Id.* at 20.

89. *Id.* at 44.

90. *See id.* at 20, 44.

As an American law professor (but not a Yale graduate), I am relatively familiar with Yale Law School (YLS) through my colleagues while clerking, in practice, and most of all in teaching; however, I have only a passing familiarity with HBS. As a result, I was amazed when I first began to research this topic. As many a lazy researcher has done before, I started with a Google search: "What is the case method?" The very first result was an HBS website called, helpfully enough, "The Case Method."[91] This area of the HBS website features a full description of the case method, a video of a class, a portion of an actual HBS case file, and a glowing description of the case method.[92] The site describes how central the case method is to the HBS mission and experience.[93] Similarly, the many books written about HBS centrally and glowingly feature the case method.[94]

The YLS website, on the other hand, states nothing about the law school case method or the Socratic method. Both the YLS website and Anthony Kronman's *History of the Yale Law School* spend a significant amount of time discussing Yale's influence on legal academia and legal thought but are strikingly silent concerning pedagogy,[95] especially when compared to HBS's pride in the case method.

In sum, the business school case method has performed what a law school professor might think a miracle. It has made pedagogy a central element of an Ivy League professional school.

III. ADVANTAGE, LAW SCHOOLS: CLINICAL LEGAL EDUCATION

There are areas of legal education that use methods similar to the business school case method. One in particular, the legal clinic, is worth mentioning.[96] Clinical classes are actually slight improvements over the business school case

---

91. Harvard Business School, The Case Method, http://www.hbs.edu/case/ (last visited Nov. 2, 2007).

92. *Id.*

93. *Id.*

94. *See, e.g.*, CHRISTENSEN, *supra* note 32, at 24–25; COPELAND, *supra* note 18, at 254–72; CRUIKSHANK, *supra* note 56, at 347–50; EWING, *supra* note 23, at 19–21.

95. *See* Anthony T. Kronman, *Introduction* to HISTORY OF THE YALE LAW SCHOOL: THE TERCENTENNIAL LECTURES at ix, ix–xii (Anthony T. Kronman ed., 2004); Yale Law School, Intellectual Life, http://www.law.yale.edu/intellectuallife/intellectuallife.htm (last visited Nov. 2, 2007) (describing the great variety of Yale's centers and programs). Humorously, YLS does have a page entitled "Law Teaching," but it only describes how many YLS graduates become professors and how YLS prepares them for academia, with no mention of how the law is actually taught at Yale. Yale Law School, Law Teaching, http://www.law.yale.edu/lawteaching.htm (last visited Nov. 2, 2007). Some other law schools do describe their teaching methods on their websites. *See, e.g.*, The University of Chicago Law School, The Socratic Method, http://www.law.uchicago.edu/socrates/method.html (last visited Nov. 2, 2007).

96. *Cf.* Gary L. Blasi, *What Lawyers Know: Lawyering Expertise, Cognitive Science, and the Functions of Theory*, 45 J. LEGAL EDUC. 313, 387–88 (1995) (noting the similarities between law school clinics and the business school case method, but also noting differences).

method because clinics avoid its greatest weakness, the difficulty of "canning" real life into a written case file. Clinic students address real issues, interact with real clients, and present cases before real tribunals.[97] Thus, there is no need to attempt capturing the full complexity of life's rich pageant; it appears before the students in the form of an actual client with an actual case.[98]

Furthermore, these cases keep learning and teaching fresh for clinical professors and students in a way that surpasses even the business school case method.[99] Business schools have long incorporated self-directed field placements into their second-year curricula,[100] but a well-done clinical experience offers a richer, more reflective experience than even a field placement can.[101]

Nevertheless, as long-time observers of legal education know, clinical legal education is relatively new, less respected, and frequently treated as an add-on to the "core" mission of law schools.[102] Legal clinics are often no more than a grudging nod to the work the students will do once they graduate.[103] In this respect, the business school case method is clearly superior; it forms the very core of what business schools are and hope to accomplish.[104]

IV. WHY IS IT THUS?

So why is business school so much more practical than law school? I think it is because of the different legal and professional standings of the two institutions. Anyone who wants to be a lawyer in America must attend law school as a precursor to taking the bar and obtaining a license.[105] Law schools have a captive market of those desiring to enter the profession: No one can choose to skip law school and become a lawyer on her own.[106]

---

97. *See* Paulette J. Williams, *The Divorce Case: Supervisory Teaching and Learning in Clinical Legal Education*, 21 ST. LOUIS U. PUB. L. REV. 331, 334–36 (2002).

98. *See id.*

99. *See id.* at 371–73 (describing the interplay between the students and professors in clinical work).

100. *See* CRUIKSHANK, *supra* note 56, at 349–50 (noting one HBS instructor's use of field studies).

101. This is because the clinic faculty have the time, energy, and training to help the students unpack and reflect upon their experiences. *See, e.g.*, Williams, *supra* note 97, at 372–73.

102. Except at the University of Tennessee College of Law and a handful of other schools. Have I mentioned our Clinic's sixtieth anniversary?

103. *Cf.* Uphoff et al., *supra* note 51, at 381–82 (noting the dissonance between legal education and the practice of law).

104. *See* Atkinson, *supra* note 62, at 979–80 (describing the success and importance of the case method in business schools).

105. *See* Benjamin Hoorn Barton, *Why Do We Regulate Lawyers?: An Economic Analysis of the Justifications for Entry and Conduct Regulation*, 33 ARIZ. ST. L.J. 429, 431, 441–43 (2001).

106. *See id.*

For this reason, law schools can afford to be—and are—inflexible in dealing with the desires of both incoming and current law students.[107] One of the most prevalent examples of this is the relative lack of differentiation or innovation among law schools.[108] Except for a few notable exceptions, like Northeastern's Cooperative Legal Education Program,[109] each American law school is quite similar.[110] For the vast majority of American law schools, first-year curricula are the same, similar candidates are hired as professors, and we professors teach in a style similar to how we learned.[111] In fact, given the massive changes in the country as a whole since the late nineteenth century, it is striking how similar today's law schools are to the Harvard Law School of the 1870s.[112]

This static nature of legal education makes a lot of sense for many of the various players involved. It makes sense for law professors, because innovation is hard and at times painful. In fact, in researching this Essay I came across several references to earlier consideration of using the business school case method in law schools. These were either limited-scope calls for reform[113] or short descriptions of failed efforts.[114] Interestingly, the failed efforts basically floundered because of the perception that it would be a hassle to develop and teach business school-type cases.[115]

---

107. *Cf.* Sturm & Guinier, *supra* note 35, at 520 (discussing how the law school "culture of competition and conformity" is often discouraging to students).

108. *See id.* (noting the law school "culture is remarkably static, non-adaptive, and resistant to change").

109. *See* Northeastern University School of Law, Cooperative Legal Education Program, http://www.slaw.neu.edu/coop/default.htm (last visited Oct. 29, 2007); Northeastern University School of Law, Curriculum, http://www.slaw.neu.edu/course/ (last visited Oct. 29, 2007) (describing Northeastern's unique program for legal education).

110. *See, e.g.*, Sturm & Guinier, *supra* note 35, at 515–20 (describing the American law school as a "tradition-bound institution" with routines and values that foster conformity).

111. *See* Uphoff et al., *supra* note 51, at 397–98 (describing the background of most law professors and its effect on their teaching).

112. *See* Robert W. Gordon, *The Geologic Strata of the Law School Curriculum*, 60 VAND. L. REV. 339, 340 (2007).

113. *See, e.g.*, Douglas L. Leslie, *How Not to Teach Contracts, and Any Other Course: Powerpoint, Laptops, and the CaseFile Method*, 44 ST. LOUIS U. L.J. 1289, 1306–13 (2000) (describing the "CaseFile Method" of teaching contracts); Janet Reno, *Lawyers as Problem-Solvers: Keynote Address to the AALS*, 49 J. LEGAL EDUC. 5, 6–9 (1999) (arguing for the use of "transactional case studies adapted from [the business school] method" in select courses to teach problem-solving to law students); George J. Seidel, *Legal Complexity in Cross-Border Subsidiary Management*, 36 TEX. INT'L L.J. 611, 615 (2001) (using the business school case method to explore legal complexities in international business).

114. *See, e.g.*, Erwin N. Griswold, *Intellect and Spirit*, 81 HARV. L. REV. 292, 303–04 (1967) (describing a failed effort to create business school-type cases for use in legal education).

115. *See id.* (describing the "difficulties" and expense of gathering the cases); *see also* William J. Carney, *Teaching Problems in Corporate Law: Making It Real*, 34 GA. L. REV. 823, 826 (2000) (noting that "few professors engage in teaching problem-solving in transactional subjects . . . [because] it is enormously costly" and "so time-consuming").

The current method of legal education also makes sense for law schools and universities as a whole because the traditional case law method is relatively affordable.[116] A bare-bones law school is inexpensive to staff and maintain; as a result, many universities are able to make a profit on their law schools,[117] unlike many other graduate programs. Furthermore, faculty members from the nation's law schools work with the two licensing bodies for American law schools, the American Bar Association and the Association of American Law Schools, to create detailed educational standards.[118] These standards have the effect of limiting variation and innovation within law schools.[119] This benefits law professors, as well as the law schools themselves, because change would be hard and expensive.

Some practicing lawyers, judges, and bar associations have begun complaining that law schools do too little to educate new lawyers.[120] Yet these complaints are few and far between, partly because some believe that it actually benefits many practicing lawyers for new lawyers to be poorly trained.[121] Since it takes time for a new graduate to become an actual competitive threat to

---

116. *See* George Anastaplo, *Legal Education, Economics, and Law School Governance: Explorations*, 46 S.D. L. REV. 102, 104 (2001).

117. *See id.*

118. *See* SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, preface (2007), *available at* http://www.abanet.org/legaled/standards/20072008StandardsWebContent/Preface.pdf; American Bar Association, Standards Committee Members, http://www.abanet.org/legaled/committees/comstandards.html (last visited Feb. 17, 2008); Association of American Law Schools, AALS Handbook: Membership Requirements, http://www.aals.org/about_handbook_requirements.php (last visited Feb. 17, 2008); Association of American Law Schools, Executive Committee, http://www.aals.org/about_ec.php (last visited Feb. 17, 2008).

119. By comparison, the Association to Advance Collegiate Schools of Business (AACSB) has more flexible standards based upon each school's individual mission. AACSB International, Accreditation, http://www.aacsb.edu/accreditation/ (last visited Oct. 23, 2007); AACSB International, Frequently Asked Questions About the Accreditation Standards, http://www.aacsb.edu/accreditation/business/std_faq5.asp (last visited Oct. 23, 2007). *See generally* Robert H. Jantzen, *AACSB Mission-Linked Standards: Effects on the Accreditation Process*, 75 J. EDUC. FOR BUS. 343 (2000) (providing an overview of the mission-linked accreditation method and the resulting trend toward including teaching-oriented and demographically diverse schools); James Yunker, *Doing Things the Hard Way—Problems With Mission-Linked AACSB Accreditation Standards and Suggestions for Improvement*, 75 J. EDUC. FOR BUS. 348 (2000) (discussing problems resulting from the mission-linked accreditation method).

120. *See generally* SECTION OF LEGAL EDUC. & ADMISSIONS TO THE BAR, AM. BAR ASS'N, REPORT OF THE TASK FORCE ON LAW SCHOOLS AND THE PROFESSION: NARROWING THE GAP (1992), *available at* http://www.abanet.org/legaled/publications/onlinepubs/maccrate.html (describing the shortcomings of current legal education).

121. *Cf.* Benjamin H. Barton, *An Institutional Analysis of Lawyer Regulation: Who Should Control Lawyer Regulation—Courts, Legislatures or the Market?*, 37 GA. L. REV. 1167, 1189–90, 1190 nn.79–80 (2003) [hereinafter Barton, *Institutional Analysis*] (noting the competitive advantage to existing lawyers of raising the standards for entry to the bar).

established lawyers,[122] the longer a new lawyer's start-up period lasts, the better for currently practicing lawyers.[123] Another possible reason for the minimal complaints is that every practicing lawyer came up through the same system; it may be that the current lawyers simply cannot imagine another, better style of legal education.

While law school is mandatory for those who desire to become lawyers, no one is required to go to business school. On the contrary, many business professionals would advise a student against going to business school if avoidable.[124] If the student can get a good job without an MBA, many would advise her to skip the investment of time and money.[125] Moreover, no particular business must hire students with MBAs. Most businesses have the choice to hire economists, college graduates, MBA graduates, or whomever they want. Accordingly, business schools are very responsive to both sides of their customer base: They work hard to make the education worth it to the students, and they work hard to ensure that MBA graduates will be a significant asset to the employers who hire them.

As an aside, I have come to know well some of the fine professors and administrators at the University of Tennessee College of Business Administration over the last few years. In discussing their relatively stronger budgetary situation, several mentioned the business school's very successful and lucrative Center for Executive Education.[126] The Center's motto is "Proven. Results. Faster." and it features various short, but expensive, executive training programs.[127] My friends at the business school wondered whether the law school offers any similar programs. The law school does offer Continuing Legal Education (CLE) programs,[128] but those who attend are basically required to do so by state law. Even with that quasi-advantage, my opinion is that we are not the primary or even one of the better CLE providers in the state.[129] Regardless, no lawyer would pay the rates for law school CLEs that business executives pay for business school classes.[130]

---

122. *Cf.* Barton, *Why Do We Regulate Lawyers?*, *supra* note 105, at 445 (noting the failure of law school and the bar exam to ensure the skill level of entry-level attorneys).

123. *Cf.* Barton, *Institutional Analysis*, *supra* note 121, at 1189–90 (noting the competitive advantage to existing lawyers of raising barriers to entry to the bar).

124. *See, e.g.*, Louise Story, *Bye, Bye B-School*, N.Y. TIMES, Sept. 16, 2007, at Business 1 (describing hedge fund managers who were skipping business school altogether).

125. *See id.* (quoting a Wall Street hedge fund manager as saying, "Going to business school is a way for people to try to open the door, to try to get into a company or hedge fund. But if you're already there, it doesn't make sense to go.").

126. *See* The University of Tennessee College of Business Administration, Center for Executive Education, http://thecenter.utk.edu/ (last visited Oct. 23, 2007).

127. *See id.*

128. The University of Tennessee College of Law, CLE, http://www.law.utk.edu/CLE/CLEhome.htm (last visited Feb. 17, 2008).

129. Three of our professors do teach in excellent CLE programs that are not affiliated with the law school. Professor Penny White is part of the Tennessee Justice Programs CLE. Justice Programs, Justice Programs Faculty, http://www.tennjusticeprograms.com/faculty.htm (last

In thinking about whether the law school could package some kind of program for our alumni, I came to the conclusion that the law school has little to sell when compared to the business school. In fact, the law school probably has a better chance of getting our alumni to pay professors *not* to teach them what we teach the law students. This comparison alone, between the law school's and business school's continuing education programs, gave me pause about what we are teaching and doing in legal education.

MBA programs must justify their existence to both students and employers; this fact alone makes them more innovative and leads to a focus on teaching the skills that people actually need. By comparison, law schools are protected by the licensing requirement and thus are quite hidebound and resistant to change.

CONCLUSION

I teach both Torts I and Torts II here at the College of Law, as well as teaching primarily in the Clinic. I use my semesters teaching the doctrinal courses as semesters to recharge and spend some needed time on my scholarship. When I am teaching Clinic, it is enormously hard to find chunks of uninterrupted time to write. Nevertheless, in writing this Essay I have unwittingly set a personal challenge for myself: If I am convinced that the business school case method is superior, how should I modify my Torts courses? Given the time commitment necessary for a true conversion to the business school case method, however, we shall see whether I am willing to put my money where my mouth is. If the 1L readership of the *Tennessee Law Review* remains similar to earlier years, I can at least rest easy knowing that few of my students will know how far I fall short of my own ideals.

---

visited Oct. 23, 2007). Adjunct professors Don Paine and Sarah Sheppeard are the heart and soul of the Tennessee Law Institute. *See* Tennessee Law Institute, Profiles, http://www.tnlawinstitute.com/faculty.htm (last visited Oct. 23, 2007).

130. Interestingly, there are some legal programs that are expensive and somewhat similar to the Executive Education at business schools. The advocacy classes given by the National Institute for Trial Advocacy (NITA) are one example. *See, e.g.*, National Institute for Trial Advocacy, Programs, http://www.nita.org/programs (last visited Oct. 23, 2007) (listing the categories of NITA's CLE offerings); National Institute for Trial Advocacy, Trial Advocacy, http://www.nita.org/page.asp?id=7&catid=25 (last visited Feb. 17, 2008) (listing CLEs in the Trial Advocacy category and their prices). Once again, though, the difference between the nuts and bolts skills training that NITA offers and the more esoteric education that law schools offer is quite instructive on what lawyers actually want out of continuing legal education.

TEACHING PROFESSIONALISM

BRIDGET MCCORMACK*

I. INTRODUCTION

Complaints about lawyers, by lawyers and non-lawyers alike, sound in professionalism.[1] The critique is so familiar that it is no longer uncomfortable. Among professionals, lawyers are comfortably in last place in the race for respect. Some of this is not deserved; for our system of justice to operate fairly, lawyers must take on an adversarial role, and adversarialism often appears unseemly.[2] But when lawyers play their proper adversarial roles, they ultimately provide loyal counsel and advocacy to their clients; these are the underlying goals of the legal profession, and they are not unseemly at all.[3]

Unfortunately, however, part of lawyers' poor reputation is earned fair and square, and the examples of lawyers acting unprofessionally are too common. Mike Nifong's conviction and disbarment for his unethical and illegal conduct in the recent Duke lacrosse case garnered significant attention from both the media and the government.[4] But the attention directed at Nifong's glaring example of professional misconduct is largely due to the confluence of many uncommon factors, most significantly the case's high-profile status even before allegations against Nifong surfaced and the involvement of wealthy defendants in a position to fight back.[5] Countless examples of lawyers in Nifong's position

---

* Associate Dean for Clinical Affairs and Clinical Professor of Law at the University of Michigan Law School.

1. *See generally* Andrew M. Perlman, *Toward a Unified Theory of Professional Regulation*, 55 FLA. L. REV. 977, 992–1010 (2003) (outlining the history of criticisms of the American bar and the legal profession's longstanding effort to enhance its image).

2. *See* Stephen L. Pepper, *The Lawyer's Amoral Ethical Role: A Defense, A Problem, and Some Possibilities*, 1986 AM. B. FOUND. RES. J. 613, 614 (1986) (explaining the traditional view that when an attorney-client relationship exists, the attorney is expected to favor the client's interests over the interests of others, even if the client's desired goals or means are morally unacceptable).

3. *See generally* Charles Fried, *The Lawyer as Friend: The Moral Foundations of the Lawyer-Client Relation*, 85 YALE L.J. 1060, 1061–62 (1976) (stating that while lawyers generally are not restricted in their choice of clients, once that choice has been made, "the professional ideal requires primary loyalty to the client whatever his need or situation").

4. *See, e.g.*, Titan Barksdale, *Nifong Influenced Criminal Justice Bills*, THE NEWS & OBSERVER, Aug. 4, 2007, *available at* http://www.newsobserver.com/politics/v-print/story/659590.html; Duff Wilson, *At Ethics Hearing, Duke Prosecutor is Called Unprofessional*, N.Y. TIMES, June 15, 2007, at A19.

5. David Feige, *One-off Offing: Why You Won't See a Disbarment Like Mike Nifong's Again*, SLATE, June 18, 2007, http://www.slate.com/id/2168680/ (last visited Apr. 2, 2008).

who commit similar violations never make the newspapers,[6] which proves that the legal profession's system of self-regulation has its costs.

Just as important as the high-profile stories of professional missteps are the unsatisfying experiences that clients regularly have with well-intentioned but overworked lawyers. These stories typically involve unreturned phone calls, a lack of information-sharing, and ultimately a sense of dissatisfaction at the end of the representation. These are the stories that all attorneys hear from their family members and friends, and they are the unfortunate experiences that people, upon being introduced to lawyers, feel compelled to unfold.

Regrettably, current trends in legal education focus little attention on the development of this "everyday professionalism." The traditional law school curriculum focuses primarily on one set of skills lawyers need to succeed, which is comprised exclusively of doctrinal analysis, synthesis, and effective argument.[7] In recent decades, law schools have added other skill sets to their required curricula—legal research and writing and applied problem-solving.[8] Interestingly, law school clinical programs also have grown substantially in number during recent years;[9] the curricula of these programs create a bridge between the more traditional skill set and the newer applied skills curricula, as well as develop additional skills in their participants.

While all ABA-accredited law schools offer a course in professional responsibility that meets ABA requirements,[10] most do not require students to

---

6. *See generally* ANGELA J. DAVIS, ARBITRARY JUSTICE: THE POWER OF THE AMERICAN PROSECUTOR 176-77 (2007) ("For the most part, the media, the electorate, the judiciary, and the legislature have taken a 'hands-off' approach towards the American prosecutor . . . . [I]t is important not only that those who perform this critical function do so in a manner that is legal and fair but also that they perform their duties and responsibilities in accordance with the highest ethical standards.").

7. *See* Anthony G. Amsterdam, *Clinical Legal Education—A 21st- Century Perspective*, 34 J. LEGAL EDUC. 612, 613 (1984) (describing three traditional types of analytic thinking taught in law schools: case reading and interpretation, doctrinal analysis and application, and logical conceptualization and criticism).

8. Over 180 law schools currently have legal research and writing programs. ASS'N OF LEGAL WRITING DIRS./LEGAL WRITING INST., 2007 SURVEY RESULTS, *available at* http://www.alwd.org/surveys/survey_results/2007_Survey_Results.pdf. Many law schools have broader required skills courses. *See, e.g.*, New York University School of Law Lawyering Program, http://www.law.nyu.edu/lawyeringprogram/ (last visited Feb. 20, 2008) (stating that the Lawyering Program complements the first year students' doctrinal classes, giving them "closely structured, collaborative experiences of law in use").

9. *Compare* John S. Bradway, *The Nature of a Legal Aid Clinic*, 3 S. CAL. L. REV. 173, 174 (1930) (listing less than fifteen law schools with functioning legal aid bureaus in 1930), *with* Kimberly McKelvey, Note, *Public Interest Lawyering in the United States and Montana: Past, Present and Future*, 67 MONT. L. REV. 337, 339 (2006) (citing LAW SCHOOL PUBLIC INTEREST LAW SUPPORT PROGRAMS: 1999–2000 DIRECTORY 3, 9–23 (Elissa C. Lichtenstein ed., 1999)) (stating that 172 of the 183 ABA-accredited law schools offered clinical programs in 2006).

10. *See* SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N,

participate in a "professionalism" program.[11] The framework of such a program is difficult to envision given the broad subject matter covered by the term *professionalism*, but law schools can begin to bear more responsibility in this area and thereby dramatically improve their current methods of teaching professionalism by making one simple change: requiring clinical education.

Legal educators should be required to answer difficult questions concerning the steps they currently are taking to teach professionalism. While views of what constitutes professionalism differ,[12] for my purposes, professionalism includes acting in accordance with the skills of a competent lawyer and with the values essential to lawyering. The skills required for professionalism include the broad set of skills needed to practice law competently, such as doctrinal analysis, synthesis, engagement with counter-argument, and problem-solving, and the more concrete lawyering skills, including legal research and writing, interviewing, counseling, negotiation, and trial skills.[13] Included in the values required for professionalism are those that lawyers would generally agree that they share: loyalty, integrity, confidentiality, competence, diligence, excellence, and public service, among others.[14]

STANDARDS FOR APPROVAL OF LAW SCHOOLS, Standard 302(a)(5) (2007), *available at* http://www.abanet.org/legaled/standards/20072008StandardsWebContent/Chapter%203.pdf (requiring "substantial instruction in . . . the history, goals, structure, values, rules and responsibilities of the legal profession and its members").

11. Some law schools require a clinical education, which would fulfill this requirement. *See, e.g.*, The City University of New York School of Law, Academic Programs and Resources, http://www.law.cuny.edu/cns/law/live/academics/curriculum.html (last visited Feb. 20, 2008); The University of New Mexico School of Law, Clinical Law Programs, http://lawschool.unm.edu/clinic/index.php (last visited Feb. 20, 2008).

12. *See, e.g.*, MONROE H. FREEDMAN & ABBE SMITH, UNDERSTANDING LAWYERS' ETHICS 6–9 (3d ed. 2004) (surveying various propositions for the distinctive features of ethics in the legal profession and positing that lawyers' ethics are rooted in the moral values expressed in the Bill of Rights); *id.* at 71–127 (exploring the professional value of zealous representation); Fried, *supra* note 3, at 1080 ("I have defined the lawyer as a client's legal friend, as the person whose role it is to insure the client's autonomy within the law."); Pepper, *supra* note 2, at 614 (stating that the generally accepted view of a lawyer's proper role "'is to prefer . . . the interests of client or patient' over those of other individuals. '[W]here the attorney-client relationship exists, it is often appropriate any many times even obligatory for the attorney to do things that . . . an ordinary person need not, and should not do.'" (quoting Richard Wasserstrom, *Lawyers as Professionals: Some Moral Issues*, 5 HUM. RTS. 1, 5 (1975))); Perlman, *supra* note 1, at 979 ("The dominant view posits (roughly) that attorneys should pursue all lawful strategies in order to achieve clients' objectives, even if those strategies produce immoral or unjust results in particular cases.").

13. *See* Roy T. Stuckey, *Education for the Practice of Law: The Times They Are A-Changin'*, 75 NEB. L. REV. 648, 654 (1996) (stating that a 1979 ABA task force recommended that law schools "should provide instruction in those fundamental skills critical to lawyer competence" and listed those skills as "being able to analyze legal problems and do legal research . . . write, communicate orally, gather facts, interview, counsel, and negotiate").

14. *See* Roy Stuckey, *Teaching with Purpose: Defining and Achieving Desired Outcomes in Clinical Law Courses*, 13 CLINICAL L. REV. 807, 820 (2007) ("Among the values that we

The bulk of teaching professionalism in law schools occurs through clinical courses. Yet the second-class role of clinical programs in the law school curriculum undermines the importance of the pedagogy of professionalism.[15] While I agree with the more general arguments for requiring clinical education, the observation that clinical pedagogy places more emphasis on professionalism than any other part of the law school curriculum is its own justification for requiring clinical education alongside traditional courses like constitutional law, property, or contracts.

II. THE CURRENT LAW SCHOOL CURRICULUM AND PROFESSIONALISM

In the 1960s, Phillip Jackson's important work on elementary education identified two primary curricula through which lessons are taught and learned.[16] The official, or explicit, curriculum encompasses that which is expressly stated;[17] teachers specifically tell their students that this curriculum is the subject of their education. The hidden, or implicit, curriculum describes the lessons that students learn in the classroom aside from the expressly labeled school subjects.[18] This curriculum includes "the crowds, the praise, and the power that combine to give a distinctive flavor to classroom life . . . which each student (and teacher) must master if he is to make his way satisfactorily through the school."[19] Unlike the explicit curriculum, which aims at teaching subjects such as reading, writing, and arithmetic, the implicit curriculum focuses on the important lessons students learn from structure, process, and social interactions.[20] Elliot Eisner has also recognized a third type of curriculum.[21] The null curriculum encompasses that which is not taught at all,[22] and through

---

should include in our instructional design are the lawyer's obligations to truth, honesty, and fair dealing; the responsibility to improve the integrity of the legal system within which the lawyer exercises the skills that are taught; the obligation to promote justice; and the obligation to provide competent representation."). *See generally* Robert MacCrate, *Yesterday, Today and Tomorrow: Building the Continuum of Legal Education and Professional Development*, 10 CLINICAL L. REV. 805, 807–11 (2004) (discussing the bar's various canons of ethical rules and values in the twentieth century).

15. *See* Kenneth R. Kreiling, *Clinical Education and Lawyer Competency: The Process of Learning to Learn From Experience Through Properly Structured Clinical Supervision*, 40 MD. L. REV. 284, 316–18 (1981) (discussing the "second-class status" of clinical teachers and its effect on clinical education).

16. *See* PHILLIP W. JACKSON, LIFE IN CLASSROOMS 33–37 (1968).

17. *See id.* at 34.

18. *See id.*

19. *Id.* at 33–34.

20. *See id.* at 34–35.

21. *See* ELLIOT W. EISNER, THE EDUCATIONAL IMAGINATION: ON THE DESIGN AND EVALUATION OF SCHOOL PROGRAMS 97–107 (3rd ed. 1994).

22. *Id.* at 97 ("[I]f we are concerned with the consequences of school programs and the role of curriculum in shaping those consequences, then . . . we are well advised to consider not only the explicit and implicit curricula of schools but also what schools do not teach. It is my

this "curriculum," students make important judgments about which values are important and which are not.[23] For better or for worse, law schools use each of these three curricula to teach lessons about professionalism.

A. The Explicit Law School Curriculum and Professionalism

Law schools maintain an explicit curriculum in professionalism by formally teaching the subject through doctrinal courses. However, in addition to coursework in professional responsibility, law schools offer students many other opportunities for formally learning the norms of the profession. Students also learn the explicit curriculum more generally through expectations, experiences, and evaluation.[24]

1. Expectations

Legal educators communicate professional expectations to students through orientation sessions, mission statements, policies and procedures, and codes of conduct. Professionalism is highlighted in all of these fora. Students must conform to these expectations if they wish to successfully complete their coursework, and these obligations instill professionalism.

New medical students traditionally end their first year of medical school by participating in what is commonly known as the "white coat" ceremony.[25] During this ceremony the medical students receive their first white coat—a symbol of their profession—and are asked to take the Hippocratic Oath[26] for

---

thesis that what schools do not teach may be as important as what they do teach.").

23. *See id.* at 97–107.

24. *Cf.* David Stern, *Outside the Classroom: Teaching and Evaluating Future Physicians*, 20 GA. ST. U. L. REV. 877, 896 (2004) (discussing teaching medical students professionalism through clinical training).

25. Carol A. Heimer, *Responsibility in Health Care: Spanning the Boundary Between Law and Medicine*, 41 WAKE FOREST L. REV. 465, 495 n.114 (2006).

26. A Modern Hippocratic Oath by Dr. Louis Lasagna states:

> I swear to fulfill, to the best of my ability and judgment, this covenant:
>
> I will respect the hard-won scientific gains of those physicians in whose steps I walk, and gladly share such knowledge as is mine with those who are to follow;
>
> I will apply, for the benefit of the sick, all measures which are required, avoiding those twin traps of overtreatment and therapeutic nihilism.
>
> I will remember that there is art to medicine as well as science, and that warmth, sympathy, and understanding may outweigh the surgeon's knife or the chemist's drug.
>
> I will not be ashamed to say "I know not," nor will I fail to call in my colleagues when the skills of another are needed for a patient's recovery.
>
> I will respect the privacy of my patients, for their problems are not disclosed to me that the world may know. Most especially must I tread with care in matters of life and death. If it is given me to save a life, all thanks. But it may also be within my power to take a life; this awesome responsibility must be faced with great humbleness and awareness of my own frailty. Above all, I must not play at God.

the first time. Law schools have begun experimenting with a tradition that mirrors this one. At the University of Michigan Law School, for example, the incoming first year law students take part in a "Commitment to Integrity" ceremony. At this ceremony a federal judge addresses the incoming class concerning integrity as a core value of the profession and then administers the integrity oath.[27] This type of ceremony sets a tone of professionalism by impressing upon students that they are part of a larger profession with shared values.

2. Experiences

Students learn a great deal about professionalism through their educational experiences in and out of the classroom. In the traditional classroom, law faculty set a professional tone and expect their students to respect and maintain the professional atmosphere. Faculty commonly set forth specific classroom expectations through syllabi or oral communication to class participants. To emphasize professionalism, law school faculty regularly refer to students by their last names. Classroom dialogue generally follows the norms of the Socratic method, which emphasizes thorough preparation, presentation, and accountability and mirrors the interactions that students can expect in a professional setting.[28] Each of these experiences in the classroom teaches professionalism while the students are also learning traditional doctrine.

> I will remember that I do not treat a fever chart, a cancerous growth, but a sick human being, whose illness may affect the person's family and economic stability. My responsibility includes these related problems, if I am to care adequately for the sick.
>
> I will prevent disease whenever I can, for prevention is preferable to cure.
>
> I will remember that I remain a member of society, with special obligations to all my fellow human beings, those sound of mind and body, as well as the infirm.
>
> If I do not violate this oath, may I enjoy life and art, respected while I live and remembered with affection thereafter. May I always act so as to preserve the finest traditions of my calling and may I long experience the joy of healing those who seek my help.

Association of American Physicians and Surgeons, Inc., http://www.aapsonline.org/ethics/oaths.htm (last visited Feb. 20, 2008).

27. The text of the integrity oath was as follows:

> Because the strength of the legal profession depends on the character of its members, during my career as a law student, I commit to comport myself honorably and with integrity. Specifically, I promise to maintain high standards of: Academic conduct in all academic relationships with the Law School and the University; Professional Conduct while functioning in a lawyer-like capacity at any time during my Law School career; and Personal conduct in all matters that touch or affect the Law School, the University, and their community members and guests.

Committing to Integrity and Professionalism, LAW QUADRANGLE NOTES (Univ. of Mich. Law School, Ann Arbor, Mich.), Fall 2007, at 55, *available at* http://www.law.umich.edu/newsandinfo/lqn/pasteditions/fall2007/ Documents/fall2007.pdf.

28. *See* James H. Backman, *Practical Examples for Establishing an Externship Program*

Students' experiences outside the traditional law school classroom are an even richer tool for teaching professionalism. When students participate in a clinical course, professionalism is at the heart of their learning.[29] This remains true whether the students are involved in a live-client course, in which they often learn by fire,[30] or a simulation course that employs more traditional clinical pedagogy.[31]

In a clinical setting, students are forced to ascertain the applicable professional expectations and norms as they muddle through solving a client's problems.[32] When a client enters a clinic student's office, the student must immediately respond to the client. When opposing counsel calls, the student must pick up the phone and engage in a conversation. When the judge asks the student a difficult question that requires upholding the duty of confidentiality while still satisfying the judge, the student is forced to reconcile these competing interests consistent with the norms of the profession. Through the demands of a clinical setting and their interactions with clients, judges, and other lawyers, students learn a great deal about the skills and values of their profession.

3. Evaluation

Law schools are largely unsuccessful at evaluating professionalism through the traditional curriculum. The letter grades students are assigned at the end of their contracts course, for example, might measure how well they answered the questions on a particular exam and might even measure such important skills as doctrinal synthesis and issue-spotting. But mastery of a limited skill set in a law school contracts course does not indicate a student's competence in professionalism more generally.

Clinical programs are designed not only to teach professionalism, but to evaluate a student's development in this area; this is the distinct value of clinical education. Evaluation is accomplished through supervision sessions in clinical and simulation-based courses,[33] as well as through feedback from peers,

---

Available to Every Student, 14 CLINICAL L. REV. 1, 21 (2007).

29. *See* Stuckey, *supra* note 14, at 828 ("Students' observations and experiences in all types of externships can provide rich fodder for discussing and reflecting on professionalism issues . . . .").

30. *See generally id.* at 830–36 (discussing the unique professionalism lessons of client representation courses).

31. *See generally id.* at 824–28 (describing how simulation-based courses allow students to replicate and work through professionalism issues).

32. *See* JoNel Newman, *Re-conceptualizing Poverty Law Clinical Curriculum and Legal Services Practice: The Need for Generalists*, 34 FORDHAM URB. L.J. 1303, 1322–23 (2007).

33. *See generally* Kreiling, *supra* note 15, at 318–19 (outlining a "supervision cycle" for clinical education comprised of "(1) initial conference, (2) preperformance conference, (3) observations, (4) preconference analysis and strategy, (5) post-performance conference, and (6) final evaluation and termination").

clients, and faculty.[34] Given the rich material with which students are working in a clinical setting, meetings between the students and their supervising faculty are never short of opportunities to review and reflect upon professional skills and values.[35] Clinical programs offer students valuable interactions with clients, judges, and other lawyers and therefore allow the students to be actively engaged in learning professionalism.[36]

Clinical faculty are in an excellent position to discover and address professionalism shortcomings in students.[37] They can personally monitor these issues in a way that faculty teaching only doctrinal courses simply cannot.[38] Most of these professionalism shortcomings can be easily addressed with feedback and continued monitoring. If more serious issues arise with a particular student, clinical faculty are well-poised to track, record, and intervene in these problems as necessary.[39]

Medical researchers who have tracked doctors who develop professionalism problems have found that the clinical setting is the only educational experience that is useful in predicting which students will later develop such problems in their practices.[40] Medical students who had difficulties with professionalism during their clinical training were the most likely to have similar problems in their post-graduation practice.[41] Likewise,

---

34. *See* Stuckey, *supra* note 14, at 829 ("'Feedback from more accomplished performers directs the learner's attention, supporting improved attempts at a goal.'" (quoting WILLIAM M. SULLIVAN, ANNE COLBY, JUDITH WEGNER, LLOYD BOND & LEE S. SHULMAN, EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW 8 (2007))).

35. *See generally* Kreiling, *supra* note 15, at 330–36 (discussing the structure and benefits of post-performance conferences between clinical students and their supervisors).

36. *See* Stuckey, *supra* note 14, at 828–29.

37. *See* Kreiling, *supra* note 15, at 331–32.

38. *See* David Luban & Michael Millemann, *Good Judgment: Ethics Teaching in Dark Times*, 9 GEO. J. LEGAL ETHICS 31, 69 (1995) ("[C]linical ethics teaching can reveal points where the moral shoe pinches that typically go undiscovered and undiscussed in the pure classroom . . . .").

39. *See* Kreiling, *supra* note 15, at 331–32.

40. David T. Stern et al., *The Prediction of Professional Behavior*, 39 J. MED. EDUC. 75, 80 (2005) (concluding that significant predictors of professionalism could only be found "in domains where students had had opportunities to demonstrate conscientious behaviour or humility in self-assessment"). *See generally* Steven H. Miles et al., *Medical Ethics Education: Coming of Age*, 64 ACAD. MED. 705, 708 (1989) ("There is wide agreement that ethics instruction should be case-centered, especially during the clinical phase of education. Case discussion serves many of the objectives of ethics education; it teaches sensitivity to the moral aspects of medicine, illustrates the application of humanistic or legal concepts to medical practice, and shows physicians acting as responsible moral agents."); *id.* at 710 ("Clinical exposure to patients in the first and second years assists the development of a patient-centered rather than a disease-centered approach to patients. Such early clinical experiences can foster ethical sensitivity, help the student examine values she or he brings to clinical care, teach clinical-ethical reasoning, and foster effective collaboration with nurses, lawyers, or other professionals.").

41. Stern, *supra* note 40, at 75.

law schools would gain a greater understanding of their students' professional competence if all law students were evaluated in a clinical setting. Without the opportunity to evaluate students as they perform in a professional setting, law schools will continue to inadequately assess and address professionalism shortcomings.[42]

B. The Implicit Law School Curriculum and Professionalism

The implicit, or hidden, curriculum is an important tool for teaching professionalism in law schools. Indeed, many of the values and some of the skills of professionalism are taught largely in the hidden curriculum.[43] The traditional law school classroom's implicit curriculum is full of lessons about professional values such as preparation, diligence, engagement in counter-argument, and thoroughness.[44] Law faculty generally do not explicitly tell their students that thorough preparation will make their time in the classroom far more valuable and less stressful, but it does not take the students long to appreciate this fact. Students also learn how to switch sides in an argument, predict the follow-up question, and play the devil's advocate, although they are not formally taught these skills.[45] Students learn these skills through their professors' modeling and the praise they receive for superior performance.[46]

Clinical education adds a valuable dimension to the professionalism that is taught in the hidden law school curriculum generally. The nature of clinical teaching makes clinical coursework a rich learning ground for the implicit professionalism curriculum.[47] Clinic students simply have far greater access to

---

42. *Cf.* David Thomas Stern, *A Framework for Measuring Professionalism*, *in* MEASURING MEDICAL PROFESSIONALISM 3,5 (David Thomas Stern ed., 2006) (discussing the current lack of an adequate set of tools with which to measure professionalism in the field of medical education and noting that "[t]he ability to accurately measure professionalism will allow us to detect and dismiss those students or physicians with extremes of deviant behavior. Measuring professionalism will allow us to provide formative feedback to physicans across the educational continuum."). *See generally* Louise Arnold & David Thomas Stern, *What is Medical Professionalism?*, *in* MEASURING MEDICAL PROFESSIONALISM, *supra*, at 15 (explaining the different definitions of professionalism posited by leaders in the medical field and calling for an explicit, consistent definition in order to adequately assess professionalism in students).

43. *See* Stern, *supra* note 24, at 896 (discussing how medical students learn professionalism and stating that "[w]hile students obtain much of the knowledge and the skills of medicine during the formal time of the curriculum, members of the faculty predominately teach the professional behaviors (or lack thereof) through the informal, or 'hidden' curriculum" (citing Fredrick W. Hafferty & Ronald Franks, *The Hidden Curriculum: Ethics Teaching and the Structure of Medical Education*, 69 ACAD. MED. 861, 865 (1994))).

44. *See generally* Vincent Blasi, *Teaching Reasoning*, 74 CHI.-KENT L. REV. 647 (1999) (describing his approach to teaching legal analysis and argumentation through seminar courses and noting the lessons that students learn from his methods).

45. *See id.* at 647–48.

46. *See id.*

47. *See generally* Nathaniel C. Nichols, *Modeling Professionalism: The Process from a*

faculty and other legal professionals, thereby enhancing their opportunities to "bump up" against the hidden curriculum.[48]

In clinical courses, students ride with faculty members as they drive to court or to meet with clients.[49] They are present in courthouse hallways when other lawyers or judges pull their supervising attorneys aside for conversation, and sometimes the students themselves are involved in those conversations. The students regularly remain in recessed courtrooms to experience the interactions among professionals and between lawyers and their clients, and occasionally they are invited into judges' chambers. The students work in the clinical law office alongside the professional paralegal staff who operate within their own set of norms,[50] and they often spend late nights working on their clients' cases with more seasoned clinical students, listening to and learning from their experiences. In all of these settings, the professionalism curriculum, though not explicit, is front and center.

The clinical setting emphasizes professionalism not only through students' increased exposure to the interactions among other professionals, but also through the intimacy that characterizes their relationships with clinical faculty. Students' relationships with clinical faculty are often significantly different than those with doctrinal faculty.[51] Whether I like it or not, my clinical students spend many hours in my office. They see pictures of my family and my kids' schoolwork, and they are often present for family interruptions. In short, they simply have more personal access to me than to their other professors. This level of intimacy provides a useful platform for students to ask difficult questions, evaluate their own missteps comfortably, and ask me about my own choices. Consequently, their understanding of professionalism is expanded

---

Clinical Perspective, 14 WIDENER L.J. 441 (2005) (describing how the clinical setting is ideal for "affective learning," which requires students to internalize the values of the profession). Professor Nichols describes how the clinic is a "model ethical law office," teaching students professionalism through office procedures, case rounds, court observations, and individual case experiences. *Id.* at 444–48.

48. *See, e.g.*, Lisa Torraco, *The New Mexico District Attorney Clinic: Skills and Justice*, 74 MISS. L.J. 1107, 1125 (2005) (stating that faculty supervisors in the University of New Mexico's District Attorney Clinic are "responsible for the competency of the student prosecutors and their compliance with professional obligations").

49. Or they ride the subway together. When I was a clinic student in New York City, the ride with my clinical supervisor on the A train from Washington Square to the courthouse in Brooklyn was an extremely important part of my own professionalism training. For my students in Ann Arbor, my mini-van provides the setting for important conversations about professionalism as we drive to and from court.

50. *Cf.* Nichols, *supra* note 47, at 445 (noting that by receiving an office procedures manual and by setting up a "dummy file based upon a mock client interview," clinical students develop "an awareness of the importance of the office procedures and a willingness to devote controlled attention to the procedures").

51. *See* Kathleen A. Sullivan, *Self-Disclosure, Separation, and Students: Intimacy in the Clinical Relationship*, 27 IND. L. REV. 115, 119 (1993) ("Most clinical teachers interact with their students more frequently and work with them more closely than traditional law teachers.").

through our interactions in ways that cannot occur in a traditional law school classroom.[52]

No textbook can teach the critical lessons learned in these settings, but it is primarily through the implicit curriculum that students internalize important professional values such as confidentiality, loyalty, and the value of public service. The opportunities for teaching these values in the hidden curriculum are enhanced in a clinical setting, where role-modeling, parables of past experiences, and intimate relationships with seasoned faculty supervisors are readily accessible.

C. The Null Law School Curriculum and Professionalism

The null curriculum is equally as important in the development of professionalism in law schools as both the explicit and the implicit curricula. This is true because professionalism is not taught uniformly in law schools; students receive varying lessons in professionalism depending upon the courses they choose to take. Whether or not they consciously recognize it, students internalize important lessons from what their law schools choose not to teach.

At most law schools, some portion of the student body chooses not to participate in a clinical course.[53] At the University of Michigan Law School that percentage hovers somewhere around 50%.[54] The students who choose to enroll in a clinic receive an excellent education in professionalism through both the explicit and implicit curricula. However, for the other half who choose not to participate, their primary lesson about professionalism is that the law school does not value the professionalism training that a clinical experience provides.

III. THE FUTURE LAW SCHOOL PROFESSIONALISM CURRICULUM

Law schools could teach professionalism more effectively by mandating participation in clinical courses. If all students spent some part of their legal education in clinical practice, law schools would provide more uniformity in the teaching of professionalism. Additionally, such a requirement would allow

---

52. *See generally id.* at 117 (discussing how clinical teaching is more intimate than traditional teaching and how the often complicated interactions between clinical teachers and their students can ultimately lead to powerful educational experiences).

53. *See* Neal Kumar Katyal, Comment, *Hamdan v. Rumsfeld: The Legal Academy Goes to Practice*, 120 HARV. L. REV. 65, 68–69 (2006) ("The truth is that very few law schools today prepare students to be lawyers: this responsibility is shunted off to law firms, the judges for whom students clerk, prosecutors' offices, and others. The obvious exception is law clinics, which do offer crucial lessons in the art of good lawyering. But clinics, despite their many virtues, still do not reach most law students, and their connection to the theoretical law taught elsewhere in the school is often left murky.").

54. E-mail from David Baum, Assistant Dean and Senior Manager of Student Affairs, Univ. of Mich. Law School, to author (Nov. 13, 2007, 13:11:30 EST) (on file with the Tennessee Law Review).

faculty to identify those students for whom developing the skills and values of the profession is difficult and would position them to intervene and take the necessary steps to remedy those problems.[55]

If law schools' unwillingness to expand clinical offerings was the only obstacle to providing more opportunities to teach professionalism, such change would likely occur within a generation. But the willingness to initiate these programs is not the major hurdle; it is the expense of such programs that makes their adoption difficult. Traditional, live-client clinics with student-faculty ratios of eight to one are costly. If clinical faculty can produce only eight widgets a term while contracts teachers can produce 100, mandatory clinical education will necessitate greater budgets regardless of any widespread agreement concerning the value of such curricular reform.

Fortunately, clinical pedagogy is available in contexts beyond the traditional live-client clinic. Many law teachers use clinical methodology in non-clinical courses; this is true both in courses which focus on lawyering skills and in courses in which other doctrine is the main focus.[56] In fact, clinical methodology can be incorporated into almost any law school course, and while it is less practical in the courses with the largest enrollments, the methodology can be engaging and can add depth to even the largest lecture class. If professionalism can be effectively taught in contexts other than a live-client clinic, we must insist that this be done.

For example, I am currently teaching criminal law for first-year students. At the end of each unit, I require the students to complete an assignment relating to the material, which forces them to apply their knowledge to an actual case. Upon completion of the homicide unit, for example, I provided the students with a four-page incident report involving a woman who shot and killed her husband after an altercation in a nearby county. I asked the students to write a memorandum to their "bureau chief" advising her of the most appropriate charge for the woman. In class, three students, each with different recommendations, took turns making their recommendations to their supervisor. The discussion which followed focused as much on the ethics of overcharging criminal defendants as it did on the substantive law. Each of the other end-of-unit exercises has been similarly followed by class discussions which explore the values of the profession as much as the substantive material assigned.[57]

---

55. *See* Kreiling, *supra* note 15, at 331–32.

56. For example, a clinical colleague at the University of Michigan Law School successfully incorporated clinical methodology into a domestic violence course. She hired an actor to play the role of the students' "client" as they "litigated" her civil and criminal cases, while learning the substantive domestic violence law. The "client" provided feedback to the students, which was invaluable in teaching professionalism.

57. The students have reported enjoying these assignments, even though they require additional work. Each of the exercises requires the students to take on a different lawyering role, and they must complete the assignment in their role instead of in the abstract. Role-playing, more than an abstract debate, advances the professionalism discussion.

In large lecture courses during the first year of law school and beyond, using clinical methodology exclusively is probably impractical, but importing pieces of such methodology into these courses can expand the extent to which they reach the skills and values of professionalism.[58] Because the traditional first-year classroom plays an important role in students' legal education, especially in instilling professionalism, clinical methodology should be employed early in their law school careers. The substantive lessons taught in doctrinal courses will not be undercut by expanding the teaching in this direction, and experimenting with clinical pedagogy will only add dimension to those courses.

By their second year, law students have learned the skill set taught in large lecture courses, if not the doctrine that is the subject of each course. They can respond to and find weaknesses on both sides of most any argument. They have also absorbed important professionalism lessons conveyed in the first-year curriculum, such as diligence, preparedness, and thoroughness. The teaching of professionalism, however, should not end after the first year; rather, it should be expanded in upper-level courses to include additional skills and values of the profession. For example, criminal procedure certainly can be taught by lecture in a large, Socratic-style course. But it could also be taught exclusively in a simulation-based course in which the students are assigned fictional clients and are required to make decisions based on their knowledge of both the applicable legal doctrine and their professional obligations in their respective roles.[59]

The teaching of professionalism can also be expanded by looking outside the realm of traditional doctrinal courses to capitalize on the professional experiences students have while still enrolled in law school. For example, a fall semester course offered to students upon their return from summer legal employment could be designed to explore professionalism norms and would be an easy addition to the upper-level curriculum. Students often return from these experiences in law practice thinking about issues of professionalism, and law schools should utilize these experiences to explicitly teach professionalism in this context.

---

58. For example, another colleague who teaches civil procedure devotes one day each semester to professional training. The professor invites a local federal court judge to conduct his or her actual motion docket at the law school. The pleadings related to the motions to be argued are posted on the course website to allow the students to read them in advance. The discussion in the class meeting preceding the docket day generally focuses exclusively on the legal issues raised in the pleadings. However, the class discussion following the docket day is almost wholly about the lawyers, their decisions, their mistakes, and their successes.

59. In fact, New York University School of Law has offered such a course, entitled "Criminal Litigation," for almost two decades. *See* New York University School of Law, Course Management System, http://its.law.nyu.edu/StudentCourseInfo.cfm (follow "Course Title" hyperlink; then follow "Criminal Litigation" hyperlink) (last visited Feb. 20, 2008). NYU also offers a similar evidence course entitled "Evidence and Professional Responsibility." New York University School of Law, Course Management System, http://its.law.nyu.edu/StudentCourseInfo.cfm (follow "Course Title" hyperlink; then follow "Evidence" hyperlink) (last visited Feb. 20, 2008).

Finally, as law curricula evolve and teaching professional skills and values gains heightened importance in law schools, clinical teachers should prepare themselves to provide clinical experiences to students in new and expansive ways. If the traditional in-house clinic is simply too expensive to provide live-client experiences to all students, clinical teachers must consider alternatives that will allow all students to at least gain the same benefits of traditional clinical programs. The ideas for countering this problem likely will be different at each school, depending on school-specific factors. Identifying qualified lawyers in the community who are willing to supervise students can be one part of the solution. Partnerships with thoughtful practitioners can greatly benefit law students as they learn professionalism, whether they are working in traditional externships and processing their experiences with full-time clinical faculty members, or working with practitioners in the field as add-on credits to their substantive coursework.[60]

Clinical faculty should be at the forefront of these advances in legal education. Law school clinics have evolved tremendously in the last thirty years.[61] They will (and must) continue to do so. Discovering new avenues for providing clinical experiences to law students benefits those students, their law schools, and the entire legal profession.

IV. CONCLUSION

Current law school curricula fails to teach professionalism consistently to all students. However, many opportunities for growth and improvement are within reach. Adding clinical methodology to courses where it has traditionally been absent and expanding clinical opportunities for students during law school would complement the teaching of professionalism, thereby promoting professionalism among lawyers. Clinical faculty should shepherd this evolution.

60. *See generally* Stuckey, *supra* note 14, at 812 (noting that in externships, students represent clients under the supervision of practicing lawyers, or they observe lawyers or judges in their work).

61. *See generally* Nina W. Tarr, *Current Issues in Clinical Legal Education*, 37 HOW. L.J. 31 (1993) (examining the many contributions made by clinical education since its inception in the 1960s and 1970s).

LEGAL ADVOCACY AND EDUCATION REFORM: LITIGATING SCHOOL EXCLUSION

DEAN HILL RIVKIN*

I. INTRODUCTION

Public education has become a crucible for fundamental debates about the nature of American democracy. This is especially true with issues surrounding exclusion of children from school. Excluding students from our "open" public school systems has sparked a robust discourse about the core purposes of public education. Litigation over exclusion highlights the critical importance of education to our children and our nation.

In *Plyler v. Doe*,[1] the United States Supreme Court invalidated a Texas law that withheld state funds for the education of children who were not "legally admitted" into the United States.[2] Justice Brennan, writing for a 5-4 majority, emphasized the importance of educating this "underclass" of children:

> Public education is not a "right" granted to individuals by the Constitution. But neither is it merely some governmental "benefit" indistinguishable from other forms of social welfare legislation. Both the importance of education in maintaining our basic institutions, and the lasting impact of its deprivation on the life of the child, mark the distinction. The "American people have always regarded education and [the] acquisition of knowledge as matters of supreme importance." . . .
>
> . . . Paradoxically, by depriving the children of any disfavored group of an education, we foreclose the means by which that group might raise the level of esteem in which it is held by the majority. . . . Illiteracy is an enduring disability. The inability to read and write will handicap the individual deprived of a basic education each and every day of his life. The inestimable toll of that deprivation on the social, economic, intellectual, and

---

* College of Law Distinguished Professor, University of Tennessee College of Law. A.B. Hamilton College (1968); J.D. Vanderbilt Law School (1971). This Article is dedicated to attorney Brenda McGee, my spouse. She single-handedly educated me about zealous education advocacy. I also hugely benefited from the insights of attorney Barbara Dyer, the staff attorney for the University of Tennessee College of Law's Children's Advocacy Network-Lawyers Education Advocacy Project (CAN-LEARN). CAN-LEARN, which I direct, is a support project for lawyers in Tennessee who represent families and children in education-related cases. *See* www.lawschoolconsortium.net. My research assistant, Madeline McNeeley, contributed greatly to the research and editing of this Article. Many of the practices and stories recounted in this Article stem from countless conversations with families and lawyers about education issues. I have litigated two of the cases discussed in the Article and many more in this field. I take full responsibility for the claims made throughout.

1. 457 U.S. 202 (1982).
2. *Id.* at 224–25.

> psychological well-being of the individual, and the obstacle it poses to individual achievement, make it most difficult to reconcile the cost or the principle of a status-based denial of a basic education with the framework of equality embodied in the Equal Protection Clause.[3]

The *Plyler* Court rightly rejected the State's claim that undocumented children were not "persons" under the Constitution. The State's argument literally objectified the children excluded by the Texas law.[4]

In *Honig v. Doe*,[5] the Court confronted a special education exclusion case involving two emotionally disturbed youths who had engaged in "disruptive behavior," including stealing, extorting money from fellow students, making sexual comments to female classmates, and kicking out a glass window.[6] Writing again for a 5-4 majority, Justice Brennan interpreted the "stay-put" provision of the federal Individuals with Disabilities Education Act (IDEA). The Court's decision prevented the San Francisco Unified School District from expelling these students for their disability-fueled behavior.[7] The Court rejected the school system's argument that Congress could not have intended to require schools to retain "violent or dangerous" students in school while they contested their expulsions through the often ponderous administrative machinery of the IDEA. The majority scolded the school system by underscoring "that Congress very much meant to strip schools of the *unilateral* authority they had traditionally employed to exclude disabled students, particularly emotionally disturbed students, from school."[8] Reading like an education primer, the opinion catalogued various methods that schools could use to educate students "who are endangering themselves or others."[9] The decision conveyed the message that continuing education—even for the most difficult students—trumped the ossified discipline practices of certain school administrators.

Despite the import of these cases, educational institutions continue to devise mechanisms for removing students from schools, which has sounded the

---

3. *Id.* at 221–22 (citations omitted).

4. *Id.* at 210. Authorities often objectify children and youths who are excluded from school for behavioral reasons. Their narratives portray these students as disruptive predators or out-of-control troublemakers, rather than persons whose developmental problems need to be understood and accounted for. A step in the right direction is the requirement of a functional behavioral assessment, followed by the development of a behavior implementation plan under the Individuals with Disabilities Education Act (IDEA). 20 U.S.C. § 1415(k)(1)(D)(ii) (Supp. 2004). Advocates today are beginning to frame school exclusion as a human rights issue. *See* Statement of Dignity in Schs. Campaign, A Project of the Educ. Subcomm. of the Am. Bar Ass'n. Children's Rights Litig. Comm., http://www.abanet.org/litigation/committees/childrights/docs/dsc_statement.pdf (last visited Jan. 18, 2008).

5. 484 U.S. 305 (1988).

6. *Id.* at 312–15.

7. *Id.* at 316–17.

8. *Id.* at 323.

9. *Id.* at 325–26.

death knell for many students' academic careers. As will be discussed in Part II of this Article, many of these mechanisms are not transparent. They play on parents' lack of sophistication about their child's education. Others invoke higher norms—like school safety—to justify exclusion. Still others impose penalties on non-conformist behaviors simply because some students' unique personalities are poorly understood by school administrators. These systems of discipline are riddled with unfair rules, procedures, and practices.

Part III will discuss the evolving legal landscape of school exclusion. It begins by exploring the mixed motivations behind school exclusion. This Part will analyze a sample of the growing number of cases that seek to turn "failure in the classroom into success in the courtroom,"[10] and it will explicate the pros and cons of using litigation to prevent school exclusion.

The Conclusion of the Article will evaluate the suitability of law school legal clinics and other public interest law firms for school exclusion work. Education as a whole is under-represented as a substantive area for legal clinics and other nonprofit firms.[11] These firms have not embraced this work for a variety of pedagogical and political reasons, but the time has come to rethink this approach. If undertaken, attorneys must pursue these cases within a framework of systemic, long-term reform. The task presents a formidable challenge.

II. PATHWAYS TO SCHOOL EXCLUSION

Historically, schools have used a number of methods to expel, suspend, or otherwise push out students whose behaviors do not meet the rules, norms, or expectations of school systems.[12] These methods range from the obvious to the obscure. Some are legitimate protections of the safety and learning environment for the majority of students. Yet, history has shown that these legitimate methods often migrate into a system of exclusion, turning the "falling through the cracks" case into a lacuna loaded with students that have

---

10. Michael Heise, *Educational Adequacy as Legal Theory: Implications from Equal Educational Opportunity Doctrine* 11 (Cornell Law Sch., Research Paper No. 05-028, Sept. 23, 2005), *available at* http://ssrn.com/abstract=815665. Heise referred to the phenomenon of litigants in school adequacy cases using data generated by the No Child Left Behind Act to prove their cases.

11. Patricia A. Massey & Stephen A. Rosenbaum, *Disability Matters: Toward a Law School Clinical Model for Serving Youth with Special Education Needs*, 11 CLINICAL L. REV. 271, 297–98 (2005). The authors ascribe the "dis-awareness" of this field to "lack of awareness [of its] civil rights implications" and latent disability bias. *Id.* at 271, 285.

12. The historic examples of school exclusion are embodied in two cases that led to the enactment of the IDEA in 1975. Mills v. Bd. of Educ., 348 F. Supp. 866 (D.D.C. 1972); Pa. Ass'n for Retarded Children v. Pennsylvania, 334 F. Supp. 1257 (E.D. Pa. 1971) (per curiam). In both of these cases, students with disabilities were excluded from educational opportunities through "warehousing" and the absence of procedural safeguards. *Pa. Ass'n for Retarded Children*, 334 F. Supp. at 1258–60, 1265; *Mills*, 348 F. Supp. at 868.

few prospects of returning to school and completing a vital credential for leading productive lives.[13]

A. Criminalizing Students: The School to Prison Pipeline

The "school to prison pipeline" describes a number of practices by school systems that can cause exclusion. Much like the term "environmental justice" described resistance against environmental practices that disproportionately affected low-income communities and communities of color,[14] the concept of the "school to prison pipeline" has galvanized civil rights groups. Many activists have formed campaigns that discourage schools from using the juvenile delinquency system as the only means of redressing problematic behavior by students, especially students with disabilities.[15] Several high profile episodes of school arrests, especially of very young children,[16] have led to calls for more sensitivity in handling students whose behaviors are symptomatic of emotional distress.[17]

After the tragic episode at Columbine High School, more schools turned to juvenile courts as corrective institutions. Many schools hired school resource officers,[18] and school safety became the mantra for arresting students for

---

13. *See generally* CHILDREN'S DEF. FUND, AMERICA'S CRADLE TO PRISON PIPELINE (2007) (reporting on risk factors and offering solutions to prevent neglect, abandonment, and criminalization).

14. *See generally* Dean Hill Rivkin, *Environmental Justice: A Universal Discourse*, 24 TEMPLE J. SCI. TECH. & ENVTL. L. 249 (2005) (describing Professor Ke Jian's linking of the environmental justice movement "to its animating cognate, the civil rights movement").

15. *See, e.g.*, ACLU Criminal Justice Project, School to School Pipeline—An Overview, http://www.aclu.org/crimjustice/juv/24704res20060321.html (last visited Jan. 18, 2008); S. Poverty Law Ctr., Legal Action, School-to-Prison Pipeline, http://www.splcenter.org/legal/schoolhouse.jsp (last visited Jan. 18, 2008); THE ADVANCEMENT PROJECT, EDUCATION ON LOCKDOWN: THE SCHOOLHOUSE TO JAILHOUSE TRACK 45 (Mar. 2005), *available at* http://www.advancementproject.org/reports/FINALEOLrep.pdf [hereinafter EDUCATION ON LOCKDOWN] (concluding that schools districts are "overreaching by inappropriately adopting law enforcement strategies" to address delinquency); NAACP Legal Def. Fund, School to Prison Pipeline, http://www.naacpldf.org/issues.aspx?issue=3 (last visited Jan. 18, 2008) (discussing and following the issue of the "School to Prison Pipeline").

16. *E.g.*, Tom Marshall & Johathan Abel, *In Class or Custody*, ST. PETERSBURG TIMES, Jan. 20, 2008, at 1A, *available at* 2008 WLNR 1138858.

17. *See* FLA. ST. CONFERENCE NAACP ADVANCEMENT PROJECT, NAACP LEGAL DEF. & EDUC. FUND, INC., ARRESTING DEVELOPMENT: ADDRESSING THE SCHOOL DISCIPLINE CRISIS IN FLORIDA 53–54 (2006) (recommending changes that local officials, state officials, juvenile court personnel, parents advocates, and education advocates implement).

18. *See* Nat'l Assoc. of Sch. Res. Officers, Introduction, http://www.nasro.org/about_nasro.asp (last visited Jan. 18, 2008) (describing school resource officers as "school based law enforcement officers, school administrators, and school security/safety professionals working as partners to protect students, school faculty and staff and the schools they attend"); *see also* OFFICE OF SCH. SAFETY AND LEARNING SUPPORT, TENN. DEP'T OF EDUC., SCHOOL POLICE DEPARTMENTS (2008) (reporting to the Tennessee General Assembly on the law and

education-related infractions. However, this practice existed before Columbine. In *Morgan v. Chris L.*,[19] a middle school filed a juvenile court petition against a student for allegedly kicking and breaking a water pipe in the school bathroom.[20] The student had been diagnosed with ADHD, a neurobiological disorder that can lead to impulsive, uncontrollable behavior.[21] Despite knowing about the diagnosis, the school system never identified the student as eligible for the protections of the IDEA.[22] Instead, the school filed a delinquency petition in the local juvenile court based on criminal vandalism.[23]

The parents filed for a due process hearing under the IDEA, claiming that Chris should have been certified as eligible for IDEA protections and that the school circumvented IDEA procedures.[24] The IDEA required that a school conduct a manifestation hearing to determine whether Chris's behavior was connected to his disability before initiating a potential change of placement.[25] The parents prevailed at the due process hearing, and the hearing officer ordered the school system to dismiss the petition, which had been stayed by the juvenile court.[26] On appeal, the District Court and the Court of Appeals affirmed the judgment of the hearing officer. Both courts indicated that the school system had ducked its special education responsibilities by shunting Chris's behavior problems to a forum that did not have the resources or the expertise to assist him.[27] In the case's aftermath, Congress amended the IDEA in 1977 by enacting a provision that allowed school systems to "report[] a crime committed by a child with a disability to appropriate authorities . . . ."[28] The sparse legislative history of the provision admonished schools not to "circumvent" the procedural safeguards of the IDEA, should a petition be filed.[29]

---

implications of employing school resource officers). The role and status of law enforcement officers in the schools remains controversial. *See, e.g.*, R.D.S. v. State, No. M2005-00213-SC-R11-JV, 2008 WL 315568, at *9–10 (Tenn. Feb. 6, 2008) (remanding for determination of whether SRO was a school official or a law enforcement officer before finally ruling on a motion to suppress).

19. 927 F. Supp. 267 (E.D. Tenn. 1994), *aff'd per curiam*, 106 F.3d 401 (6th Cir. 1997).

20. *Morgan*, 927 F. Supp at 269; Morgan v. Chris L., No. 94-6561, 1997 WL 22714, at *1 (6th Cir. 1997) (per curiam).

21. *Morgan*, 927 F. Supp. at 268.

22. *Id.* at 269.

23. *Id.*

24. *Id.* at 268.

25. *Id.* at 269 (quoting from the record of the hearing before the administrative law judge).

26. *Id.*

27. *Id.* at 271–72; *Morgan*, 1997 WL 22714, at *5–6.

28. 20 U.S.C. § 1415(k)(6)(A) (Supp. 2004) (originally enacted as 20 U.S.C. § 1415(k)(9)(A) (1997)).

29. 143 Cong. Rec. S4403 (daily ed. May 14, 1997) (statement of Sen. Harkin) (stating that schools should not use referrals to "circumvent [their] responsibilities under IDEA").

The incidence of school petitions is not well documented.[30] Since Columbine, courts have not been sympathetic to claims that juvenile courts do not have jurisdiction over school-filed petitions.[31] The degree of cooperation between juvenile courts and school systems varies dramatically on the local level. Some juvenile courts are not receptive to school-filed petitions, believing that the system is "dumping" children into the judicial systems. The courts understand that they lack the resources that schools have when it comes to developing plans for treatment and rehabilitation of youth offenders. Other juvenile courts only see their role as facilitating correction and punishment. In these courts, juveniles are often subjected to probation plans that rigidly require adherence to school rules and strict attendance. These plans are often recipes for serial violations based on minor infractions of school rules. They also place juveniles at risk of incarceration, especially those with mental or emotional impairments.

B. School Discipline

School discipline policies and practices have been the subject of intense controversy for some time.[32] Critics have argued that they fuel school exclusion and unfairness. First, studies show that school discipline is disproportionately leveled against students of color.[33] In a recent study, a Task Force appointed by the Mayor of Knox County, Tennessee, found that "[t]he data on school discipline shows clear disparities based on race."[34] Poverty, which in Knox County is correlated with race, was determined to be "a more significant indicator of disciplinary incidents than race."[35] Among other

---

30. *See generally* Eileen L. Ordover, *When Schools Criminalize Disability: Education Law Strategies for Legal Advocates* (April 2002) (unpublished manuscript), *available at* http://www.cleweb.org/Downloads/when_schools_criminalize_disabil.htm (discussing this phenomenon).

31. Joseph M. v. Se. Delco Sch. Dist., No. CIV.A.99-4645, 2001 WL 283154, at *5–6 (E.D. Pa. Mar. 19, 2001); Commonwealth v. Nathaniel N., 764 N.E.2d 883, 887 (Mass. App. Ct. 2002); *In re* Beau II, 738 N.E.2d 1167, 1171 (N.Y. 2000).

32. *See generally* TEXAS APPLESEED, TEXAS' SCHOOL-TO-PRISON PIPELINE: DROPOUT TO INCARCERATION (2007) (drawing a convincing connection between school discipline policies and practices and involvement in the juvenile justice system).

33. RUSSELL J. SKIBA ET AL., THE COLOR OF DISCIPLINE: SOURCES OF RACIAL AND GENDER DISPROPORTIONALITY IN SCHOOL PUNISHMENT 2 (Indiana Educ. Policy Ctr., Policy Research Report No. SRS1, June 2000), *available at* http://www.indiana.edu/~safeschl/cod.pdf; ADVANCEMENT PROJECT & CIVIL RIGHTS PROJECT AT HARVARD UNIV., OPPORTUNITIES SUSPENDED: THE DEVASTATING CONSEQUENCES OF ZERO TOLERANCE AND SCHOOL DISCIPLINE POLICIES vi (June 2000), *available at* http://www.civilrightsproject.ucla.edu/research/discipline/opport_suspended.php; EDUCATION ON LOCKDOWN, *supra* note 15, at 7.

34. DISCIPLINE TASK FORCE, KNOX COUNTY SCHS., RACIAL DISPARITY IN SCHOOL DISCIPLINE TASK FORCE—FINAL REPORT 7 (Mar. 12, 2007) (unpublished report), *available at* http://www.kcs.k12tn.net/reports/taskforce/discipline_task_force.pdf.

35. *Id.* at 3, 7.

suggestions, the Task Force recommended more training of school personnel in multicultural awareness and increased opportunities for dialogue addressing race issues.[36]

School discipline has a number of deep-seated problems.[37] First, the racial aspects of school discipline virtually guarantees that the students who are expelled live in neighborhoods with less community supports and services. Family incomes in these areas are generally lower. Once a student is suspended or expelled, the impetus to return to school is diminished. Long-term suspensions often lead to the practical termination of a student's educational career.

Second, school disciplinary rules are often fatally overbroad. "Behavior prejudicial to the good order" of the school[38] is hardly a standard that gives guidance to a student (or parents) on what types of behavior are subject to school discipline. Yet, standards such as this give administrators virtually unregulated discretion to exclude students for even minor misconduct. These codes provide a recipe for imposing exclusion on students who do not fit into the regimented nature of most public schools.[39]

Third, the minimal due process protections that were articulated in *Goss v. Lopez*[40] have become a facade for arbitrariness in determining both liability and punishment.[41] "Some kind of hearing"[42] has not protected students from administrators who impose their own idiosyncratic interpretations of school rules. In serious cases, where the prospect of long-term exclusion is high, the full panoply of due process procedures is often not afforded to students.[43] The vast majority of students are not represented by counsel at these base school hearings. Providing representation at these hearings could greatly improve students' chances. At least one concerted effort to supply counsel to students has yielded success in dropping the rates of expulsions and long-term suspensions.[44]

---

36. *Id.* at 1, 7.

37. *See* Marc Levin, *Schooling a New Class of Criminals? Better Disciplinary Alternatives for Texas Students*, POLICY PERSPECTIVE (Texas Pub. Policy Found., Mar. 2006), *available at* http://www.texaspolicy.com/pdf/2006-03-PP-DAEP-ml.pdf.

38. Tenn. Code Ann. § 49-6-3401(b)(1)(C) (Supp. 2007).

39. The broad discretion given to base-school administrators and appeals bodies—to set the duration of a suspension—mirrors the issue of sentencing discretion in criminal cases.

40. 419 U.S. 565 (1975).

41. Susel Orellana, *Advocacy at School Expulsion Hearings*, 9 AM.BAR ASS'N CHILD. RTS. LITIG. COMM. NEWSL. 5 (Winter 2007); Simone Marie Freeman, Note, *Upholding Students' Due Process Rights: Why Students Are in Need of Better Representation at, and Alternatives to, School Suspension Hearings*, 45 FAM. CT. REV. 638, 641–42 (2007).

42. *Goss*, 419 U.S. at 579.

43. *See, e.g.*, C.B. v. Driscoll, 82 F.3d 383, 386 (11th Cir. 1996) ("[O]nce school administrators tell a student what they heard or saw, ask why they heard or saw it, and allow a brief response, a student has received all the process that the Fourteenth Amendment demands."); Freeman, *supra* note 41, at 641–42.

44. Libby Sander, *In School Expulsion Cases, a Little Legal Advice Goes a Long Way*, 29

Finally, zero-tolerance policies have left a taint on schools from their prior misuse, though they are on the wane and often limited to serious offenses, such as gun possession or drug peddling.[45] Under these strict liability rules, where no finding of individual culpability or intent is necessary, school administrators do not have to exercise any discretion before excluding a student.[46] This mentality has bled into non-zero tolerance practices, giving administrators subtle power to make questionable findings in non-zero tolerance cases.

C. Special Education

Students with disabilities are especially vulnerable to the mechanisms of exclusion.[47] Exclusion of students with disabilities takes many forms. Initially, families may not recognize that their child qualifies for special education services and protections. Often, warning signs are overlooked. Behaviors are attributed to notions that the student is simply choosing inappropriate actions, is lazy, lacks motivation, or comes from bad genes. Students fortunate enough to cross the threshold for evaluation often are improperly found not to have a qualifying disability. If a disability is diagnosed, students can be denied eligibility by a finding that the disability does not adversely impact a student's education. Evaluations that result in a finding of no disability often are marred by not being sufficiently comprehensive, with not all suspected areas of disability being evaluated. Even if the evaluation was sufficiently comprehensive, all areas of suspected disability may not be addressed in the Individualized Education Program (IEP).[48] In these cases, if the family is not apprised of their right to request an Independent Education Evaluation (IEE),[49] the family will forfeit what may be the student's last chance to be identified for special education services. Additionally, a student may qualify under the first requirement that they have a disability but not be found to satisfy the succeeding requirements of eligibility for special education services—namely a need for special education services in order to succeed not only academically but also functionally and developmentally.[50] Many decisionmakers only look at

---

CHI. LAW. 60, 61 (2006).

45. ADVANCEMENT PROJECT & CIVIL RIGHTS PROJECT, *supra* note 33; RUSSELL SKIBA ET AL., AM. PSYCHOLOGICAL ASS'N COUNCIL OF REPRESENTATIVES, ARE ZERO TOLERANCE POLICIES EFFECTIVE IN THE SCHOOLS? AN EVIDENTIARY REVIEW AND RECOMMENDATIONS (2006), *available at* http://www.apa.org/ed/cpse/zttfreport.pdf.

46. *See* Seal v. Morgan, 229 F.3d 567 (6th Cir. 2000).

47. Daniel J. Losen & Kevin G. Welner, *Disabling Discrimination in Our Public Schools: Comprehensive Legal Challenges to Inappropriate and Inadequate Special Education Services for Minority Children*, 36 HARV. C.R.-C.L. L. REV. 407, 419 (2001).

48. *See generally* U.S. DEP'T OF EDUC., A GUIDE TO THE INDIVIDUALIZED EDUCATION PROGRAM (July 2000), *available at* http://www.ed.gov/parents/needs/speced/iepguide/iepguide.pdf (providing an overview of the IEP process).

49. 34 C.F.R. § 300.502(a) (2007).

50. 20 U.S.C. § 1401(3)(A)(i)–(ii) (2007); 20 U.S.C. § 1414(b)(3)(A)(ii) (2007); 34 C.F.R. § 300.301(c)(2).

adverse impact on the student's academic achievement and do not consider the adverse impact on the student's functional and developmental progress. Whatever the reasons for the determination of ineligibility, students who need assistance are frequently bypassed.

Another group of students are not identified because some believe that aggressive intervention strategies might forestall the need to label a student as disabled. The 2004 IDEA Amendments allow schools to use 15% of IDEA funds to provide early intervening services (EIS)[51] to students at risk of needing special education services, prior to referral for evaluation. However, few rules prescribe which students fit this category, when or how parents are made aware of the potential need to evaluate, or how EIS squares with a referral to evaluate. Such an option can distract the team from referring a student for evaluation. This can cause an even longer delay before students receive appropriate support and special education services.

School personnel may suspect that students have one or more of the so-called "hidden disabilities," such as ADHD, ADD, Specific Learning Disabilities (SLDs), language processing disorders, and others. In these cases, the team can avoid using a trial period to ascertain whether the student has learning problems. The team may decide to use special Response to Intervention (RTI)[52] practices that are specifically recommended in conjunction with SLDs.[53] Problems arise when these methods become protracted and are never re-evaluated. As of yet, these methods have insubstantial scientific or objective grounds and few evidence-based procedures, which leaves students vulnerable to subjective variables.[54] Without a focused set of goals and strategies, a student may drift over time. If a school does not attend to the student's problems, he or she may never attain comprehensive assistance through an IEP or a 504 plan.[55]

Some students exhibit challenging behaviors that cause them to be perceived as "just bad kids." School administrators have used this as an excuse to exclude them or deny them evaluation. When these students are referred for evaluation, often the outcome is delay and an inaccurate and incomplete identification of disability is formed.[56] Also, school systems may, through less

---

51. 20 U.S.C. § 1413(f)(1) (2007); 34 C.F.R. § 300.226(a).

52. 34 C.F.R. § 300.307(a)(2).

53. *See* NAT'L JOINT COMM. ON LEARNING DISABILITIES, RESPONSIVENESS TO INTERVENTION AND LEARNING DISABILITIES (2005), *available at* http://www.ncld.org/index.php?option=content &task=view&id=497.

54. *See* CTR. FOR MENTAL HEALTH IN SCHS., MENTAL HEALTH IN SCHOOLS AND SCHOOL IMPROVEMENT: CURRENT STATUS, CONCERNS, AND NEW DIRECTIONS 4-1 to 4-10 (2008).

55. A 504 Plan specifies accommodations and modifications for students with qualifying impairments as defined by § 504 of the Rehabilitation Act. 29 U.S.C. § 701 et seq.; *see* 34 C.F.R. § 104.33.

56. A classic example of this practice involves students with ADHD. Despite specific categorization as a qualifying disability under the IDEA's "Other Health Impaired" classification, school systems may refuse to certify any student diagnosed with ADHD on the

than aggressive outreach, avoid their IDEA "child find"[57] obligation proactively to identify and recommend students for evaluation.

School officials commonly use school discipline actions illegally to exclude students who they know are at risk of having a disability, instead of referring them for evaluation.[58] These students rise through the grades with little academic success, while frequently being disciplined, suspended, or expelled. Many of these students also have problems in other parts of their lives, such as traumatic family circumstances, multiple moves resulting in different school settings, parental divorce, family drug abuse, and more. Even if finally evaluated, many students with a history of "behavior difficulties" also are not comprehensively assessed. This results in non-identification of hidden disabilities like learning disabilities, speech and language processing disorders, depression, and bipolar disorder.

When students with disabilities violate school rules or act in inappropriate ways, administrators may suspend them for no more than ten school days without it being a change in educational placement.[59] These students may be deprived of educational services during this time. If the suspension lasts for more than ten days, the school must conduct a manifestation hearing to determine whether a change of placement is appropriate.[60] The rules governing manifestation hearings changed in the 2004 IDEA Amendments. They gave greater latitude for schools to find that a student's behavior is not a manifestation of the student's disability.[61] As a consequence, although the student is still entitled under IDEA to receive continuing educational services, he or she may be transferred to an interim alternative educational setting.[62] These settings are places where virtually all students have exhibited challenging behaviors, and the quality of education is questionable. In these placements, a student's IEP may be difficult, if not impossible, to implement. Some refer to these settings as "warehouses." They are schools characterized by a maze of punitive processes and very little in the way of Positive Behavior Support,[63] procedures, or effective behavior intervention techniques. As a consequence of this neglect, students may be inhibited from making meaningful educational

---

theory that, with medication or therapy, the student's behaviors can be manageably corralled. The corollary theory for excluding this entire segment of students is that the ADHD is not adversely affecting the student's education because the student has passable grades.

57. 34 C.F.R. § 300.111.

58. Joseph B. Tulman, *Disability and Delinquency: How Failures to Identify, Accommodate, and Serve Youth with Education-Related Disabilities Leads to Their Disproportionate Representation in the Delinquency System*, 3 WHITTIER J. CHILD & FAM. ADVOC. 3, 36 (2003).

59. 34 C.F.R. § 300.536(a).

60. 20 U.S.C. § 1415(k)(1)(E)(i) (2007).

61. *Id.* § 1415(k)(1)(E)(i)(I) (requiring that the student's conduct be caused by or have a "direct and substantial relationship to" the disability).

62. 34 C.F.R. § 300.530(d)(2).

63. 20 U.S.C. § 1414(d)(3)(B)(i) (2007); 34 C.F.R. § 300.324(a)(2)(i).

progress. Alienation from the education process is a logical consequence of such treatment.

Standardized test performance is another way to exclude students with disabilities. Many students with disabilities find standardized tests to be a frustrating barrier. Since the enactment of the accountability requirements in the No Child Left Behind Act of 2001,[64] all states have developed protocols that include standardized testing that students must successfully complete before they may graduate with a regular high school diploma. Often, challenged students need supplemental assistance to prepare them to take and to succeed in standardized testing. First, administrative staff must recognize that students have these needs. Second, they must create strategies to assist in preparation and successful execution of state tests. Students who are eligible for special education services should have this incorporated in their overall program far in advance of testing.[65] Without the existence of adequate programs, many students fail these tests, and thus, they do not receive regular diplomas. Future gainful employment could hang in the balance.

Students with disabilities also must have transition services plans incorporated into their IEPs by age sixteen.[66] These services must include "[a]ppropriate measurable postsecondary goals based upon age appropriate transition assessments related to training, education, employment, and, where appropriate, independent living skills; and . . . [t]he transition services (including courses of study) needed to assist the child in reaching those goals."[67] Such services are crucial to the futures of students with disabilities, considering that students with disabilities have more trouble fitting into real-life roles without preparation and transition. Many educators do not provide adequate transition services to students with disabilities, despite this being their last chance for a successful transition from high school into higher education, the working world, or independent living. The 2004 IDEA Amendments significantly tightened schools' responsibilities to ensure that a meaningful transition plan is created and implemented.[68]

D. Truancy

Compulsory education laws compel schools to enforce attendance policies. State funding and NCLB requirements have heightened the focus on ensuring

---

64. 20 U.S.C. § 6301 (2007) (reauthorizing the Elementary and Secondary Education Act of 1965).

65. 34 C.F.R. § 300.324(a).

66. 34 C.F.R. § 300.320(b).

67. 43 C.F.R. § 300.320(b)(1)-(2).

68. The goals of a transition plan must now be measurable, and the transition services designed to achieve these goals must be included in the student's IEP. 34 C.F.R. § 300.320(b). Also, if an outside agency fails to provide the student with the required transition services, an IEP team must be reconvened to identify alternative strategies to meet the transition objectives. 34 C.F.R. § 300.324(c)(1).

that students attend school regularly. The concept of truancy is an old one.[69] Today, the once feared truant officer has transformed into a team composed of school personnel, juvenile court staff, district attorneys, and social services representatives. Parents are warned about their child's poor attendance, excoriated for the child's behavior, and sometimes prosecuted for neglect.

However, truancy laws fail to address the root causes of a student's aversion to school.[70] Some truants are actually students with unidentified special education needs.[71] Tighter rules for screening and evaluation are a necessary step for identifying why these students stop attending school. For students already certified under the special education laws, truancy penalties are not the answer.[72] Instead, an IEP or 504 team should meet promptly to ascertain what in the student's IEP or section 504 plan needs to be modified. The team may need to introduce or intensify services, such as social work or psychological counseling. The team may even formulate wrap-around services, which are a heavy regime of support for the student and her family.

Before prosecuting parents or students, truancy enforcers should exhaust a number of other explanations.[73] For example, bullying has received attention both in the popular press and by school systems and legislatures.[74] School systems and courts should first protect students vulnerable to bullying before taking action against the family. Likewise, schools should explore whether an insensitive or poorly trained teacher could be the cause of a student's skipping school before punishing the student for truancy.

---

69. *See, e.g.*, Harold O. Levy & Kimberly Henry, Op-Ed., *Mistaking Attendance*, N.Y. TIMES, Sept. 2, 2007, § 4, at 11. The article states, "America is awash in casual truancy," further noting that "[s]kipping school has been going on since biblical times," and that insufficiently meaningful statistics perpetuate denial about the problem and failure to identify appropriate solutions. *Id.*

70. *See* Lorenzo A. Trujillo, *School Truancy: A Case Study of a Successful Truancy Reduction Model in the Public Schools*, 10 U.C. DAVIS J. JUV. L. & POL'Y 69, 83 (2006) (describing a successful early intervention program for reducing truancy).

71. *See West Lyon Community Sch. Dist. and Northwest Area Educ. Agency*, 48 INDIVIDUALS WITH DISABILITIES EDUC. L. REP. 232 (State Education Agency Iowa 2007) (finding school violated IDEA by failing to evaluate student's psychological needs based on chronic absenteeism).

72. *See, e.g.*, Independent Sch. Dist. No. 284 v. A.C., 258 F.3d 769 (8th Cir. 2001) (holding that student's truancy was caused by an emotional disability and required a residential placement).

73. The tragic death of 16-year-old Kaleb Shelton illustrates this point. Jim Balloch, *DCS Defends Omni Visions*, KNOXVILLE NEWS-SENTINEL, Dec. 17, 2007, at B1, *available at* http://license.icopyright.net/user/viewFreeUse.act?fuid=NjQxNTyO. Kaleb had a learning disability. *Id.* He was prosecuted for truancy, eventually expelled from school, and placed in foster care. *Id.* He died at the hands of his foster father. *Id.*

74. TENN. CODE ANN. §§ 49-6-1014 to -1019 (2002).

E. Push-Out Practices

Some schools resort to "push-out" practices with students who perform poorly and are not eligible for special education protections. These schools appear more concerned about test scores and higher achievement than reaching troubled students. Sometimes these practices of exclusion are subtle. For example, a school administrator tells a student that her best option is to drop out and take the GED because she is behind in credits for graduation. These "drop-outs" often fail to receive either a GED or regular diploma. Litigators in New York City have successfully challenged one type of exclusionary practice,[75] but most are under the radar of effective accountability.

F. Alternative Education

Many school districts do not maintain alternative schools to educate students whom they suspend or expel for infractions of school rules.[76] Lengths of suspensions vary, but suspensions and expulsions last anywhere from one to 180 days. Even the shorter suspensions can harm a student's educational progress when students miss tests and school work. A 180-day expulsion for a zero tolerance offense often means that a student will miss two or three semesters of school, a sure incentive to drop out.

Even in school systems that offer alternative education, barriers to participation exist. Many systems do not provide transportation to alternative schools.[77] This is especially burdensome when the alternative programs meet at night or in parts of town not readily accessible by public transportation. Students who can attend an alternative school program often face inadequate instruction.[78] *C.S.C. v. Knox County Board of Education*[79] involved a three-hour, four-day-a-week Night Alternative Program (NAP), which only offered computer-based programs and did not cover all aspects of the State's required curriculum.[80] A challenge to the adequacy of the instruction under state regulations failed. The court held that such instruction was within the

---

75. *See generally* Elisa Hyman, *School Push-Outs: An Urban Case Study*, 38 CLEARINGHOUSE REV. 684 (2005) (describing the litigation and awareness campaign).

76. *See, e.g.*, DIGNITY IN SCHOOLS CAMPAIGN, ALTERNATIVE SCHOOLS AND PUSHOUT: RESEARCH AND ADVOCACY GUIDE (2007).

77. *See, e.g.*, Kadrmas v. Dickinson Pub. Schs., 487 U.S. 450, 461–62 (1988) (upholding a North Dakota law allowing local school boards to charge families a user fee for bus service). Justice O'Connor remarked, "The Constitution does not require that such service be provided at all . . . ." *Id.* at 462.

78. David J. D'Agata, *Alternative Education Programs: A Return to "Separate but Equal?"*, 29 NOVA L. REV. 635, 640 (2005).

79. C.S.C. v. Knox County Bd. of Educ., No. E2006-00087-COA-R3CV, 2006 WL 3731304, at *1 (Tenn. Ct. App. Dec. 19, 2006).

80. *Id.* at *1–3, 7.

exclusive province of the school system.[81] Likewise, state regulations required educators to make counseling services "accessible" to students in alternative schools.[82] Despite testimony showing that the school system had not provided such behavioral services in two-years, the court did not require the system to develop a written plan for providing such services.[83] Such inattention to deficient services in alternative schools is all too common.[84]

G. Education in Jails and Correctional Institutions

Incarcerated youths in juvenile detention facilities or state youth prisons often receive inferior educational services.[85] Students who are or should be in special education services are particularly vulnerable. The prison staff have free reign to fit students in an education program, rather than tailoring a program for individual students. The lack of advocates for students exacerbates this situation.

Youths who have been transferred from juvenile courts to adult jails often do not receive any education, despite long waiting periods for a plea or trial.[86] Criminal defense lawyers customarily do not push for educational services in jail out of genuine concern that their clients might disclose harmful information. As a consequence, these youths never receive even basic instruction.

H. General Inadequacy of Educational Opportunities

The inadequate distribution of resources in many school systems (not the statewide issue of school financing, but its local counterpart) creates pockets of schools that lack a number of ingredients for good education. These schools are invariably in poorer areas of the community, those that disproportionately house a sizeable number of households of color. Although, in part, NCLB was designed to redress these inequities, they persist all around the country. Schools in this category lack decent facilities, advanced courses, experienced teachers, guidance counselors, meaningful early intervention programs, diverse extracurricular activities, and other criteria of quality education. Some school administrators fail to implement even the remedial measures of NCLB, such as after-school tutoring programs for students in "failing" schools.[87] The

---

81. *Id.* at *7–9.

82. *Id.* at *9.

83. *Id.*

84. *See generally* JOHN G. MORGAN, COMPTROLLER OF THE TREASURY, TENNESSEE'S ALTERNATIVE SCHOOLS (April 2005), *available at* http://www.comptroller1.state.tn.us/repository/re/final_alt_school.pdf (last visited Jan. 18, 2008).

85. *See, e.g.*, Marcus X. v. Adams, 856 F. Supp. 395 (E.D. Tenn. 1994) (describing the inadequate services provided to one student while in a juvenile correction facility).

86. Doe v. Knox County, No. 143196-2 (Tenn. Ch. Ct. filed June 7, 1999).

87. Rhea R. Borja, *Companies Want Changes in NCLB Tutoring Policies*, EDUC. WK.,

connection between inadequate distribution of resources and school drop-out rates is not easily documented. Some blame entrenched local politics of school boards and municipal government—matters largely insulated from judicial review.

III. CONFRONTING SCHOOL EXCLUSION IN THE COURTS: OPPORTUNITIES AND ISSUES, COMPLEXITIES AND CONTRADICTIONS

A. Motivations Behind School Exclusion

Motivations behind exclusionary policies and practices are mixed and complex. They implicate profound questions about democracy and education. Despite glowing rhetoric about the importance of education to our economic, political, and social systems, school exclusion remains a well-kept secret, except to the families that it affects. To understand the roles that lawyers and courts play—or should play—one must begin by examining the motives underlying school exclusion.

1. Racial and Ethnic Currents

Many schools have not yet embraced the vast social and cultural changes that are transforming public education today. Institutional racism is prevalent. Its subtlety makes it difficult to discuss, much less root out. The same holds true with the cross-cultural currents that are infused into public education. Elected school boards and politicians sometimes make change slow and difficult.

2. Regimentation

For a long time, the scheme of public education has conflicted with the needs and expectations of growing segments of the school-age population. Critics fault the NCLB for allowing rigid testing to push weaker students by the wayside. Until the education community appreciates the necessity of plans for all students, not just for students with disabilities, the failure to adapt to different learning needs will continue to frustrate many students and their families. Without intending to oversimplify, this frustration on the part of the students often manifests itself as "bad behavior," which is then used as a justification to exclude students.

January 24, 2007, at 10, *available at* http://www.edweek.org/ew/articles/2007/01/24/20tutor.h26.html.

3. Politics

Schools exclude students with problematic behaviors because parents of other students complain that these students create disruptions in the classroom or pose safety problems. These concerns should not be dismissed. They usually stem from good-faith efforts on the part of the majority of parents in a school to protect their children. However, parents can overreact, and they can put serious pressure on school administrators. If a principal fails to remove or isolate a problematic student, complaints to the superintendent or the school board could stall that principal's career. Majority rule has driven public education throughout history. The topic of exclusion is no exception, no matter how vulnerable a particular student or class of students may be.

4. Failure to Adopt Evidence-Based Practices

School systems change slowly. Many universities' schools of education maintain cozy relationships with their local and statewide school systems for a variety of self-serving purposes. Nevertheless, some institutions are conducting cutting-edge research on issues of behavioral support. School reform advocates say that anti-exclusion solutions should be directed primarily at school employees, and not solely focused on students. Educational pioneers are testing promising reforms, such as school-wide programs of positive behavioral intervention and supports, at school systems across the country.[88] These programs focus on ending the practice of referring normal disciplinary action to the courts and reducing the number of suspensions and expulsions. They also attempt to create a climate of tolerance and good citizenship among students and teachers. Yet, supporters of the status quo often resist these innovations because they conflict with the prevailing philosophy about school discipline.

B. Why Litigation?

Whether conducted by legal clinics, legal services programs, nonprofit public interest law firms, private attorneys, or government agencies, litigation over school exclusion must be carefully thought through. Blending direct

88. *See, e.g.*, Press Release, All American Patriots, Barack Obama: Obama, Durbin, Hare Introduce Bill to Improve Student Behavior in Schools (Oct. 3, 2007), *available at* http://www.allamericanpatriots.com (search "All American Patriots" for "Bill to Improve Student Behavior"; then follow "Barack Obama" hyperlink); Press Release, S. Poverty Law Ctr., SPLC Wins Special-Education Services for Baton Rouge Students (Nov. 16, 2006), *available at* http://www.splcenter.org/news/item.jsp?aid=224. Positive behavior intervention and supports is a school-wide program in which all school personnel are trained to recognize students with emotional problems and to respond to these students using proven methods of behavioral support. *See generally* Positive Behavioral Interventions and Supports, School-Wide PBS, http://www.pbis.org/schoolwide.htm (last visited Jan. 18, 2008) (describing and promoting this discipline).

representation of individual students with systemic reform strategies requires a comprehensive understanding of the local context. This includes becoming familiar with the school system, the state and federal courts, the advocacy community, grassroots groups, and the political landscape. A court-focused, rights-based approach may set back reform efforts if the conditions are not ripe for change. On the other hand, restraint from litigation sometimes means ignoring individual needs, which creates cruel paradoxes for lawyers in this field. This highlights some of the challenges of modern day public interest lawyering.

In *Law and School Reform: Six Strategies for Promoting Educational Equity*,[89] various authors adduced several rationales in favor of using litigation to confront educational inequities like school exclusion. Their justifications include the following: (1) to compel additional resources and accountability to fill gaps education to vulnerable groups, (2) to correct market failures in the distribution of educational resources, (3) to correct bureaucratic failures, (4) to challenge political power, and (5) to give parents a "voice" in educational decision-making.[90] As described in the book's six case studies of education reform litigation, realizing these goals can precipitate meaningful changes for marginalized students. But the test of litigation's mettle is whether a judge's decree will bring about lasting reform. The question remains as to whether litigation will alienate future "collaborative" efforts for change or stifle relationships necessary for lasting success.

Other compelling reasons support turning to litigation to redress school exclusion. A major reason is the opportunity to respond to a specific child and family in crisis. Parents do not turn to lawyers willy-nilly just to sue their children's schools. They hire lawyers as a last resort when their concerns have been ignored for a long period. Even a short time out of school can harm a student's academic progress and cause emotional distress. Labeling a student as "bad" can cause long-term consequences for a child's development. A lawyer might be in an ideal position to protect a student from exclusion by litigating existing rules.

Still, a myopic focus on individual cases, however successful under conventional measures, may stymie changes that would benefit the same

---

89. LAW AND SCHOOL REFORM: SIX STRATEGIES FOR PROMOTING EDUCATIONAL EQUITY (Jay P. Heubert ed., Yale Univ. Press 1999). This excellent volume strikes chords of hope and doubt about reliance on litigation to achieve education reform. There is little question that in certain areas—school finance or special education—litigation precipitated profound changes in the provision of educational services to children in poor school districts and to disabled students. But proper skepticism about the fine balance of law, policy, politics, and advocacy in this field pervades the volume, as it does in this Article. *See also* Michael Heise, *Litigated Learning, Law's Limits, and Urban School Reform Challenges*, 85 N.C. L. REV. 1419 (2007); James S. Liebman & Charles F. Sabel, *A Public Laboratory Dewey Barely Imagined: The Emerging Model of School Governance and Legal Reform*, 28 N.Y.U. REV. L. & SOC. CHANGE 183 (2003).

90. LAW AND SCHOOL REFORM, *supra* note 89.

student throughout her school career. Public interest lawyers must weigh this possibility in light of a system where the unmet need for legal representation outpaces the available supply of knowledgeable lawyers. Because the bulk of for-profit education representation is conducted by small firms or solo practitioners, these lawyers' financial needs may prevent them from taking clients who are unable to pay even a reduced fee. Turning away a potential client is a serious matter. In some settings, non-lawyer advocacy organizations may assist families in representing their child's needs, especially in special education cases that do not reach the due process administrative hearing level.[91] But many cases require skilled lawyers with working knowledge of the technical and institutional dimensions of education representation. Meeting the immediate needs of a child may be the right course to take from both an ethical and moral standpoint. All legal players in this field must set priorities; not every individual case can be served. Lawyers must carefully examine the waterwheel of cases to determine where the limited legal resources can be most effectively allocated.

C. Why Not Litigation?

Deciding between using either litigation or extrajudicial advocacy to combat school exclusion is unnecessary. A blend of strategies is the hallmark of modern public interest advocacy, regardless of the field. But striking an appropriate balance depends on trained vision and good timing.[92] Legal strategists must account for pitfalls in anti-exclusion litigation when developing a long-range advocacy strategy.

First, as *Brown v. Board of Education*[93] and its progeny have demonstrated, establishing a new rule that will benefit an individual client or even a class of persons is not always enough to fix the underlying policy or practice. Unless careful attention is paid to implementation, bureaucratic resistance and

---

91. *See* Stephen A. Rosenbaum, *The Juris Doctor Is* In*: Making Room at Law School for Paraprofessional Partners*, 75 TENN. L. REV. 315, 323–29 (2008) (describing the unique opportunities of paralegals to engage in advocacy on behalf of parents and students in the special education setting).

92. *See* Monique L. Dixon, *Combating the Schoolhouse-to-Jailhouse Track Through Community Lawyering*, 39 CLEARINGHOUSE REV. 135, 141–43 (2005); Amy M. Reichbach, *Lawyer, Client, Community: To Whom Does the Education Reform Lawsuit Belong?*, 27 B.C. THIRD WORLD L.J. 131 (2007) (describing the difficulty of keeping focus on the client-lawyer relationship and the significance of client autonomy during the course of complex NCLB reform litigation). The debate over the proper balance of legal and extralegal strategies for reform is ongoing. *See* Orly Lobel, *The Paradox of Extralegal Activism: Critical Legal Consciousness and Transformative Politics*, 120 HARV. L. REV. 937 (2007); Scott Cummings, *Critical Legal Consciousness in Action* (UCLA Sch. of Law Pub. Law & Legal Theory Research Paper Series, Research Paper No. 07-24; NYLS Clinical Research Inst., Research Paper No. 07/08-5, 2007), *available at* http://ssrn.com/abstract=998040 (last visited Jan. 18, 2008) (responding to the Lobel article).

93. 347 U.S. 483 (1954).

maneuvering can attenuate new rules. To ensure full relief, lawyers must coordinate with savvy clients or advocacy organizations. This takes vigilant, time-consuming, and resource-intensive monitoring.[94]

Second, enforcement is often necessary but difficult. Returning to the court that granted the relief is sometimes problematic. Judicial attention and resolve can wane. State judges, most of whom are elected, inevitably keep an open eye on the impact of a decision against the education system, which is often a community's largest municipal agency. The history of serial enforcement in prison litigation cases shows how political backlash can erase an otherwise promising judgment.[95] Concerns about judicial expertise, separation of powers, and the cost of implementation can intrude on the enforcement process, freezing the relief that was granted.

Third, long-term implementation can become too lawyer-centric. Without mechanisms for involving affected clients in the post-decree, there is real danger in leaving too much decision-making power in lawyers' hands, which removes cases from the evolving realities of the clients' needs and concerns.[96] Questions of accountability, endemic to class action litigation, arise.[97] Lawyers must account for the time it takes to address these dynamics when planning their commitment to a case.

Fourth, anti-exclusion litigation, especially in special education cases, may be too specialized for lawyers who do not concentrate in the field or who do not have steady back-up resources to consult on an ongoing basis. For example, public defenders would be a natural corps of lawyers to litigate anti-exclusion cases regularly. But many public defender offices already stretch staffing

---

94. Implementation committees, paid for by government defendants and composed of representatives of the plaintiffs, have been effective in public institutional cases involving the environment and prisons and jails. *See generally* Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015 (2004). In the context of school exclusion, meaningful reform should involve the institution of school-wide practices that focus on preventing conflict and chronic behavior issues. Today, a program called "Positive Behavior Interventions and Supports" (PBIS) represents a promising alternative to the current system of exclusion. National Technical Assistance Center on Positive Behavioral Interventions and Supports, http://www.pbis.org/ (last visited Feb. 18, 2008). Implementing such a program must involve a lawyer with the school community in sustained ways.

95. Harvey Berkman, *Proud and Wary, Prison Project Director Bows Out*, NAT'L L.J., Jan. 8, 1996, at A12 (observations of ACLU Prison Project Director Alvin Bronstein).

96. *See* Jennifer Gordon, Essay, *The Lawyer Is Not the Protagonist: Community Campaigns, Law, and Social Change*, 95 CAL. L. REV. 2133 (2007).

97. *See* Derrick A. Bell, Jr., *Serving Two Masters: Integration Ideals and Client Interests in School Desegregation Litigation*, 85 YALE L.J. 470 (1976); *see also* Gary Bellow & Jeanne Kettleson, *From Ethics to Politics: Confronting Scarcity and Fairness in Public Interest Practice*, 58 B.U. L. REV. 337, 341 (1978) (arguing that in the public interest arena, "the ability of clients who are being represented to keep their attorneys accountable is limited by their lack of economic leverage").

beyond advisable capacity. Also, litigating education cases would take a back seat to the daily grist of criminal defense.[98]

Fifth, anti-exclusion cases inevitably drift into litigation that challenges the adequacy of education. As illustrated in *C.S.C.*, such challenges are rarely successful, even with access to representation.[99] In that case, after the court established the right to alternative education under state constitutional and statutory rules, the plaintiffs were compelled to challenge the adequacy of the alternative program that the school system created in response to the threshold ruling.[100] As is discussed above, the effort failed.[101] History has shown that adequacy challenges require intensive fact investigation, close client communications, substantial discovery, and expert testimony. Such challenges have taken decades to litigate in New York City and Boston[102] and thus represent a daunting prospect for many clinics and nonprofit public interest firms.

IV. CONCLUSION

The campaign to reduce or eliminate school exclusion involves complicated local and national strategies. Litigation is one component, but it should not be the exclusive focus. Rather, litigation should be reserved for situations where individual needs are critical. The growing recognition of the legal needs in this field will bring much trial and error. Candid sharing of experiences among the involved lawyers, advocates, and clients will be indispensable to long-term reform.

Law schools' legal clinics can play an important role in this effort.[103] Clinical law teachers understand and care about pedagogy. They are uniquely suited to judge the quality of education. For clinics in universities with

---

98. Thoughtful public defender programs such as the Public Defender Service in Washington, D.C., and Washington state's public defender service have staff attorneys dedicated to education work for their juvenile clients, but these programs are the exception.

99. C.S.C. v. Knox County Bd. of Educ., No. E2006-00087-COA-R3CV, 2006 WL 3731304, at *1, *16 (Tenn. Ct. App. Dec. 19, 2006).

100. *See supra* notes 79–82 and accompanying text.

101. *See supra* notes 83–84 and accompanying text.

102. *See, e.g.*, Hancock v. Comm'r of Educ., 822 N.E.2d 1134 (Mass. 2005); Campaign for Fiscal Equity, Inc. v. State, 801 N.E.2d 326, 348–50 (N.Y. 2003) (issuing a remedial order as the culmination of more than a decade of litigation); Campaign for Fiscal Equity, *CFE v. State*: Ensuring Every New York Child Their Constitutional Right to a Sound Basic Education, *available at* http://www.cfequity.org/Litigation_Update_1page.pdf (last visited Jan. 18, 2008).

103. Three recent articles offer sophisticated analyses of the importance of broad-gauged education and advocacy strategies in law school legal clinics. *See* Anthony V. Alfieri, *(Un)covering Identity in Civil Rights & Poverty Law*, 121 HARV. L. REV. 805, *available at* http://www.harvardlawreview.org/issues/121/jan08/alfieri.shtml; Sameer M. Ashar, *Law Clinics and Collective Mobilization*, 14 CLINICAL L. REV. (forthcoming Apr. 2008), *available at* http://ssrn.com/abstract=1022366; Louise G. Trubek, *Crossing Boundaries: Legal Education and the Challenge of the "New Public Interest Law"*, 2005 WIS. L. REV. 455 (2005).

progressive education schools, opportunities for interdisciplinary learning and collaboration exist. Clinics reluctant to embark on cases that logically lead to "impact" work can focus on target schools. School disciplinary hearings are typically short matters, ideal for law student representation. Aggregating a number of these cases may reveal patterns in disciplinary practices that could persuade even entrenched school administrators—without pursuing "impact" litigation—to revise their policies and practices.

There are downsides, however, that must be addressed. Once a community learns that a law school legal clinic is occupying this field, the clinic could experience a deluge of clients seeking representation. Also, on a personal level, clinicians and clinic students frequently have children in the same school system and would understandably not be immune to concerns about retaliation, however remote. These considerations should be weighed prudently, deliberatively, and collaboratively. As Justice Brennan recognized, the work itself is an expression of democracy that often does not inhere in private litigation. Being part of the solution to school exclusion, not part of the problem,[104] requires creativity, sensitivity, and vision. Clinics should cultivate, inculcate, and model these attitudes and qualities.

104. *See* Gary Bellow, *Turning Solutions into Problems: The Legal Aid Experience*, 34 NLADA BRIEFCASE 106, 108 (1977) (arguing that many federal aid programs "may be supporting the very inequalities that brought a federally financed legal aid program into being").

IN RE GAULT AND THE PROMISE OF SYSTEMIC REFORM

KATHERINE R. KRUSE*

Systemic reform begins when an observer perceives a gap between the ideals upon which a system was founded and that system's actual mode of operation. Such was the situation when the United States Supreme Court observed the Arizona juvenile justice system's treatment of Gerald Gault, a fifteen-year-old who was committed to the State Industrial School for up to six years as punishment for making a lewd telephone call to a neighbor lady.[1] The state's extreme and needless intervention in Gerald Gault's life was anathema to the Supreme Court.[2] But, it signaled more than the mere inattention of an individual judge in the case of a single boy. For the Court, Gault's treatment revealed a gap between the juvenile court's founding ideals of "careful, compassionate, individualized treatment"[3] and its operation, which was characterized by inadequate fact-finding[4] and systemic inattention to the individualized needs of children.[5] As the Court famously noted, "Under our Constitution, the condition of being a boy does not justify a kangaroo court."[6]

Gault represents an effort at systemic reform—a purposeful alteration of the structure, procedure, or resources of a law-administering system that aims to better align the system's operation with the principles or ideals on which it is based.[7] Because the Court diagnosed the systemic problem as excessive

---

* Associate Professor of Law, William S. Boyd School of Law. JD University of Wisconsin Law School, MA Universtiy of Wisconsin-Madison (Philosophy), BA Oberlin College. I would like to thank Benjamin Barton, Mae Quinn and the other fine clinicians at Tennessee for inviting me to participate in this symposium. I would also like to thank my colleagues Annette Appell, Megan Chaney, Joan Howarth, Leticia Saucedo, David Thronson and other members of the Thomas & Mack Legal Clinic for allowing me to present this essay at a clinic research forum at Boyd School of Law and providing helpful feedback to me. I would also like to thank Michael Smith for numerous discussions about systemic reform and Pamela Towers for including me as a "stakeholder" in the JDAI reforms in Clark County Nevada.

1. *In re* Gault, 387 U.S. 1, 4, 7 (1967).
2. *Id.* at 28–29.
3. *Id.* at 18.
4. *Id.* at 5, 25, 32 & n.21.
5. *See id.* at 51 (noting mounting evidence that the system's practice of encouraging juveniles to confess without informing them of their right to remain silent did not contribute to individualized treatment).
6. *Id.* at 8.
7. *Cf.* RAYMOND T. NIMMER, THE NATURE OF SYSTEM CHANGE: REFORM IMPACT IN THE CRIMINAL COURTS 4 (1978) (defining reform as "any planned change in the structure, rules, or resources of the judicial process" and describing reform as "a stimulus intended to produce in various individuals, groups, or organizations a particular behavioral change that is regarded as desirable").

informality in juvenile court proceedings, it introduced a host of procedural due process rights as a remedy.[8] Importantly, the Court maintained that procedural formality was not intended to transform the juvenile justice system into something new, but to return it to its founding ideals.[9] It critically examined the "claimed benefits" of the informal processes in juvenile court and concluded that recent studies showed "with surprising unanimity" that "the appearance as well as the actuality of fairness, impartiality and orderliness—in short, the essentials of due process—may be a more impressive and more therapeutic attitude so far as the juvenile is concerned."[10]

Like most top-down law reform efforts, *Gault*'s attempt at systemic reform was problematic because it was not self-executing. For a reform effort to be effective, the individuals responsible for implementing changed policy at lower levels—in this case, prosecutors, probation officers, and juvenile court judges and masters—must buy into the reform enough to change their behavior.[11] Moreover, the realization of *Gault*'s promised systemic reform depends largely on action by the legal profession. While the right to counsel was just one of the Court's constitutionally-mandated procedural reforms, its overall reform scheme placed faith in the provision of counsel to assure the implementation of the other constitutional rights it imposed.

The Court articulated the benefits of counsel for juveniles in terms of individual representation.[12] Counsel, the Court wrote, is required in juvenile cases "to cope with problems of law, to make skilled inquiry into the facts, to insist upon regularity of proceedings, and to ascertain whether [the client] has a defense and to prepare and submit it."[13] However, juvenile defenders are increasingly called upon to expand their role to include broader forms of advocacy aimed at reforming juvenile justice system practice and procedure.[14] This expansion of the juvenile defender's role occurs within the context of

---

8. *In re Gault*, 387 U.S. at 31–57. These include the right to notice of charges, the right to counsel, the right to confront and cross-examine prosecution witnesses, and the right to invoke the privilege against self-incrimination. *Id.*

9. *See id.* at 27–28.

10. *Id.* at 21, 26. Moreover, the Court viewed the introduction of procedural formality as a direct response to the misguided result in Gault's case. *See id.* at 21. In the Court's view, the procedural rules fashioned from the due process standard "are our best instruments for the distillation and evaluation of essential facts from the conflicting welter of data that life and our adversary methods present" and "enhance the possibility that truth will emerge from the confrontation of opposing versions and conflicting data." *Id.*

11. *See* Heidi M. Hsai & Marty Beyer, *System Change Through State Challenge Activities: Approaches and Products*, JUV. JUST. BULL. (U.S. Dep't of Justice/Office of Juvenile Justice and Delinquency Programs, Wash., D.C.), Mar. 2003, at 3, *available at* http://www.ncjrs. gov/pdffiles1/ojjdp/177625.pdf.

12. *In re Gault*, 387 U.S. at 36.

13. *Id.*

14. *See, e.g.*, ELIZABETH CALVIN, NAT'L JUVENILE DEFENDER CTR., LEGAL STRATEGIES TO REDUCE THE UNNECESSARY DETENTION OF CHILDREN 49–54 (2004), *available at* http://www.njdc.info/pdf/detention_guide.pdf.

juvenile justice reform programs and initiatives, which offer grant money and technical assistance to local jurisdictions for the purpose of addressing systemic problems.[15]

When a local juvenile court engages in a systemic reform program, juvenile defenders have the opportunity to expand their advocacy beyond the representation of juveniles in individual proceedings. Although the predominant systemic reform model of collaboration among juvenile justice stakeholders is seemingly counter to adversarial defense advocacy in individual representation, I argue that the stakeholder model is actually more strategic than it is made to appear. Hence, it provides opportunities for defense advocacy at the systemic level. However, to ensure that the promise of *Gault* is fulfilled through such advocacy, juvenile defenders must become conversant in the process of systemic reform and properly define their role within these initiatives.

In Part I of this essay, I argue that participation in a stakeholder collaboration reform is not necessarily antithetical to a defense role; rather, systemic reform advocacy is consistent with the juvenile defender's role as delineated in *Gault*. Part II suggests three basic "building blocks" of systemic reform advocacy that will facilitate juvenile defenders' reform efforts: (1) the ability to act as a spokesperson for the client's perspective, (2) the cultivation of an acute awareness of the underlying interests and incentives that shape the status quo, and (3) the capacity to frame proposed changes in ways that allow system participants to legitimate change. I argue that these "building blocks" are continuous with the juvenile defender's advocacy in individual cases, and that engaging as a stakeholder in systemic reform efforts can enhance individual advocacy. Finally, Part III shares strategies for use in clinical teaching to help students draw connections between advocacy in individual cases and advocacy for broader systemic reform.

I. SYSTEMIC REFORM ADVOCACY IN THE STAKEHOLDER COLLABORATION MODEL

The prevailing model according to which many current systemic reform initiatives proceed is what I will call the "stakeholder collaboration model."[16]

---

15. *See generally, e.g.*, DAVID STEINHART, THE ANNIE E. CASEY FOUND., PLANNING FOR JUVENILE DETENTION REFORMS: A STRUCTURED APPROACH (1999), *available at* http://www.aecf.org/upload/publicationfiles/planning%20for%20detention%20reforms.pdf (describing the Annie E. Casey Foundation's Juvenile Detention Alternative Initiative); Hsai & Beyer, *supra* note 11 (describing the initiatives of the Office of Juvenile Justice and Delinquency Prevention).

16. *See generally, e.g.*, STEINHART, *supra* note 15; Hsai & Beyer, *supra* note 11 (discussing the use of this model in juvenile justice system reform). The basic components of this model are gaining prevalence in settings beyond juvenile justice reform. *See generally, e.g.*, Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015 (2004) (describing a trend in the administration of consent

The stakeholder collaboration model promotes inter-agency collaboration and data-driven, research-based reforms to juvenile justice system practice and procedure.[17]

Taken at face value, the stakeholder collaboration model appears to be a value-neutral, consensus-based process that may threaten or undermine a juvenile defender's commitment to adversarial client advocacy in individual cases. Upon closer examination, however, the stakeholder collaboration model is most often employed strategically, not to gain consensus among stakeholders, but to influence system officials and line workers in favor of reforms with pre-determined values and goals that are typically consistent with those of juvenile defenders. As such, participation in a stakeholder collaboration allows juvenile defenders to utilize many of the same tools they use for advocacy in individual cases to shape policies and procedures on the systemic level.

This Part explores the process and methods of the collaborative stakeholder model by examining the Annie E. Casey Foundation's Juvenile Detention Alternatives Initiative (JDAI). The JDAI is a multijurisdictional reform network that began in 1992 by issuing planning grants to five sites that were interested in experimenting with systemic changes to reduce their juvenile detention populations.[18] As of October 2007, the JDAI has grown to encompass eighty-seven sites in twenty-one states and the District of Columbia.[19] In addition to being one of the most extensive juvenile justice systemic reform networks, it is also one of the best documented.[20] The Annie E. Casey Foundation has published a series entitled *Pathways to Juvenile Detention Reform*, which consists of fourteen separately authored reports that describe and analyze its approach to systemic reform, drawing lessons from the challenges and successes of the JDAI's participants.[21]

---

decrees in many areas of public law litigation that relies on stakeholder deliberation and experimental implementation of reform strategies whose success is measured in outcomes). I have previously examined the process of criminal justice reform in response to wrongful convictions, which also relies on inter-agency collaboration and scientifically-tested best practices. *See generally* Katherine R. Kruse, *Instituting Innocence Reform: Wisconsin's New Governance Experiment*, 2006 WIS. L. REV. 645 (2006).

17. *See, e.g.*, STEINHART, *supra* note 15, at 13–14; Hsai & Beyer, *supra* note 11, at 3–4.

18. Bart Lubow, *Series Preface* to STEINHART, *supra* note 15, at 4, 8 [hereinafter Lubow, *Series Preface*]. The same preface appears at the beginning of each of the fourteen reports in the Annie E. Casey Foundation's *Pathways to Juvenile Detention Reform* series.

19. *Announcing New JDAI Sites*, JDAI NEWS (Annie E. Casey Found./Juvenile Det. Alternatives Initiative, Balt., Md.), Oct. 2007, at 5, *available at* http://69.18.145.86/upload/PublicationFiles/JDAI%20News%20October%202007.pdf.

20. Bart Lubow, *Preface Update* to RICHARD A. MENDEL, THE ANNIE E. CASEY FOUND., BEYOND DETENTION: SYSTEM TRANSFORMATION THROUGH JUVENILE DETENTION REFORM 15 (2007), *available at* http://www.aecf.org/upload/PublicationFiles/JDAI_Pathways14.pdf [hereinafter Lubow, *Preface Update*].

21. This entire series can be found on the Annie E. Casey Foundation's website. *See* The Annie E. Casey Foundation, Pathways to Juvenile Detention Reform, http://www.aecf.org/KnowledgeCenter/PublicationsSeries/JDAIPathways.aspx (last visited Jan. 17, 2008).

A. The Stakeholder Collaboration Model

The starting assumption of stakeholder collaboration, as articulated in the *Pathways to Juvenile Detention Reform* series, is that the term "juvenile justice system" is something of a misnomer.[22] To call it a system suggests that it is a "complex whole" consisting of interacting, interdependent constituent parts.[23] More often, the juvenile justice system is characterized by a variety of agencies—police, prosecution, detention, probation, judges, and the defense bar—each of which are separately administered and act independently, with little understanding of the policies, procedures, or assumptions under which the other agencies proceed.[24] The stakeholder collaboration model is designed to facilitate understanding and coordination among the agencies that will deal with a juvenile defendant as a case unfolds.[25] The key components to reform are inter-agency collaboration, the rigorous collection and analysis of data from one's own jurisdiction, and the experimental implementation of innovative practices that have been successful in other jurisdictions.[26]

The keystone of the JDAI reform model is the process of collaboration, described as "the coming together of disparate juvenile justice system stakeholders and other potential partners (like schools, community groups, the mental health system) to confer, share information, develop system-wide policies, and to promote accountability."[27] Juvenile justice system agencies often have disparate cultures, as well as differing perspectives and attitudes about the treatment of juveniles.[28] The JDAI stakeholder collaboration model seeks to foster initial consensus among stakeholders that the "limited purposes of secure detention" are "to ensure that alleged delinquents appear in court at the proper times and to protect the community by minimizing serious delinquent acts while their cases are being processed."[29] It encourages stakeholders to agree on a specific plan for reform based on "an accurate description of the current system;" a description of the principles and values of the proposed reformed system; and "an action plan [with] carefully delineated .

---

22. KATHLEEN FEELY, THE ANNIE E. CASEY FOUND., COLLABORATION AND LEADERSHIP IN JUVENILE DETENTION REFORM 10 (1999), *available at* http://www.aecf.org/upload/PublicationFiles/collaboration%20and%20leadership.pdf.

23. *Id.*

24. *Id.*

25. *See* STEINHART, *supra* note 15, at 15–19; Hsia & Beyer, *supra* note 11, at 2–3, 6.

26. *See* STEINHART, *supra* note 15, at 13–14. The first two components of collaboration and data collection run throughout the JDAI series and are described in detail in separate reports. *See generally* DEBORAH BUSCH, THE ANNIE E. CASEY FOUND., BY THE NUMBERS: THE ROLE OF DATA AND INFORMATION IN DETENTION REFORM (1997) (discussing the use of data); FEELY, *supra* note 22 (discussing collaboration). The third is implicit in the Pathways series itself, which seeks to share the "innovations and lessons" of the initial program sites. Lubow, *Series Preface*, *supra* note 18, at 9.

27. Lubow, *Series Preface*, *supra* note 18, at 7.

28. FEELY, *supra* note 22, at 15.

29. *Id.* at 7.

. . time frames, budgets," and allocations of responsibility.[30] In order for the plan to be successful, the stakeholders involved must include policymakers from each of the primary agencies: "the judiciary, prosecution, defense, probation, detention, and related service providers."[31]

The second key to reform under the JDAI stakeholder collaboration model is data-driven decisionmaking. Each agency's approach toward the treatment of juveniles may be based on assumptions or rationalizations about juveniles, their families, or changing crime or arrest patterns that are founded in anecdote or impression rather than fact.[32] JDAI promotes a juvenile justice system's rigorous collection and analysis of information about its actual operation.[33] The collection of information includes quantitative data about arrests, as well as demographic information about who is detained, the types of charges upon which they are detained, daily bed counts, and the number of days spent in detention for various types of children and cases.[34] Data analysis then becomes integral to identifying the issues that should be addressed by targeted reforms.[35] For example, in Cook County, Illinois, data analysis showed that children accused only of violating the conditions of their probation were being held for an average stay of 28 days in detention; thus, reform efforts focused on the handling of these violations.[36] Furthermore, before a reform is initiated, data can be used to project the anticipated effects of a changed policy or procedure.[37] After a reform is initiated, data can be used to test the impact of policy or procedural changes.[38] JDAI jurisdictions break down all data by race and sex to help identify the disproportionate effects of juvenile detention policy on ethnic minorities and girls.[39]

---

30. *Id.* at 31.

31. *Id.* at 22.

32. BUSCH, *supra* note 26, at 10–11; FEELY, *supra* note 22, at 17 ("In the past, anecdote and 'fingertip knowledge' have guided change in juvenile justice systems").

33. *See* STEINHART, supra note 15, at 20–28.

34. *Id.* The collection additionally contains a "systems analysis" of flow charts to map out the way juveniles move through various agencies in the system, *id.* at 28–32, a "conditions analysis" of the legal regulations that govern decision-making at different stages, *id.* at 32–35, and a "cost analysis" of secure detention beds and community-based alternatives, *id.* at 35–36.

35. *Id.* at 41–42; *see also* BUSCH, *supra* note 26, at 14 (discussing the effects of "the power of data.")

36. STEINHART, *supra* note 15, at 42.

37. *See id.* at 54–56 (describing JDAI sites' use of a methodology developed by the National Council on Crime and Delinquency for projecting the use of detention bed space under one or more reform scenarios).

38. *See* FRANK ORLANDO, THE ANNIE E. CASEY FOUND., CONTROLLING THE FRONT GATES: EFFECTIVE ADMISSIONS POLICIES AND PRACTICES 36–38 (1999), *available at* http://www.aecf.org/upload/PublicationFiles/controlling%20front%20gates.pdf (describing the use of data to validate risk assessment instruments used in making initial detention decisions).

39. *See generally* ELEANOR HINTON HOYTT ET AL., THE ANNIE E. CASEY FOUND., REDUCING RACIAL DISPARITIES IN JUVENILE DETENTION (2001), *available at* http://www.aecf.org/upload/PublicationFiles/reducing%20racial%20disparities.pdf (discussing

Finally, JDAI provides technical assistance by fostering communication among the jurisdictions that have undergone JDAI reform. The desire to share strategies for success was the initial impetus for JDAI, which began as an effort to replicate detention reforms funded by the Annie E. Casey Foundation in Broward County, Florida.[40] In 1993, JDAI launched five sites with three-year planning grants.[41] By 1998, three of these sites—Cook County, Illinois; Multnomah County, Oregon; and Sacramento County, California—had engaged in "fundamental, system-wide changes" and "absorbed the JDAI innovations into their regular juvenile justice budgets and procedures."[42] These three sites became JDAI model sites and were willing to be "nagged, measured, scrutinized and endlessly visited" by sites seeking to replicate their successful detention reform innovations.[43] As it has grown in size and scope, JDAI has spawned an elaborate network of information-sharing. This network hosts national all-site conferences, produces JDAI newsletters with articles detailing projects at various JDAI sites, and maintains a web-based Help Desk on which jurisdictions can post their policies, procedures, or other innovations as examples for other sites undertaking reform.[44]

More than a means to reform, proponents of the JDAI model view stakeholder collaboration itself as a reform. The JDAI model assumes that the habits of inter-agency collaboration and data-driven decisionmaking will take hold in juvenile justice systems and continue to thrive even after grant funding has ceased. Small victories will build confidence in the process of collaborative, data-driven decisionmaking; this confidence will generate collaborative efforts that go beyond detention reform to transform juvenile justice systems more broadly.[45]

---

the problem of racial disparities in juvenile detention and the efforts to reduce such disparities); FRANCINE T. SHERMAN, ANNIE E. CASEY FOUND., DETENTION REFORM AND GIRLS: CHALLENGES AND SOLUTIONS (2005), *available at* http://www.aecf.org/upload/PublicationFiles/jdai_pathways_girls.pdf (examining the upward trend in the juvenile detention system's female population and the social problems related to this trend).

40. ROCHELLE STANFIELD, THE ANNIE E. CASEY FOUND., THE JDAI STORY: BUILDING A BETTER JUVENILE DETENTION SYSTEM 8 (1999), *available at* http://www.aecf.org/upload/PublicationFiles/jdai%20story.pdf.

41. *Id.*

42. *Id.*

43. Douglas W. Nelson, President of The Annie E. Casey Found., Remarks at the Juvenile Detention Alternatives Initiative Inter-Site Conference (April 10, 2003), *in* JDAI NEWSL. (The Annie E. Casey Found./Juvenile Det. Alternatives Initiative, Balt., Md.), Nov. 2003, at 10, 11, *available at* http://www.aecf.org/upload/PublicationFiles/november2003.pdf. Bernalillo County, New Mexico was added as a fourth model site in 2005. Bart Lubow, *From the Foundation*, JDAI NEWS (The Annie E. Casey Found./Juvenile Det. Alternatives Initiative, Balt., Md.), Winter 2005, at 2, *available at* http://www.aecf.org/upload/PublicationFiles/march2005.pdf.

44. Lubow, *Preface Update*, *supra* note 20, at 10–11.

45. *See generally* MENDEL, *supra* note 20 (discussing the JDAI's widespread work to effect systemic changes).

B. The Role of the Juvenile Defender in Stakeholder Collaboration

The stakeholder collaboration model is attractive in principle. It resonates with theories of good governance and deliberative democracy. It suggests that well-intentioned public officials can come together, agree on the goals of the system, use empirical methods to identify problems with the system's operation, and implement comprehensive changes to policy and procedure based on that data.

However, reform under the stakeholder collaboration model is not as simple as it sounds. Stakeholder collaboration in initiatives like the JDAI seek advocacy to reform a local juvenile justice system according to pre-determined values, rather than to reach a true consensus among stakeholders about the values upon which the system ought to be based. Despite the stated focus on collaboration, efforts such as the JDAI's are not value neutral. For example, although the JDAI recognizes that various jurisdictions may articulate the goals of their juvenile detention systems differently, the JDAI literature does not suggest that stakeholders deliberate about the underlying values served by juvenile detention reform.[46] Nor does the JDAI literature entertain the possibility that stakeholders might reach consensus on goals that run counter to JDAI values—for example, that detention can be legitimately used to "teach children a lesson" or that a jurisdiction should hold children in secure detention for their own protection.[47] Rather, the discussions of stakeholder consensus focus on ways to advocate and persuade system officials and line workers with different ideologies to buy into the JDAI values.[48] These values include limiting the use of detention to children who need to be in secure custody to protect the public and promoting community-based alternatives for children who do not fit that description.[49]

Juvenile defenders are awkwardly situated for participation in the stakeholder collaboration model. As the JDAI literature demonstrates, reform efforts can easily be stymied by resistance from agency officials or employee unions, or hindered by mutual distrust among agencies.[50] When it comes to specific reform proposals, juvenile defenders can quickly find themselves

---

46. *See, e.g.*, STEINHART, *supra* note 15, at 37–38. On the one hand, "values and attitudes" will differ from one local jurisdiction to another. *Id.* at 37. However, on the other hand, the "goal-setting process should include discussion and self-education by planners on the legal and constitutional purposes of secure juvenile detention" and "the use of secure detention for purposes beyond protection of the public or prevention of flight is highly suspicious." *Id.*

47. *See* ORLANDO, *supra* note 38, at 10–12. In fact, such goals are considered illegitimate from the JDAI perspective. *Id.*

48. *See generally* ROBERT G. SCHWARTZ, THE ANNIE E. CASEY FOUND., PROMOTING AND SUSTAINING DETENTION REFORMS (2001), *available at* http://www.aecf.org/upload/PublicationFiles/promoting%20sustaining%20reforms.pdf (discussing JDAI sites' promotion of reforms).

49. Lubow, *Series Preface*, *supra* note 18, at 7.

50. STEINHART, *supra* note 15, at 59–62 (discussing barriers to reform).

resistant to collaboration, because defense sensibilities about matters such as procedural formality and expedient case processing are likely to differ from the perspectives of rival agencies.[51] Additionally, collaboration with prosecutors and probation departments over new programs and procedures may threaten juvenile defenders' more traditionally adversarial role in individual cases. For example, defenders' agreement to target particular diversion programs or services toward a special population of youth may restrain their ability to advocate for those services in exceptional cases falling outside that group.[52]

However, the values that the JDAI assumes as its starting point—centering on the idea that juvenile detention should be used sparingly, in cases where detention is necessary to protect the public while juvenile cases proceed—coincide with the values and perspectives of juvenile defenders. The JDAI's suggested reforms have also met resistance since their inception from political forces intent on being tough on juvenile crime.[53] Through participation in stakeholder collaboration, juvenile defenders can support systemic changes designed to facilitate the kind of treatment they often advocate in individual cases: carefully tailored interventions in the lives of children and families that favor community placement over institutional custody. If successful, JDAI reforms are likely to create a system more hospitable to the clients that juvenile defenders represent.

To navigate the waters of systemic reform while retaining their role as advocates, juvenile defenders need to view participation in collaborative stakeholder ventures as a form of advocacy on the systemic level. Furthermore, to become sophisticated advocates on the systemic level, juvenile defenders must develop a vocabulary for discussing strategic engagement in stakeholder collaboration reform efforts, just as they have developed ways of discussing what it means to be effective advocates for individual clients. While the project of fully developing the advocacy role and strategies for effective advocacy within collaborative reform efforts is beyond the scope of this essay, the following Part describes three building blocks of systemic reform advocacy for juvenile defenders.

II. BUILDING BLOCKS OF SYSTEMIC REFORM ADVOCACY

This Part describes three principles of advocacy that juvenile defenders can employ in systemic reform efforts under a stakeholder collaboration model. These principles are: (1) understanding the system from the perspective of the client, (2) analyzing the status quo as an accommodation of competing interests, and (3) listening for the narratives of system officials that legitimate their

51. *See infra* Part II.A. (discussing case processing reforms).

52. FEELY, *supra* note 22, at 38.

53. *See* STANFIELD, *supra* note 40, at 6–8 ("The JDAI effort to reduce the numbers of confined youth went against a popular tide of mounting arrests and skyrocketing detentions."); Lubow, *Series Preface*, *supra* note 18, at 8 (noting the shift toward stricter juvenile justice policies that occurred after a string of high-profile cases in the 1990s).

behavior and place blame for systemic problems elsewhere. Rather than being incongruent with individual advocacy, these building blocks are extensions of the same advocacy tools that defenders already employ in their representation of individual clients.

A. Developing a View of the System From the Client's Perspective

One of the defining features of juvenile defender advocacy at the systemic level is an extension of the traditional role of juvenile defenders to provide a voice for their clients. In their traditional role as individual client advocates, juvenile defenders act as spokespersons for children, whose voices would otherwise go unheard.[54] If the juvenile justice system is indeed an uncoordinated "non-system" of separately governed and administered agencies, the one place where these agencies converge is the clients into whose lives the juvenile justice system intervenes. Clients bear the burdens of the lack of coordination between agencies—burdens that may negatively affect their education, their family and community relationships, and their receipt of services.

Participation in the stakeholder collaboration process gives juvenile defenders the opportunity to observe and articulate the burdens that the system places on their clients. As the JDAI literature notes, more than other stakeholders, juvenile defenders are concerned with whether a proposed reform will be best for the children, rather than with the savings in time or dollars that specified reforms may achieve.[55] As JDAI proponents acknowledge, when juvenile defenders are marginalized in stakeholder collaboration, debates over changes to juvenile systems and policies can lose the perspectives of the children accused.[56]

One of the most important areas in which juvenile defenders need to be heard is in discussions about case processing reforms. JDAI reform focuses on ways to "streamline the processing of cases" and "recommend changes in case processing that can accelerate the movement of cases and reduce stays in detention."[57] Although admirable, the goal of efficient case processing can run counter to the values promoted by adversary adjudication.[58] If juvenile

---

54. *See generally* Annette R. Appell, *Children's Voice and Justice: Lawyering for Children in the Twenty-First Century*, 6 NEV. L.J. 692 (2006) (analyzing the various ways in which legal advocates give voice to children).

55. D. ALAN HENRY, THE ANNIE E. CASEY FOUND., REDUCING UNNECESSARY DELAY: INNOVATIONS IN CASE PROCESSING 37 (1999), *available at* http://www.aecf.org/upload/PublicationFiles/reducing%20unnecessary%20delay.pdf. JDAI sites' experience suggests that the defense perspective is critical to success in case processing reform. *Id.* ("Although other system participants may focus on time and dollar savings, the defense is ambivalent about the first and has little interest in the second.").

56. FEELY, *supra* note 22, at 37.

57. STEINHART, *supra* note 15, at 13–14.

58. *See* HENRY, *supra* note 55, at 14 (noting the "pitfall of speeding up case processing as

defenders are to carry out the role specified for them in *Gault*—"to make skilled inquiry into the facts . . . [and] to ascertain whether [the client] has a defense and to prepare and submit it"[59]—then they need time to investigate. They also need time to develop relationships with their clients that allow the clients to make informed decisions about whether to take a case to trial or accept a plea.[60] As the stakeholders aim to reach consensus on ways to expedite case processing for children in detention, the procedural formality required by *Gault* may be perceived as an impediment, rather than an aid, to the JDAI's goals. Defenders can help to bring the voice of children back into these discussions, explaining the difficulties of attempting to effectively counsel an adolescent client at a juvenile detention hall who "just wants to go home" about the benefits and detriments of accepting a plea offer.

Yet, to effectively play the role of client spokesperson in systemic reform efforts, juvenile defenders must understand their clients' lives more broadly than required by traditional representation.[61] The burdens of heavy caseloads often limit juvenile defenders' ability to understand their clients within the broader context of their neighborhoods, schools and families. Moreover, if representation ends with disposition, juvenile defenders do not get a chance to evaluate how or whether the system is delivering on its promise of effective, well-tailored, individualized treatment, rather than punishment. Participation in collaborative reform efforts may draw juvenile defenders' attention to the impact of the system on their clients in ways that individual advocacy does not reveal; in turn, this more contextual conception of individual clients may enhance juvenile defenders' ability to voice clients' perspectives on a broader range of systemic issues.

B. Analyzing the Status Quo as an Accommodation of Competing Interests

As advocates for individual clients, juvenile defenders are often in the position of insisting on procedural regularity—holding system officials to statutory and constitutional requirements in the face of their competing desires to protect the public and act in the perceived best interests of children. To effectively insist on procedural regularity in individual cases, juvenile defenders must understand what the rules require and what the decisionmaker will be inclined to do in a particular case, and must strategically decide when to demand and when to cajole.

---

an end in itself" without attention to improving the justice of proceedings).

59. *In re* Gault, 387 U.S. 1, 36 (1966).

60. *See generally* Abbe Smith, *Defending and Despairing: The Agony of Juvenile Defense*, 6 NEV. L.J. 1127 (2006) (describing the complex relationship between a particular juvenile defender and her client).

61. *See generally Recommendations of the UNLV Conference on Representing Children in Families: Child Advocacy and Justice Ten Years After* Fordham, 6 NEV. L.J. 592 (2006) (recommending and describing broader, more holistic views of juvenile defense representation).

Despite the veneer of consensus that cloaks stakeholder collaboration reform, similar strategic processes govern juvenile defender advocacy at the systemic level. Writing nearly thirty years ago, Raymond Nimmer analyzed the process of systemic change and noted that reformers too often approach reform with the naïve assumption that system participants desire change.[62] The efforts of such reformers, he observed, are often "characterized by a failure to distinguish between the substance and the tactics of reform."[63] Reformers focus on removing obstacles to the ideal operation of a system, overlooking the fact that the status quo reflects an accommodation of competing interests and incentives that may have little to do with the ideals on which the system is based.[64] Unless the reformer can understand and address the underlying interests that shape the status quo, Nimmer argued, the reform effort is likely to flounder in the face of systemic resistance to change.[65]

In the JDAI process, the importance of addressing underlying interests and incentives is reflected in the creation of objective "risk assessment instruments" (RAIs), which are used in making initial detention decisions. A core strategy of the JDAI is to replace subjective discretionary decisionmaking concerning juvenile detention with objective criteria.[66] Systems applying RAIs make initial detention decisions by assigning points to various objective criteria, including the seriousness of the offense for which a child is arrested, the child's prior history of arrest or adjudication, and any history of failure to appear in court.[67] Children with a low point score are presumptively released, children with a mid-range score are considered for detention alternatives, and only children with a high score are presumptively detained pending further proceedings.[68]

To effectuate change, an RAI must be respected by system officials and implemented by line workers. Yet, if Nimmer is correct, the behavior of these parties is often influenced by underlying interests and incentives that may run counter to the limited goals of detention. For example, police agencies and actors in the child welfare, mental health, or school systems may use detention facilities as a dumping ground of last resort for children who are either too difficult to manage or who have nowhere else to go.[69] Detention officials or

62. NIMMER, *supra* note 7, at 1–2.

63. *Id.* at 2.

64. *See id.* at 176–77. Professor Nimmer argues that the "basic fallacy" of reform planning is that it assumes the incentive for change exists and focuses only on removing barriers to the desired behavior, rather than viewing the status quo as an accommodation of competing interests in which each system participant derives a benefit. *Id.*

65. *Id.* at 177.

66. The Annie E. Casey Foundation, Juvenile Detention Alternatives Initiative: Core Strategies, http://www.aecf.org/MajorInitiatives/JuvenileDetentionAlternativesInitiative/CoreStrategies.aspx (last visited Jan. 17, 2008) (listing the core strategies).

67. ORLANDO, *supra* note 38, at 24–27.

68. *Id.* at 27.

69. *Id.* at 11; *see also id.* at 20–22 (discussing the dilemma of police using detention as a holding place for children with no other place to go).

line staff who accept children under such circumstances may prescribe detention to "teach [children] a lesson,"[70] to provide access to services, or protect children from themselves—goals that contradict the values of limited detention endorsed by the JDAI. Furthermore, long-standing relationships between detention staff members and law enforcement or field probation agents may be disrupted when detention staff members are asked to turn children away because the children fail to meet objective criteria.[71]

Moreover, juvenile court personnel may have their own interests and incentives to keep children detained when the limited purposes of detention do not apply. For example, judges, as publicly accountable officials, may seek to avoid releasing individuals who will make the headlines; thus, their detention and release decisions may reflect risk-averseness, rather than adherence to objective detention criteria.[72] Although ostensibly adversarial, prosecutors and defense attorneys may subtly collude in a system in which offenders are routinely overcharged and then charges are reduced through plea bargaining, because such a system portrays the prosecutor as tough on crime while allowing defense attorneys to claim that they have secured a benefit for their clients through negotiation.

As evidenced by the JDAI's experience with RAIs, in order to effectuate meaningful changes, reforms must appeal not only to logic and good public policy, but also to the existing incentives and interests of system participants. Logic and good public policy dictate that the objective criteria institutionalized in the RAI should closely track the limited purposes of detention endorsed by the JDAI: "(1) to ensure the alleged delinquents appear in court and (2) to minimize the risk of serious reoffending while current charges are being adjudicated."[73] In New York, an independent professional nonprofit agency was consulted to prepare an RAI that scored arrested youths based on objective factors that were known to bear a statistically significant relationship to the risk of re-offense or failure to appear in court.[74] However, the research-based criteria were so far removed from established practice that the criteria were simply ignored by the line staff, who implemented the RAI, and judges, who decided whether to release detained children pending further proceedings.[75]

---

70. *Id.* at 11.

71. *See id.* at 18 (discussing difficulties in implementing statutory criteria in Florida that unsettled common practice by limiting detention). In one incident, "a very frustrated sheriff's deputy tried to arrest an intake supervisor for obstruction of justice after the supervisor refused to allow the deputy to leave a particular youth (charged only with a traffic offense) at the detention center." *Id.*

72. *See id.* at 23–24 (noting the political vulnerability of detention criteria in light of elections and highly publicized cases); STANFIELD, *supra* note 40, at 18 (quoting a judge regarding his awareness of the potential for public reaction to his decision-making).

73. ORLANDO, *supra* note 38, at 10.

74. *Id.* at 29.

75. SCHWARTZ, *supra* note 48, at 20–21 (contrasting the resistance of line staff to the New York instrument with the acceptance of an RAI in Multnomah County, Oregon, where members

Moreover, the RAI is not meant to be inflexible; to this end, the RAIs of many jurisdictions allow line staff within the detention center to override the presumptive determination based on the listed criteria.[76] If an RAI's objective criteria do not reflect the detention officials' subjective understanding about which children ought to be detained, the use of such overrides is likely to become prevalent.[77] This was the experience in Multnomah County, Oregon, where the RAI was carefully structured to prevent the detention of certain targeted populations based on data about who was being incarcerated; in that case, "the initial reliance on data perhaps helped obscure some lack of consensus" among system officials about who should be detained.[78]

Most JDAI jurisdictions opt for the much less rigorous "normative" method of developing their RAIs, which involves looking to the RAIs used in other jurisdictions and adjusting them to local conditions through the work of a committee comprised of representative stakeholders.[79] Rather than changing common practice, this normative method may simply codify the attitudes and preconceived notions of probation agents, judges, and prosecutors about which children may pose a risk to the public or themselves pending court proceedings. Moreover, if the RAI is designed to include as many perspectives as possible, it may produce more risk-averse detention decisions than occurred prior to the RAI. For example, after Cook County, Illinois, developed an RAI according to the normative method, its rate of detention increased markedly.[80] The RAI was revised by projecting the effects of various criteria changes on re-arrests and failures to appear in court; Cook County then conducted a three-month test of the revised criteria in a sample group of cases while the original RAI remained in effect.[81] Though it met with some political resistance, the process of testing the original RAI against the new one helped move Cook County's RAI toward the JDAI's limited goals of detention.[82]

As the JDAI experience developing and implementing RAIs demonstrates, the goal of producing an RAI is in some ways less important than a commitment to the process of continual revision and testing.[83] Furthermore, this continual revision is not just a matter of fine-tuning. Rather, it is a process of education, persuasion, and accommodation that aims to identify and address the systemic obstacles to realization of the identified objectives.

The role of juvenile defenders in this process of evaluation and revision is consistent with their role in individual case representation: to ensure that the

---

of the line staff involved in creating the detention criteria).

76. ORLANDO, *supra* note 38, at 27.

77. *See id.* at 14, 27 (noting the need to monitor use of overrides in an RAI system).

78. BUSCH, *supra* note 26, at 15.

79. ORLANDO, *supra* note 38, at 29.

80. *Id.* at 31.

81. *Id.* at 37.

82. *See generally id.* (describing the sometimes unwelcome process of field testing draft RAIs).

83. *Id.* at 32.

system follows legal mandates, even when those mandates are unpopular. As the JDAI literature notes, there is no natural constituency to advocate for changing or limiting the purposes of pre-adjudication detention.[84] Although an effective RAI must reflect the attitudes and assumptions of system workers and officials, there must also be stakeholders who are willing to challenge the comfort levels of these workers and officials (who often have interests in maintaining the status quo)[85] and insist on compliance with the limited purposes of detention. In individual cases, particularly those involving suppression motions, defenders regularly challenge the comfort levels of criminal justice agents and officials in order to enforce the system's compliance with constitutional requirements. The role of the stakeholder who questions the RAI's compliance with the limited legal purposes of pre-adjudication detention is the systemic reform analogue to the juvenile defender's advocacy role.

Juvenile defenders' engagement and familiarity with the process of RAI development can also enhance their advocacy for release of clients in individual cases. When detention decisions are made on the basis of objective factors that must be scored in each case, the ostensible justification for detaining a particular client is apparent. Yet, this process also opens detention decisions to challenge where the analysis was based on erroneous information or assumptions. For example, a probable cause affidavit stating that a crime was committed by the accused may be unassailable as a basis for arrest. However, a client might receive a high point score because the facts alleged in the affidavit were stretched in an attempt to overcharge the conduct—for example, where the affidavit states that a robbery occurred, but the factual allegations more properly indicate the occurrence of a petty larceny. In that case, a juvenile defender can argue that the client should be released because, if not for the overcharging, the client would not have been detained under the RAI in the first place.[86]

C. Interpreting the Narratives of Legitimation and Blame

People usually like to believe that they are engaged in meaningful work, especially when that work affects the lives of children. As a result, most juvenile justice system participants believe that their behavior reflects the values of the system—that they are helping children and protecting the public. To the extent that participants acknowledge problems in the system, they tend to attribute those problems to other agencies. Thus, a "circle of blame" is

84. *Id.* at 40.

85. *See supra* notes 69–72 and accompanying text.

86. In some JDAI jurisdictions, an "expediter" may be charged with monitoring compliance in individual cases and changing the RAI score when new facts come to light or when charges are reduced. ORLANDO, *supra* note 38, at 34–35. If no expediter position exists, however, it falls naturally to the juvenile defender to monitor and raise compliance issues on a case-by-case basis.

created, in which each agency acknowledges the same systemic problems but acquits itself and indicts others for causing the problems.[87]

Collaboration and data-driven decisionmaking can help system participants tackle the circle of blame; however, these tactics alone may not provide sufficient motivation to break out of it. To break out of the circle of blame, system officials must realize that there exists a manageable solution that they can take a heroic role in implementing. Once such a solution is in hand, it becomes acceptable to define the problem differently. Rather than having to bear the blame for a seemingly insoluble problem, publicly accountable system officials can position themselves as the solution to the problem.[88]

To work effectively within a systemic reform effort, reformers must understand these dynamics of system change and develop advocacy tools and strategies based on narratives that deflect blame and embrace changed policies and behavior. The deployment of such narratives that depict decisionmakers such as juries as heroes is often the essence of effective defense advocacy.[89] When the stakeholder collaboration model is understood to be an elaborate arena of persuasion and advocacy, the tools of persuasion can be usefully integrated into systemic reform advocacy.

III. TEACHING WITH AN EYE TOWARD SYSTEMIC REFORM

As I continue to educate myself about systems and systemic reform, I have begun to incorporate some of this information into my Juvenile Justice Clinic teaching. This section describes two such efforts.

A. Embedding Students Within the System

Experiential learning is often divided between a live-client clinic model, in which students take a low caseload and work on individual cases, and an externship model, in which students are placed within an existing legal structure such as a governmental office or a court and are guided in their work

87. *See generally* Robert L. Nelson, *The Discovery Process as a Circle of Blame: Institutional, Professional, and Socio-Economic Factors That Contribute to Unreasonable, Inefficient, and Amoral Behavior in Corporate Litigation*, 67 FORDHAM L. REV. 773 (1998) (analyzing this phenomenon with respect to civil discovery abuse).

88. I owe this insight to Michael Smith, who utilized this theory in his work as Director of the Vera Institute of Justice. *See generally* VERA INSTITUTE OF JUSTICE, TWENTY-FIVE YEAR REPORT FROM THE VERA INSTITUTE OF JUSTICE, 1961–1986 (1986) (describing Smith's work).

89. *See, e.g.*, Anthony G. Amsterdam & Randy Hertz, *An Analysis of Closing Arguments to a Jury*, 37 N.Y.L. SCH. L. REV. 55, 64–67, 97 (1992) (describing a defense attorney's use of "classic heroic verbs with the jury as the subject" in his closing argument); Philip N. Meyer, *Making the Narrative Move: Observations Based Upon Reading Gerry Spence's Closing Argument in* The Estate of Karen Silkwood v. Kerr-McGee, Inc., 9 CLINICAL L. REV. 229, 242 (2002) (expressing agreement with Amsterdam and Hertz's interpretation of the same argument).

in that setting. About three years ago, the Juvenile Justice Clinic at the William S. Boyd School of Law, University of Nevada, began to blend these models by incorporating a Juvenile Public Defender shadowing component into our live-client clinic. The JPD shadowing was necessitated in part by our clinic structure; it was a new clinic, from a new law school, in a jurisdiction with a strong local culture of unwritten rules.[90] Shadowing the juvenile public defenders introduced the clinic students to practice in the juvenile court environment in ways that individual case representation could not and gave them ready access to a knowledge bank of unwritten rules that the clinic did not possess.

When the clinic first incorporated Juvenile Public Defender shadowing, I was skeptical because I feared that students would learn to simply replicate the less-than-optimal practices they observed. However, I found the opposite to be true in practice. Students were able to learn valuable lessons from their Juvenile Public Defender mentors, who also aspire to higher standards of practice than the norm in our juvenile court, while maintaining a critical perspective on what they observed in the system itself.[91] Students are required to write reflective memos about what they see in their shadowing, and we discuss these memos in weekly supervision meetings and in seminar classes. As my own awareness of the dynamics of systems as a whole has grown, I have been able to draw attention to the interests and incentives of system participants into these discussions.

B. Case Debriefing and Analysis

It is always difficult to comprehensively evaluate students' work on a case that was ultimately unsuccessful. Because students feel the weight of personal responsibility for their clients, my initial tendency is to emphasize their strengths and reassure them that they did the best that they could. After the initial discussion, however, I conduct a more focused supervision session to ascertain what the student wishes he or she had done differently.[92]

Last semester, when a student lost a contested hearing, I tried something different, which incorporated both individual feedback and systemic analysis. In the seminar class following the contested hearing, we made a chart on the

---

90. *See generally* Katherine R. Kruse, *Standing in Babylon, Looking Toward Zion*, 6 NEV. L.J. 1315 (2006) (providing an in-depth description the clinic's introduction to the legal culture of Clark County, Nevada).

91. The Clark County juvenile defender office recently underwent a massive change in response to a 2003 audit by the National Legal Aid and Defender Association and a threatened lawsuit. *Id.* at 1324–31. Nonetheless, despite their commitment to professionalism, the juvenile public defenders confront a strong local culture of informality that stems from years of low trial rates and almost nonexistent motion practice throughout the adult and juvenile systems. *Id.*

92. *See generally* Beryl Blaustone, *Teaching Law Students to Self-Critique and to Develop Critical Clinical Self-Awareness in Performance*, 13 CLINICAL L. REV. 143 (2006) (describing an extremely effective model for delivering feedback).

whiteboard with the following five headings: (1) "What went right?" (2) "What went wrong?" (3) "Why?" (4) "What could we have done differently in individual advocacy?" (5) "What requires systemic reform?" The clinic students in the class—many of whom had either attended the contested hearing or been intimately involved in helping to moot and prepare it—brainstormed about the information that fit into each column, effectively covering all of the individual feedback I wanted the student to receive. By basing the discussion of individual representation within the context of systemic change, the analysis reinforced that one must keep in mind the limits of what is possible to achieve through individual advocacy when attempting to learn from one's mistakes. The discussion also helped students to spot the systemic reform issues that need attention in our jurisdiction.

I am still just beginning to understand the dynamics of systems and the process of systemic reform. But the more I have seen of it, the more convinced I have become that a rich understanding of system dynamics is a necessary component both for juvenile defender advocacy and for clinical education, as we consider how to move forward in fulfilling the promise of *Gault*.

THE LAW SCHOOL CLINIC AS A PARTNER IN A MEDICAL-LEGAL PARTNERSHIP

JANE R. WETTACH*

I. INTRODUCTION

In today's complex and interconnected society, lawyers undeniably must possess the ability to solve problems in an interdisciplinary context. Lawyers in varying practice areas need to understand aspects of technical fields such as medicine, science, international trade, and banking to represent their clients effectively. The development of this cross-disciplinary problem-solving ability has increasingly become part of clinical legal education.[1] The pediatric medical-legal partnership, a recently developed model for offering legal services to low-income clients, provides a creative opportunity for clinics focused on children to teach interdisciplinary skills along with the more traditional skills taught in the law school clinic.

A medical-legal partnership incorporates attorneys into the health care team of a needy patient.[2] It is based on the premise that social and other non-medical

---

* Clinical Professor of Law and Director of the Children's Law Clinic, Duke University School of Law. B.A., University of North Carolina at Chapel Hill, 1976; J.D., University of North Carolina School of Law, 1981. The Children's Law Clinic began the Medical-Legal Partnership for Children in Durham in 2007, in collaboration with Duke Primary Care for Children (an outpatient practice of Duke University Health System), the Lincoln Community Health Center (a federally-qualified health clinic), and Legal Aid of North Carolina.

1. For example, in 2003, 2004, and 2005, Washington University School of Law's Clinical Education Program and Center for Interdisciplinary Studies hosted three national conferences on interdisciplinary teaching and practice. *See* Symposium, *Justice, Ethics, and Interdisciplinary Teaching and Practice*, 14 WASH. U. J.L. & POL'Y 1 (2004); Symposium, *Poverty, Justice, and Community Lawyering: Interdisciplinary and Clinical Perspectives*, 20 WASH. U. J.L. & POL'Y 1 (2006); Symposium, *Promoting Justice Through Interdisciplinary Teaching, Practice, and Scholarship*, 11 WASH. U. J.L. & POL'Y 1 (2003). Several others have recently written about this topic as well. *See generally* Alexis Anderson, Lynn Barenberg & Paul R. Tremblay, *Professional Ethics in Interdisciplinary Collaboratives: Zeal, Paternalism and Mandated Reporting*, 13 CLINICAL L. REV. 659 (2007) (addressing problems that arise in the interdisciplinary practice of law); Bruce J. Winick & David B. Wexler, *The Use of Therapeutic Jurisprudence in Law School Clinical Education: Transforming the Criminal Law Clinic*, 13 CLINICAL L. REV. 605 (2006) (discussing the "creative application of behavioral science research to the legal context"); Christina A. Zawisza & Adela Beckerman, *Two Heads Are Better than One: The Case-Based Rationale for Dual Disciplinary Teaching in Child Advocacy Clinics*, 7 FLA. COASTAL L. REV. 631 (2006) (collecting articles on and arguing for interdisciplinary collaborations in child advocacy).

2. Barry Zuckerman, Megan Sandel, Lauren Smith & Ellen Lawton, *Why Pediatricians Need Lawyers to Keep Children Healthy*, 114 PEDIATRICS 224, 225 (2004).

factors influence the development of childhood disease.[3] Health care professionals are insufficiently equipped to respond to such factors, but lawyers possess the skills that can help resolve some of these non-medical obstacles to a child's health.[4] For example, if a child's chronic asthma is exacerbated by mold or other toxins in his apartment, a lawyer can take action against a recalcitrant landlord in a way that a pediatric nurse cannot.[5] If a child's application for government benefits to stabilize income or health coverage is denied, an attorney is needed to represent the child in the appeal process.[6] If a child with a mental disability is not receiving appropriate support in school and is thus spiraling downward, an attorney can intervene to navigate the special education system.[7]

While most of the seventy or so pediatric medical-legal partnerships in the United States are partnerships between hospitals and local legal aid offices, a number of them involve law school clinics.[8] Although these collaborations between law schools and medical practices present some challenges,[9] they also offer rich educational opportunities for students to engage in interdisciplinary lawyering that focuses on the holistic needs of their child clients. This Essay will describe the model of a medical-legal partnership in detail and analyze the benefits of this model for a law school children's clinic, particularly a clinic focusing on education and government benefits.

II. THE MEDICAL-LEGAL PARTNERSHIP MODEL

The inspiration for a medical-legal collaboration to benefit children came from Dr. Barry Zuckerman, a pediatrician at Boston Medical Center (BMC), who was constantly frustrated with his own limitations in solving the health problems of the children he treated.[10] He treated patients for malnutrition at the

---

3. *Id.* at 224–25.

4. *Id.* at 225.

5. *See id.* at 224.

6. *See id.* at 226.

7. *See id.* at 225.

8. *See* The Medical-Legal Partnership for Children, Partnership Sites, http://www.mlpforchildren.org/partnershipsites.aspx (last visited Mar. 6, 2008) [hereinafter MLPC, Partnership Sites] (listing existing medical-legal partnerships). Law schools with clinical programs or externship programs connected with a medical-legal partnership include Duke Law School, the University of Michigan Law School, the University of New Mexico School of Law, Roger Williams University Law School, University of Iowa College of Law, Albany Law School, Vanderbilt Law School, Syracuse University College of Law, Georgia State University College of Law, the University of Connecticut School of Law, and the University of Virginia Law School. *Id.*

9. *See* Zuckerman et al., *supra* note 2, at 226–27 (discussing barriers to collaboration between lawyers and health care professionals).

10. *See* Ellen M. Lawton, *The Family Advocacy Program: A Medical-Legal Collaborative to Promote Child Health & Development*, MGMT. INFO. EXCHANGE J., Summer 2003, at 12, 12.

same time the child's parents had been denied food stamps.[11] He treated patients for asthma who lived in squalid rental housing.[12] He treated patients with Attention Deficit Hyperactivity Disorder (ADHD) who were unable to obtain special education services at school.[13] Ultimately recognizing that his patients were facing legal problems, he hired an attorney in 1993 to join the hospital clinical team and represent patients.[14] Since then, additional attorneys and other staff have been hired; the Boston partnership now aids not only BMC pediatric patients, but also those at six affiliated health centers.[15] In addition, with foundation support, the Boston partnership established the national Medical-Legal Partnership for Children and now provides technical assistance and seed money to encourage the establishment of similar partnerships around the country.[16]

Medical-legal partnerships typically consist of at least one medical practice and one law practice, with a partnership medical director and a partnership legal director. Many current partnerships involve an alliance between a children's hospital or the pediatric department of a major hospital and a local legal aid office that provides general civil legal services to low-income families.[17] The legal director or other attorneys provide training about the basic legal rights of children to doctors and other medical personnel, helping them recognize when children or their families are experiencing problems that potentially could be remedied. Together, the directors create screening tools and an efficient referral mechanism so that when doctors see a patient that could benefit from legal intervention, the patient can access the legal team. The lawyers also typically make themselves available for "case consultations"; thus, a doctor can present a question about a patient's situation and receive a quick answer or advice from the lawyer about whether further legal assistance would be beneficial. In many partnerships, the lawyers are on-site at the hospital or health clinic, which allows for informal collaboration and relationship building. Ideally, partnerships develop to the point where the lawyers and doctors can jointly identify systemic barriers to child well-being and work collaboratively to remedy those barriers.

The following story exemplifies how a medical-legal partnership works in practice.[18] J.M. was a sixth-grade student diagnosed by a pediatric psychiatrist

---

11. *Id.*

12. *Id.*

13. *Cf.* Zuckerman et. al, *supra* note 2, at 225 (noting the importance of appropriate education to children's health).

14. Lawton, *supra* note 10, at 12; The Medical-Legal Partnership for Children, Boston: About Us, http://www.mlpforchildren.org/about-us-boston.aspx (last visited Mar. 6, 2008) [hereinafter MLPC, Boston: About Us].

15. MLPC, Boston: About Us, *supra* note 14.

16. The Medical-Legal Partnership for Children, http://www.mlpforchildren.org/default.aspx (last visited Mar. 6, 2008).

17. *See* MLPC, Partnership Sites, *supra* note 8.

18. J.M.'s story is a composite of the stories of several clients represented by the

with bipolar disorder and anxiety. The psychiatrist was treating the mental health condition with drugs and psychotherapy. At home and in the community, J.M. was improving with the treatment; at school, however, his symptoms were worsening. He was failing all of his classes and refused to engage in classroom activities. He spent considerable time with his head on his desk under a sweatshirt. J.M. was labeled as "behaviorally-emotionally disabled" and was placed in a separate class with other behaviorally-emotionally disabled children. His classmates tended to be behaviorally disabled, rather than emotionally disabled like J.M., and the teacher was a strict disciplinarian who managed the class with a loud voice and firm hand. J.M. did not respond positively to this environment; the louder and firmer the teacher became, the more J.M. withdrew from classroom activities. J.M.'s mother talked to the psychiatrist about the situation. He was reluctant to prescribe more medication, since a therapeutic dose seemed to have been achieved for the home setting. They both felt the school setting was inappropriate, but the mother had been told by school officials that there were no other options. Neither the doctor nor the parent felt they could do anything to improve J.M.'s condition.

As a participant in a medical-legal partnership, the psychiatrist had attended a workshop offered by the legal team about special education and the value of advocacy to effect change. Because of training from the partnership's legal team, the psychiatrist knew the school setting could be challenged through either an IEP meeting[19] or an administrative hearing. He suggested that J.M.'s mother seek an attorney's help. The legal team had also made access for patients' parents quite simple, so a referral to the attorney was easily made.

When the partnership attorney investigated the case, she agreed to advocate for a more suitable classroom placement. Because of the partnership, J.M.'s psychiatrist was extremely cooperative about returning phone calls, talking with the attorney, and working with the attorney to sign a letter describing J.M.'s condition and his recommendation for a placement that would lessen J.M.'s anxiety. The psychiatrist's letter was crucial to the attorney's presentation at J.M.'s IEP meeting, as a result of which the IEP team placed J.M. in a regular classroom and provided him with support from a special education teacher who helped children with learning disabilities. The more nurturing teaching style and less aggressive tendencies of the children in the new classroom turned out to be the right fit for J.M., who brought his head out from under his sweatshirt and began to participate in school. His anxiety symptoms quickly decreased, eliminating the need for additional medication.

The partnership facilitated the result here in a number of ways. It gave the doctor sufficient information about special education to engage him in the topic and enable him to discuss options with J.M.'s mother. He knew that she did

---

Children's Law Clinic at Duke University School of Law.

19. An "IEP meeting" is a school meeting convened to discuss a child's Individualized Education Program. Although IEP meetings are not typically attended by attorneys, they provide a venue in which advocacy for a child can take place.

not have to accept the situation at school if it was impeding J.M.'s educational progress. He also knew there were mechanisms available to challenge the IEP team decisions. Because he had applicable print materials, he knew exactly how to refer J.M.'s mother to an attorney on the partnership's legal team who could handle her son's special education case. The attorney's job was made easier because the doctor had a vested interest in the case and understood why the attorney needed his letter. At the same time, the medical-legal partnership gave the attorney a better understanding of J.M.'s psychiatric needs, which allowed her to produce a key document supporting the mother's request for a change of placement.

III. LAW SCHOOL CLINICS AS "LEGAL PARTNERS"

Law school clinical programs can be ideal participants in a medical-legal partnership because student participants can both contribute greatly to and benefit tremendously from the relationship. Furthermore, a clinic with expertise in special education law and other public benefit programs for children is distinctly well suited for a medical-legal partnership.

Children's law clinics are likely to find very willing partners in pediatric medical practices. As a group, pediatricians already embrace the concept of advocacy[20] and understand that a child's health is affected by many economic and social influences. For example, many pediatricians are keenly aware of the interplay between the school experiences of a child and his overall health and well-being. Given the complexity of the special education system, though, they feel powerless to intervene in that arena (even if they otherwise would be inclined to do so). Likewise, pediatricians see the effects when families lack the necessary financial resources to provide for their children's needs, but the intricacies of public benefits eligibility are quite outside the pediatrician's scope of expertise. Thus, pediatricians easily comprehend the benefits of collaborating with lawyers, especially with those who focus their efforts on the problems of at-risk children experiencing disabilities and poverty.

The work of a children's law clinic can be enhanced by the medical-legal collaboration, as well. At the front end, the pediatric partners are a good source of client referrals, which every clinic needs in order to operate.[21] The doctors, who may have been unaware of the existence of free legal services in the

---

20. *See* Charles N. Oberg, *Pediatric Advocacy: Yesterday, Today, and Tomorrow*, 112 PEDIATRICS 406, 406 (2003).

21. This can be more or less successful, depending on the circumstances. Particularly in the partnership's early stages, doctors may fail to remember its existence and the opportunity to refer patients. If the law clinic has some physical presence at the hospital or medical office, the frequency of referrals will likely increase. If the law clinic does not have a physical presence in the medical facility, it will be incumbent upon the legal team to create opportunities to remind the doctors about the partnership. Having at least one doctor who champions the partnership to his or her colleagues is vitally important to the vibrancy of the partnership. Developing a quick and easy referral mechanism is another key to the success of the referral system.

community, may be eager to connect their patients with that resource. Further into the representation, the law student advocate may find that access to the client's doctor is crucial to the representation. In the special education context, for example, the issue handled by the clinic may be a child's eligibility for special education services. In such cases, producing medical documentation of ADHD or other medical conditions supporting eligibility for special education services is essential to a successful outcome. The partnership gives the law student much easier access to the doctor, who can elaborate as needed on the condition, administer required tests, or make important notations in the chart. Furthermore, the existence of a partnership opens doors for law students that can enhance their advocacy in other types of clinic practice.[22]

IV. SPECIFIC BENEFITS FOR LAW STUDENTS

A medical-legal partnership within the law school clinic context offers rich experiences for law students. Exposure to a law practice that is intentionally interdisciplinary gives students an opportunity to work directly with other professionals, which they likely will need to do frequently during their legal careers. Although in this type of collaboration the other professionals are from the medical community, the lessons drawn from working with them apply in many other circumstances. The following sections discuss specific skills that can be developed in a medical-legal partnership.

A. Communication

Every profession has its own cultural norms, such as work styles and schedules, vocabulary, patterns of communication, methods of handling information, and modes of interaction with clients or patients. If attorneys wish to maximize their interaction with members of another profession, they must learn and adapt to the norms of that profession. Law students in a medical-legal partnership quickly learn that the norms of a busy medical practice are markedly different from the norms of a law practice. Law students are forced to develop strategies for reaching a doctor by phone, for learning medical terminology and acronyms (not to mention learning to read the often illegible chart notes), and for talking to a doctor with medical vocabulary rather than legal jargon. As law students develop their skills in communicating with doctors and other members of the medical profession, they find themselves better equipped to represent their clients. With some guided reflection, the students also learn that they can use the same adaptation skills with professionals in other fields, enhancing their abilities to communicate well across professional norms.

22. Clinics handling disability cases are likely to find that this medical-legal partnership model significantly benefits their practice, given the universal need to develop medical evidence in those cases.

B. Application of Legal Standards

In cases involving eligibility for benefits, whether they be educational benefits under the Individuals with Disabilities Education Act, disability benefits, or other government benefits, attorneys must develop evidence to show that their clients meet the applicable legal standard to qualify for assistance. For example, in a special education case, an attorney may need to show that a child in school meets the legal standard for having a "behavioral-emotional disability." The information obtained from the child's medical and mental health records will contain medical diagnoses. A law student will need to develop the skill to distinguish between a legal standard (such as the criteria for being considered behaviorally-emotionally disabled for purposes of qualifying for special education services) and a medical standard (such as the laboratory and other clinical information that results in a diagnosis of a particular emotional illness) and to translate the medical information to the legal standard. The student must see the medical information as evidence used to prove that the legal criteria are satisfied and take on the responsibility of developing that evidence.

A case involving a Supplemental Security Income application provides a good example of how this works in practice. If a child is denied Supplemental Security Income benefits for a failure to meet disability requirements, reversal nearly always depends on getting the right medical information into the child's file and persuasively presenting that information to the Social Security Administration. The medical personnel, especially medical personnel in a medical-legal partnership, are likely to be willing to share information, but may be frustrated that the diagnoses included in their patient's chart are not sufficient. By using good communication skills, the clinic student can help the doctor understand what specific findings, test results, or other conclusions must be in the records to meet the legal standards. The student can then highlight those requirements in communications with the Social Security Administration, establishing eligibility for the child.

This skill of translation between medical evidence and the legal standards by which that evidence must be measured is relevant to many other fields of practice. Students who later handle medical malpractice, worker's compensation, or disability rights as attorneys will find this skill directly applicable to their work. Moreover, attorneys who practice in other areas, particularly those involving scientific or technical evidence, will apply the same principles in another milieu. Because the medical-legal partnership provides relatively easy access to the medical clinicians, the partnership creates a favorable environment in which to nurture this translation skill.

C. Presentation Skills

Lawyers involved in a medical-legal partnership are responsible for training the medical staff on various legal problems that might affect a child's health and how legal remedies could positively affect their patients' overall health.

This training encourages doctors to consider the potential of patient advocacy and encourages them to make referrals to the legal team of the partnership. Involving students in the creation of the training modules gives them an opportunity to develop an extremely transferrable skill: the ability to crystallize the basic principles of an area of law and concisely and interestingly present them to an audience.

As most teachers know, only when one must convey concepts to an audience does one truly grapple with those concepts. Teachers must grasp the whole subject matter in order to reduce it to its essence. Many attorneys act as "teachers" throughout their careers, whether in one-on-one counseling sessions with clients, in court before judges or juries, or in presentations to administrative agencies or boards. Doctors and other medical personnel are good audiences for students to teach: They are smart and sophisticated, but very busy. In creating a training session for the medical team in the partnership, law students must fully understand the subject matter and determine the most efficient and powerful method of presentation, given the characteristics of the audience. They must make choices about vocabulary, visual aids, handouts, and the like, and then must deliver the presentation. Students will have few other opportunities in law school to develop these highly useful skills.

D. Development of an Interdisciplinary Outlook

A medical-legal partnership provides law students with a unique opportunity to see how a lawyer's role fits into the workings of an interdisciplinary team. For example, when the client is an at-risk child with chronic health problems, the medical team may already include a pediatrician, a social worker, and a psychologist. The medical-legal partnership integrates an attorney into the team. Students will observe how an interdisciplinary team can address a child's situation holistically and will identify the unique contributions that can be offered by an attorney. Working in a partnership also gives clinic students an understanding of how the legal problems faced by a child and his family are interrelated with other issues affecting the child's overall well-being. The existence of a partnership invites law students to think comprehensively about a child's issues and to use the partnership to marshal resources for the child.

A school discipline case provides an appropriate example. A child facing suspension from school for marijuana use is referred by the social worker at the partnering medical practice to the law school clinic. Rather than focusing purely on the child's defense to the charges, the law student will be encouraged to talk with the social worker and medical providers to explore whether underlying issues such as depression or family instability are involved in the child's case. This collaboration may trigger additional referrals for services, as well as provide a potential approach for the clinic student to take in the child's suspension hearing.

V. BENEFITS TO THE LAW SCHOOL

Participation in a medical-legal partnership offers other advantages to the law school. If the partnership is between a university's law school and its hospital or medical school, the university administration may be particularly apt to support the effort and may even be willing to provide financial support to the law school clinic. The partnership may also generate publicity for the law school; an article about hospital residents working together with law students to provide coordinated services to local at-risk children would be a welcome addition to most any alumni magazine. In addition, a partnership provides an opportunity for the law school to show the community—particularly the medical community—that it is training its students to be compassionate and caring professionals (not just malpractice attorneys!).

VI. CONCLUSION

As law school clinicians contemplate the future of their programs, adding mechanisms for interdisciplinary lawyering may be highly important. The medical-legal partnership model offers one possibility for consideration. A clinic focused on children's issues, especially education issues, is a prime candidate for partnering with local pediatricians, who undoubtedly feel the same frustration Dr. Zuckerman felt at the limits of his effectiveness in addressing non-medical obstacles to his patients' health.[23] Not only does such a partnership provide law students with fruitful opportunities to work in an interdisciplinary setting, it also allows them to be a part of a coordinated and compassionate endeavor to improve the lives of their child clients.

23. *See supra* text accompanying notes 10–13.

THE JURIS DOCTOR IS *IN*: MAKING ROOM AT LAW SCHOOL FOR PARAPROFESSIONAL PARTNERS

STEPHEN A. ROSENBAUM*

I. FROM GOLD TO DIAMONDS

The day was much too lovely to spend inside. I ate my bagged lunch on the steps of the law school, facing Cumberland Avenue, yellow jackets, holly bushes and the sun.[1]

Ten years ago, at a celebration for the golden anniversary of clinical legal education at the University of Tennessee, Dean Richard Wirtz and Advocacy Center Director Jerry Black presented a comprehensive scheme for training future advocates. This scheme emphasized teaching practical skills to law students early in their education.[2] Wirtz and Black based some of their observations on recommendations from a commission reporting on the future of Tennessee's judicial system.[3] This commission recommended that future lawyers be "both disposed and trained: to attend conscientiously to their client's interests . . . with sensitivity to all of the human factors[4] [and] to resolve every dispute by the least combative and expensive means available . . . ."[5] Wirtz and Black drew other ideas from a faculty task force's recommendations for law schools' advocacy curriculum.[6] Together with their faculty colleagues, they

---

* Lecturer, University of California, Berkeley School of Law and Stanford Law School; Staff Attorney, Disability Rights California (DRC), a member of the federal network of protection and advocacy agencies. The views here are those of the author and not necessarily those of DRC, its staff or board of directors. The author thanks Wendy Shane for her research assistance and *Tennessee Law Review* editors Brittain Sexton and David Goodman, as well as Ms. Shane, for their editorial suggestions.

1. Personal Journal Entry of Stephen Rosenbaum (Feb. 24, 1973) [hereinafter Personal Journal Entry] (on file with author). This personal journal entry followed my visit to a University of Tennessee College of Law conference at the time of my first (undergraduate) legal services field placement with the Appalachian Research & Defense Fund.

2. Jerry P. Black & Richard S. Wirtz, *Training Advocates for the Future: The Clinic as the Capstone*, 64 TENN. L. REV. 1011, 1013–14 (1997).

3. *Id.* at 1011 (citing COMM'N ON THE FUTURE OF THE TENN. JUDICIAL SYS., TO SERVE ALL PEOPLE (1996) [hereinafter TO SERVE ALL PEOPLE]).

4. *See infra* text accompanying notes 82–90.

5. Black & Wirtz, *supra* note 2, at 1011 (quoting TO SERVE ALL PEOPLE, *supra* note 3, at 69 app. A). Dean Wirtz had served with his fellow Tennessee law school deans and board of law examiners on the Commission's eight-member working group on lawyer education and bar admission. *Id.*

6. *Id.* at 1012–13 (citing SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM.

urged law schools to take a page from the influential MacCrate Report[7] and "convey to the students a sense of what being a professional means, not only in terms of skills and knowledge, but also in terms of ethics, attitudes, and other dimensions of lawyering."[8]

We now mark the diamond anniversary of continuous clinical legal education at the University of Tennessee with new sources of guidance available. These include the Carnegie study on professionalism and a legal education road map charted by Professor Roy Stuckey and others. The latest report endorsed by the Carnegie foundation addresses the *formation*[9] of students in professional schools, or in the authors' words, the "apprenticeship" of professionalism and purpose.[10] Among other educational best practices, Professor Stuckey promotes instruction that nurtures cross-professional collaboration, teamwork, and effective communication with colleagues and other professionals.[11]

BAR ASS'N, LEGAL EDUCATION AND PROFESSIONAL DEVELOPMENT—AN EDUCATIONAL CONTINUUM, REPORT OF THE TASK FORCE ON LAW SCHOOLS AND THE PROFESSION: NARROWING THE GAP 236-45 (1992) [hereinafter MACCRATE REPORT]).

7. MACCRATE REPORT, *supra* note 6.

8. Black & Wirtz, *supra* note 2, at 1014; *see, e.g.*, MACCRATE REPORT, *supra* note 6, at 135–41, 233–68 (describing the "educational continuum" through which law school students acquire professional skills and values); TO SERVE ALL PEOPLE, *supra* note 3 (making recommendations for new directions in continuing legal education). Some of the fundamental lawyering skills identified by the Tennessee Commission's Working Group echo those detailed by the ABA Task Force, including the following: an understanding of alternative dispute resolution (Skill 8.4); familiarity with systems and procedures to ensure efficient allocation of time, resources, and effort (Skill 9.2); and development of systems and procedures for effectively working with others (Skill 9.4). MACCRATE REPORT, *supra* note 6, at 196–98, 200–01. The author of the ABA report was also one of the speakers at the Legal Clinic's 50th anniversary symposium. While reiterating the statement of skills and values contained in the report, Mr. MacCrate observed that many law schools were moving beyond traditional teaching in the development of new "coherent agendas of skills instruction" and were introducing new teaching methodologies into core courses. Robert MacCrate, *Educating a Changing Profession: From Clinic to Continuum*, 64 TENN. L. REV. 1099, 1132 (1997). For a look at the status of the MacCrate Report's recommendations almost a decade later, with an eye on the need for new skills acquisition, see Gary A. Munneke, *Legal Skills for a Transforming Profession*, 22 PACE L. REV. 105, 135–54 (2001) (stating that since the release of the MacCrate Report, "change in legal education has accelerated, not declined").

9. WILLIAM M. SULLIVAN, ANNE COLBY, JUDITH WELCH WEGNER, LLOYD BOND & LEE S. SHULMAN, EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW, 84, 128–29 (Carnegie Found. for the Advancement of Teaching 2007). The authors, who embrace the French approach to education capsulated in the term *formation*, actually view the intensive socialization, professionalization, and values-shaping inherent in traditional law school pedagogy as one of its few positive features. *Id.* at 185–86.

10. *Id.* at 97.

11. ROY STUCKEY ET AL., BEST PRACTICES IN LEGAL EDUCATION: A VISION AND A ROAD MAP, 77–79, 119 (2007) (setting instructional goals relating to professional skills and professionalism, including techniques to "communicate effectively" with colleagues and other

Anniversaries present an opportunity to retool the pedagogical machinery, reexamine and reshape the curriculum, reflect on advice not taken, and reignite what Professor Dean Rivkin calls the "insurgent movement for change"[12] in (clinical) legal education. Specifically, I challenge law schools to take the concept of *formation* one step further by extending educational opportunities to other members in the legal field, namely the paralegal community. Paraprofessionals can help lawyers accomplish their tasks with the efficiency, affordability, professional collaboration, and responsiveness to clients that is promoted by the leading legal educators. Thus, the inclusion of paralegals in the law school classrooms and corridors may not qualify as an act of insurgency, but as a valuable opportunity to generate dialogue, reflection, and criticism. The presence and engagement of nonlawyer peers would further open law schools to the public they seek to serve.

This Article promotes a modest pedagogical retooling: Law schools should offer a degree program for nonlawyer advocates.[13] This would capitalize on the many attributes that paralegals bring to the profession. In my argument, I focus on how teaching paralegals or lay advocates in law schools advances non-costly and non-adversarial dispute resolution, sensitivity to human and cultural aspects of client rapport, and co-education between members of the legal profession. I

---

professionals and the capacity to deal sensitively and effectively with colleagues and others from various backgrounds). While Professor Stuckey and the Carnegie Foundation bring a nuanced and renewed attention to longstanding concerns, "the same critiques and responses have been repeated" for seventy-five years in the vast literature on preparation of law students for practice. John S. Elson, *Why and How the Practicing Bar Must Rescue American Legal Education from the Misguided Priorities of American Legal Academia*, 64 TENN. L. REV. 1135, 1135 (1997).

12. Panel Discussion, *Clinical Legal Education: Reflections on the Past Fifteen Years and Aspirations for the Future*, 36 CATH. U. L. REV. 337, 340–41 (1987) (remarks of Professor Dean Hill Rivkin) [hereinafter Rivkin, *Clinical Education: Reflections*]. I have dubbed Rivkin the "James Dean of Clinical Education" for his desire to rekindle the early passion and rebelliousness of the legal clinical movement in which "[c]linicians claimed to be sensitive, egalitarian, nonhierarchical, mutual[ly] trusting, caring, open, etc.—offering a sharply contrasting profession[al] model to their nonclinical colleagues." *Id.* The term "insurgent" has since taken on a sinister meaning, but the original sentiment is still *à propos*. On the insurgency theme and the need to re-focus on accessing justice for underserved clients, see also Stephen Wizner & Jane Aiken, *Teaching and Doing: The Role of Law School Clinics in Enhancing Access to Justice*, 73 FORDHAM L. REV. 997, 998, 1006 (2004) (stating that "[t]hirty years ago a hardy band of public defenders and legal services attorneys stormed the academy" and it is time to consider a "return to our roots").

13. Apologies are extended to those offended by defining this person in terms of what she is *not*—a lawyer. For purposes of this Article, I use the terms *paralegal*, *lay advocate*, *legal assistant*, *lay practitioner*, and *paraprofessional* interchangeably, although I recognize there are perceived and real distinctions. *See, e.g.*, Alan W. Houseman, *The Future of Civil Legal Aid: A National Perspective*, 10 UDC/DCSL L. REV. 35, 64 (2000) (listing both *paralegals* and *lay advocates* in a legal services veteran's categorization of legal aid providers); *see also* Debra J. Monke, *What to Know Before Your Firm Hires a Legal Assistant: Why Paralegal Certification Counts*, 41 TENN. B.J. 22, 36 (2005) (using *legal assistant* and *paralegal* as synonymous terms).

use the specific situation of special education advocacy to show how trained paraprofessionals can be particularly effective in real-word scenarios. While my chief example is a lay advocate working on behalf of disabled[14] students in a public school setting, the value added by paralegals is by no means limited to that venue. Finally, I suggest what educators should emphasize in a new law school curriculum and how they might design a paraprofessional program.

II. NEW CLIÈNTELE AT THE ACADEMY

Law school sometimes gives law students the impression that they are solitary warriors [but] [l]awyers practice law as part of a team . . . .[15]

The legal educational establishment—law schools and the ABA accreditation overseers—should undertake a broader approach in preparing tomorrow's advocates. They should create opportunities for paralegals to study, train, and work side-by-side with future lawyers within the law school facilities.[16]

This reform is not explicitly recommended in any of the reports on improving the professionalization, skills development, and ethical components in the curriculum, nor is this proposal limited to *clinical* education. The concept of law-school-administered paralegal programs, however, is not entirely new. In the early 1970s, the avant-garde Antioch School of Law trained paralegals for the public sector, primarily through clinical experiences.[17]

---

14. I am obliged to make a disclaimer about use of the term "disabled clients" in contrast to "clients with a disability." For many of the disability cognoscenti, it is important always to use "people first" language to emphasize their humanity, not their disability. This linguistic predilection is somewhat analogous to that of favoring "people of color" over "colored people," although "disabled" lacks the pejorative connotation of "colored." The debate is really much more nuanced. Some disability activists actually prefer to accentuate the disability as a matter of identity and pride. *See* Patricia A. Massey & Stephen A. Rosenbaum, *Disability Matters: Toward a Law School Clinical Model for Serving Youth with Special Education Needs*, 11 CLINICAL L. REV. 271, 272 n.3, 286 n.78 (2005) (explaining reclaimed epithets and "disability first" language).

15. Munneke, *supra* note 8, at 146.

16. This program would certainly involve different admissions and graduation criteria. Presumably, applicants would not take the LSAT examination and might not have obtained a four-year undergraduate degree. *See infra* text accompanying notes 122–23. The name of this degree could be as hard to fathom as an adequate title for the paralegal candidate who earns it. Most likely, it would include the word "juris," if not "doctor."

17. *See* Brief for National Paralegal Institute as Amicus Curiae Supporting Respondents, Procunier v. Martinez, 414 U.S. 973 (1973) (No. 72-1465), 1973 WL 171721 (providing a history of paralegal training). Antioch Law School has since closed its doors. The University of West Los Angeles School of Law, accredited by the State Bar of California Committee of Bar Examiners, ostensibly offered a paralegal program with a basic curriculum structure similar to that of the law school. *Id.* Its website does not contain information about paralegal studies, but

A number of other law schools have piloted paralegal studies in specialized public law subjects, such as fair housing, consumer claims, landlord-tenant, welfare, domestic relations, Social Security, and human rights.[18]

Paralegal programs would deliver tangible benefits to traditional law students. Opening law schools to a new class of advocates would strengthen future lawyers' abilities to deliver legal services more efficiently and to communicate more effectively with clients and co-workers. With exposure to paralegal students, the traditional law students would obtain some of the fundamental paralegal legal research and drafting skills.[19] Also, the law students could observe the instinctual and experiential know-how that a

---

it does note that the school's mission is the "democratization of the legal community." Univ. of West L.A. Sch. of Law, Mission & History, http://www.uwla.edu/Welcome/Mission History.aspx (last visited Feb. 18, 2008). Capital University Law School is an ABA-accredited school with a paralegal training program. *See* Capital Univ. Law Sch., Paralegal Programs, http://www.law.capital.edu/Paralegal (last visited Feb. 18, 2008).

18. Brief for National Paralegal Institute, *supra* note 17, at *6–7. These pilot schools included Boston College Law School, Columbia Law School, and the Denver College of Law. *Id.* It does not appear that any of these schools offers a paralegal studies program at the present time. *See* Columbia Law Sch., Centers and Programs, http://www.law.columbia.edu/center_program (last visited Feb. 18, 2008); Boston Coll. Law Sch., Curriculum, http://www.bc.edu/schools/law/home.html (follow "Curriculum & Course" hyperlink) (last visited Feb. 18, 2008); Univ. of Denver Coll. of Law, Degree Programs, http://www.law.du.edu/degrees/ (last visited Feb. 18, 2008). The University of Denver College of Law, however, offers a Master of Science in Legal Administration to "train[] students in the legal culture; the process of lawyering; and principles of business management unique to the legal environment." *See* Univ. of Denver Coll. of Law, M.S. in Legal Administration Program, http://www.law.du.edu/msla/index.cfm (last visited Feb. 18, 2008). While the program's express purpose is not the training of paralegals, it could easily be a model for students seeking that kind of training and is indicative of the kind of nonlawyer degrees that law faculties may opt to offer. Boston College offers a Professional Studies Certificate in Criminal and Social Justice, but the program is not directly affiliated with the law school. *See* Boston Coll., Professional Studies Certificate, http://www.bc.edu/schools/advstudies/certificate.html (last visited Feb. 18, 2008).

19. In a brief filed with the Supreme Court more than three decades ago, arguing that paralegals can and should be used to help meet the serious need for more adequate legal services for prisoners, the National Paralegal Institute wrote:

> The duties of paralegals vary according to the setting in which they work. In private law firms, for example, paralegals draft and file corporate documents, maintain clients' tax records, and collect data relevant to estate planning. In the public sector, among other things, paralegals interview clients, investigate facts, and conduct negotiations. Paralegals are being trained to perform a variety of functions in many different areas of the law.

Brief for National Paralegal Institute, *supra* note 17, at *2–3 (footnote and citation omitted). For typical skills and courses, see, for example, *id.* at *8–9; Monke, *supra* note 13, at 23–24; and All Criminal Justice Schs., Paralegal School Accreditation, http://www.allcriminaljusticeschools.com/faqs/paralegal_accreditation.php (last visited Feb. 18, 2008) [hereinafter Justice Schools].

paralegal brings to a task. These benefits support the notion that the legal academy should reach out to a wider audience.

Among these programs, the curricular and faculty infrastructure might vary from institution to institution, but the emphasis should remain focused on the ethics of practice, clinical education, skills training, and other forms of experiential education. At the same time, law schools should continue to assure that candidates for J.D. degrees are "disposed and trained" in these same values and skills. To achieve this, the faculty must provide a more diverse curriculum that focuses as much on *formation* as it does on technical skills. The prerequisites for a paralegal degree would be less comprehensive than that for lawyers.[20] The goal would be to offer a rigorous curriculum, teach shared skills with traditional law students, and use the clinical model and interdisciplinary teaching to foster a common knowledge base that would aid postgraduate collaboration.

Activist and academic Ed Sparer and his colleagues called for the creation of "lay advocate" centers to complement the work of legal services attorneys more than forty years ago.[21] Sparer's contemporary, the practitioner and scholar Gary Bellow, also touted a system using paralegals, noting that they are "long-term service providers capable of providing first class legal representation."[22] He argued that, aside from lowering the cost of advocacy,

---

20. Presumably, students enrolled in this program would pay less in fees than traditional law students. Their earning potential also would be lower than that of J.D. candidates. This could conceivably lead to tension within the law school student body. *See infra* text accompanying notes 123–25.

21. Edward V. Sparer, Howard Thorkelson & Jonathan Weiss, *The Lay Advocate*, 43 U. DET. L.J. 493, 494 (1966).

22. Gary Bellow, *Legal Services in Comparative Perspective*, 5 MD. J. CONTEMP. LEGAL ISSUES 371, 376 (1994). For similar and distinct reasons, the independent paralegal movement has taken hold in the United Kingdom and in post-colonial developing countries. *See generally, e.g.*, Thomas F. Geraghty et al., *Access to Justice: Challenges, Models, and the Participation of Non-Lawyers in Justice Delivery*, *in* ACCESS TO JUSTICE IN AFRICA AND BEYOND: MAKING THE RULE OF LAW A REALITY 53 (2007) [hereinafter ACCESS TO JUSTICE] (offering guidance for confronting difficulties in providing legal aid within Africa's criminal justice system). In South Africa, there has even been a proposal to redefine "legal practitioner" in that country's constitution to include paralegals, which would permit them to represent clients in court. *Id.* at 66. In the United Kingdom, "legal executives," as they are called, attend police interviews aside suspects, take statements from imprisoned defendants, and follow up with witness declarations. Adam Stapleton, *Introduction and Overview of Legal Aid in Africa*, *in* ACCESS TO JUSTICE, *supra*, at 3, 20; *see also* DEBORAH L. RHODE, IN THE INTERESTS OF JUSTICE: REFORMING THE LEGAL PROFESSION 136, 138, 141 (2000). If not an all-out endorsement of a paralegal degree, Professor Rhode makes an appeal for law schools to offer a differentiated instruction. In her view, "[t]he diversity in America's legal needs demands corresponding diversity in its legal education." *Id.* at 190. She notes that, in other nations, nonlawyers with legal training provide routine services in areas such as bankruptcy, immigration, uncontested divorces, and landlord-tenant matters "without demonstrable adverse effects." *Id.* By contrast, the American law school three-year program is "neither necessary nor sufficient" to train students to be competent

"we have been able to teach paralegals to do large amounts of legal services work as well as, and sometimes better than, their lawyer counterparts."[23] He suggested altering the two-to-one lawyer to paralegal ratio, in addition to "developing a form of 'free-standing' paralegalism not unlike nurse practitioners in medicine."[24] Ethicist Deborah Rhode concurred, suggesting that law schools take a cue from other nations. That is, offer training to paralegal specialists in areas of unmet legal need, and help design appropriate licensing structures for paralegals.[25]

Nonlawyer advocates have different names and functions. Sometimes they are paralegal technicians or researchers. At other times, they may play the roles of lay advisors or community organizers.[26] Sparer and his associates described the role of "lay advocates" in a sort of primer:

> The lay advocate teaches his clients to protest when he has been deprived of his most elementary rights, rather than suffer inwardly [and] sink further into despair . . . .
>
> The lay advocate teaches his clients to protest in such situations by going to an attorney—if there is an attorney available.
>
> Frequently, the lay advocate's protest can be effectively and properly made directly to the sources which have the authority to remedy the grievance.
>
> . . . .
>
> The lay advocate is a source of education as to fundamental legal rights.[27]

---

in these areas of unmet legal needs. *Id.*

23. Bellow, *supra* note 22, at 376. Professor Louise Trubek also acknowledged the role that lay advocates can play in issues of importance to poor clients. In writing about health care advocacy, she suggested that "law school education should be modified and lay advocates encouraged and trained." Louise G. Trubek, *Making Managed Competition a Social Arena: Strategies for Action*, 60 BROOK. L. REV. 275, 299 (1994). It is not clear whether she envisioned a full-fledged professional degree granting program for advocates, as her focus was on more involvement generally of law school clinics (and legal services organizations) in tackling health care issues. *Id.*; *see also* Jane R. Wettach, *The Law School Clinic as a Partner in a Medical-Legal Partnership*, 75 TENN. L. REV. 305, 306–10 (2008) (describing the role of law school clinics in pediatric medical-legal partnerships that serve needy clients).

24. Bellow, *supra* note 22, at 376.

25. RHODE, *supra* note 22, at 190.

26. In an attempt to define the role of the nonlawyer advocate, Sparer and colleagues observed that it "involve[s] legal rights and relationships . . ." and more than administrative hearings representation. Sparer et al., *supra* note 21, at 505, 512.

27. *Id.* at 513. The authors characterize the role of the lay advocate as "at once more independent of the attorney's and at the same time closely related to that of the attorney who represents the poor." *Id.* It should go without saying that the exclusive use of the masculine pronoun should be read in its historical context, for most of us have evolved in language and mindset since the 1960s.

Lay practice can be as nondescript as the "practice of law," which has been defined simply as "what lawyers do."[28] Whatever the definition, paraprofessionals can render competent, vigorous, and commonsense legal assistance in a range of legal or quasi-legal[29] situations.[30] These include problems concerning education, housing, foster care, public benefits, immigration status, health care, consumer affairs, or employment. In some instances, paralegals cut costs for clients and law offices, as well as preserve attorney resources. Moreover, paralegals are sometimes better equipped than lawyers when communicating with, informing, advising, and generally assisting clients.

Despite these benefits, concerns about the unauthorized practice of law and the quality of advice will persist, some more legitimate than others. Professor Sparer referred to "the incantation of the dark phrases—'solicitation,' 'unauthorized practice,' 'stirring up of litigation,' 'lay intermediaries'—murmured in the lobbies and men's rooms, but rarely debated in open fashion."[31] Of course, some of these "concerns" are due to the fact that lawyers have a vested economic interest in retaining a professional monopoly over the privilege to advocate in court. Some critics argue that lawyers' selfish economic desires partially explain unauthorized practice of law statutes, which ban direct advocacy by paralegals.[32]

Professor Rhode offers sanguine advice on the subject. She counsels against a prohibition on paralegal practitioners, instead suggesting regulation.[33]

---

28. Deborah L. Rhode, *Policing the Professional Monopoly: A Constitutional and Empirical Analysis of Unauthorized Practice Prohibitions*, 34 STAN. L. REV. 1, 45 (1981).

29. Sparer and his associates referred to these as "'low-level' legal problems." Sparer et al., *supra* note 21, at 493.

30. *Id.* at 510, 514–15; *see also* Thais E. Mootz, Comment, *Independent Paralegals Can Fill the Gap in Unmet Legal Services for the Low-Income Community*, 5 UDC/DCSL L. REV. 189, 199–202 (2000) (arguing that paralegal practitioners can provide affordable services in areas of social security, immigration, veterans benefits, unemployment compensation, workers' compensation and family law matters including domestic violence restraining orders).

31. Sparer et al., *supra* note 21, at 494. Presumably, the *sotto voce* discussions also take place in ladies' rooms nowadays. Sparer and his co-authors remind us that the canons of legal ethics and statutory restrictions on lawyering "were not conceived in the context of an overriding concern with equal justice . . . ." *Id.* The subject is far from buried. *See, e.g.*, William C. Bovender, *Treating the UPL Epidemic*, 42 TENN. B.J. 26, 27 (2006) (arguing lawyers have a duty to combat unauthorized practice of laws, notwithstanding charges of protectionism). *But see generally* AM. BAR ASS'N COMM'N ON NONLAWYER PRACTICE, NON-LAWYER ACTIVITY IN LAW-RELATED SITUATIONS (1995), *available at* http://www.paralegals.org/displaycommon.cfm?an=1&subarticlenbr=338#One (explaining the utility of paralegals when public protections are in place).

32. M. Brendhan Flynn, *In Defense of Maroni: Why Parents Should be Allowed to Proceed Pro Se in IDEA Cases*, 80 IND. L.J. 881, 901–02 (2005). Judges more naturally empathize with lawyers' complaints about the damage done by the unauthorized practice of law than with consumer complaints about being denied a chance to use the legal system. *Id.* at 902.

33. RHODE, *supra* note 22, at 137–39; *see also* Mootz, *supra* note 30, at 203–04

Training and guidance from law schools could help answer legitimate concerns from the bar and the public at large.

The legal academy and the bar[34] should be involved in the education, training, and certification of paralegals, rather than adopting a position of indifference or opposition.[35] Lawyers of tomorrow should have a co-educational experience with future colleagues in a collaborative, nonhierarchical, and reciprocal setting. This is preferable to operating in separate spheres, establishing impromptu relationships, or fretting over the quality of service provided by paralegals. Even skeptics of lay advocacy programs must acknowledge that law schools presently do not prepare future attorneys for collaboration with colleagues and subordinates, who are necessary in the practice of law.[36] This proposal would make the curriculum more comprehensive and increase the quality of the development of paralegals, who will be important assets in the professional partnership for providing legal services.

III. SPECIAL EDUCATION ADVOCACY: THE IDEAL SETTING

Lawyers . . . sometimes labor under the fiction that we provide technical assistance when in fact our clients are really not equipped to advocate on their own.[37]

Lay advocates can operate successfully in many forums. I will use my specialized practice area, special education law, to exemplify the benefits of paraprofessional partnership.[38] Parents of students with disabilities, for example, must negotiate the labyrinth of education law at Individualized

---

(suggesting other remedies to protect consumers against paralegal malpractice).

34. At least one state bar association actually allows for paralegal membership. Since 2002, the Indiana State Bar has offered an associate membership to paralegals who qualify on the basis of education through a degree program, continuing legal education, or through work experience. Edna M. Wallace, *Who's On First? Paralegals with an ISBA Membership*, 49 RES GESTAE 26, 26 (Feb. 2006).

35. Professor Sparer and his co-authors put the fundamental challenge to the legal profession this way: "whether it is prepared to assist lay advocates in equipping themselves with knowledge of such basic rights and in openly offering . . . to assist in the realization of such basic rights." Sparer et al., *supra* note 21, at 512.

36. Munneke, *supra* note 8, at 146 (discussing how students are given the impression that they are "solitary warriors, doing battle for their clients" without reference to associate lawyers, legal assistants, secretaries, and non-legal professionals).

37. Stephen A. Rosenbaum, *Aligning or Maligning? Getting Inside a New* IDEA, *Getting Behind* No Child Left Behind *and Getting Outside of It All*, 15 HASTINGS WOMEN'S L.J. 1, 12 n.59 (2004) [hereinafter Rosenbaum, *Aligning or Maligning?*].

38. For the past twelve years, I have advised or represented parents and students in special education matters ranging from workshops, consultations and IEP meetings to mediation, due process hearings and litigation.

Education Program (IEP)[39] conference tables, in mediation rooms, or at parent organizing meetings. These would all be appropriate settings for paralegals to engage in individualized or group advocacy on behalf of youths and their families.

The primary role of enforcing the Individuals with Disabilities Education Act (IDEA)[40] falls on parents and their advocates, where available.[41] In a major decision interpreting the IDEA, the Supreme Court declared that parents "will not lack ardor" in making sure their children gain access to all the educational benefits entitled to them under the Act.[42] While the Court reemphasized the central role of parental decision-making in its last term,[43] it overestimated the ability of parents to act on their own. This is particularly true when families are limited by poverty, disability, language barriers, immigration status,[44] or lack of formal education.

---

39. The hallmark of special education, the IEP is a written statement of a child's educational levels of academic achievement and functional performance and measurable goals, as well as placement, instructional methodologies and services developed by a team of educators and parents for meeting these goals. *See* 20 U.S.C. §§ 1401(14), 1414(d) (2000 & Supp. 2007); 34 C.F.R. §§ 300.22, 300.320–.324 (2006). I have previously commented that "the ritual of writing lengthy IEPs . . . . seems to follow less from the law than from district or parent culture" and that "[p]arents and school staff are too busy to assemble around a table on under-sized chairs for [these] marathon session[s] . . . ." Stephen A. Rosenbaum, *When It's Not Apparent: Some Modest Advice to Parent Advocates for Students with Disabilities*, 5 U.C. DAVIS J. JUV. L. & POL'Y 159, 173 (2001) [hereinafter Rosenbaum, *When It's Not Apparent*].

40. 20 U.S.C. §§ 1400–82 (Supp. 2007) (originally enacted as Education for All Handicapped Children Act of 1975, Pub. L. No. 94-142, 89 Stat. 773 (1975)).

41. NAT'L COUNCIL ON DISABILITY, BACK TO SCHOOL ON CIVIL RIGHTS: ADVANCING THE FEDERAL COMMITMENT TO LEAVE NO CHILD BEHIND 7, 70 (2000) [hereinafter NCD, BACK TO SCHOOL]. For an overview of special education law and procedure, see, for example, LINDA D. HEADLEY (revised and updated by STEPHEN A. ROSENBAUM), *Schools and Educational Programs*, §§ 5.01–5.02, *in* AIDS AND THE LAW (David W. Webber ed., 4th ed. 2007); COMMUNITY ALLIANCE FOR SPECIAL EDUC. & PROTECTION AND ADVOCACY, INC., SPECIAL EDUCATION RIGHTS & RESPONSIBILITIES (9th ed. 2005), *available at* http://www.pai-ca.org/pubs/504001SpecEdIndex.htm; Wendy F. Hensel, *Sharing the Short Bus: Eligibility and Identity Under the IDEA*, 58 HASTINGS L.J. 1147, 1152–79 (2007); Mark C. Weber, *Reflections on the New Individuals with Disabilities Improvement Act*, 58 FLA. L. REV. 7 (2006).

42. Bd. of Educ. v. Rowley, 458 U.S. 176, 178, 209 (1982). I have previously argued that "ardor may not be enough" to achieve success in light of some of the jurisprudential and legislative setbacks in the IDEA. Stephen A. Rosenbaum, *A Renewed IDEA and the Need for More Ardent Advocacy*, 32 HUM. RTS. 3, 3 (2005).

43. Winkelman v. Parma City Sch. Dist., 127 S. Ct. 1994, 2000 (2007) (noting that the statute lays out "general procedural safeguards that protect the informed involvement of parents in the development of an education for their child"). Although the courthouse door has been opened for non-attorney parents, the undertaking of an administrative hearing, not to mention a federal court appeal, is still a daunting task and not to be assumed lightly. On the other hand, the potential for attorney-advocate collaboration on a hearing or appeal is great.

44. We know anecdotally that undocumented immigrant parents have children enrolled in the nation's schools. *See, e.g.*, Dean Hill Rivkin, *Legal Advocacy and Education Reform:*

Under IDEA, parents are equal members of the IEP planning team, along with school personnel.[45] To be successful, parents must understand both their children and their children's disabilities. They also must be able to follow the proceedings of the IEP meetings, voice disagreement, seek clarification, and be willing to use the available procedures to resolve conflicts.[46] Successful decision-making and implementation under IDEA require skills and knowledge beyond the reach of many.[47] This is precisely where advocates—with or without a J.D.—can provide assistance.

The skills required for special education counseling and advocacy are not necessarily conventional legal skills.[48] Advocates versed in instructional methodology, behavior intervention, nursing, medicine, child development, or other therapies can provide great support to parents and legal practitioners.[49] Even skills in community organizing and policy analysis are relevant and useful. Private sector practitioners and law students alike have observed that what lawyers do "isn't law, it's social work."[50] This truism has particular resonance when working with disabled clients seeking education or other services.[51]

Successful advocates working on behalf of a special-needs child must have an understanding of education rights, as well as awareness of protections

---

Litigating School Exclusion, 75 TENN. L. REV. 265, 265–66 (2008) [hereinafter *Legal Advocacy and Education Reform*] (discussing *Plyler v. Doe*, 451 U.S. 968 (1981)).

45. There are extensive procedural protections for parents as the educational representatives of their children. 20 U.S.C. § 1415 (Supp. 2006); 34 C.F.R. §§ 300.500–.505 (2006).

46. *See* Rosenbaum, *When It's Not Apparent*, *supra* note 39, at 166–67, 172–86 (describing the trials and frustrations inherent in the IEP design and implementation).

47. *See* GAIL IMOBERSTEG, EVALUATION STUDY OF SPECIAL EDUCATION DISPUTE RESOLUTION ISSUES IN CALIFORNIA, FINAL REPORT 24, 26 (2000), *available at* http://www.cde.ca.gov/sp/se/ds/documents/duprevalrpt.pdf (discussing the perceived need for lawyers in the decision-making process). Sometimes the parent needs more *support* than actual advocacy. *See, e.g.*, Louise G. Trubek & Jennifer J. Farnham, *Social Justice Collaboratives: Multidisciplinary Practices for People*, 7 CLINICAL L. REV. 227, 243 (2000) (describing collaborative relationship with counselors, victim advocates and other service providers prevalent in the domestic violence context).

48. Sparer et al., *supra* note 21, at 506 (stating that an important skill is the ability to "add[] hope and a sense of human dignity").

49. Massey & Rosenbaum, *supra* note 14, at 306–07.

50. Jane Aiken & Stephen Wizner, *Law as Social Work*, 11 WASH. U. J.L. & POL'Y 63, 63 (2003).

51. *Id.* at 74–77 (describing lawyer *qua* social worker who serves her clients holistically and seeks to understand nature of all social diversity and oppression, including those related to mental or physical disability). Professor Sparer and colleagues recognized the value that social workers bring to resolving disputes in schools many years ago. In one vignette, their article recounts that "[w]hile the lawyer prepares a court action, the social worker seeks out the acting school superintendent and argues the [suspended] boy's cause. Before the lawyer files his papers, the suspension is lifted and the boy graduates." Sparer et al., *supra* note 21, at 502.

against discrimination. This unique form of advocacy also requires an understanding of the child's disability,[52] how schools and other bureaucracies function, and how to articulate a client's objectives and objections effectively. With the proper training, a paralegal can be well suited for this task.[53]

Parents may not have negotiation skills or familiarity with legal terminology. This contributes to the power imbalance, as does their lack of training in evaluating and marshalling evidence. IEP meetings have been described as "highly formal, non-interactive, and replete with educational jargon."[54] The stress, frustration, and anger that many parents experience also may interfere with their ability to present concerns in due process administrative hearings or mediation.[55] Advocates can provide the necessary distance and composure, along with knowledge and empathy.

A disproportionate burden falls on parents from marginalized groups to deal with these systemic obstacles. Those who do manage to avail themselves of procedural due process are predominately white, upper-to-middle class, English speaking, and well educated. In situations where this parental subgroup has difficulty with special education advocacy, non-English speakers with little formal education fare far worse. Marginalized parents are also likely to have greater difficulties using compliance complaints, alternative dispute

---

52. *See, e.g.*, Stephen A. Rosenbaum, *Representing David: When Best Practices Aren't and Natural Supports Really Are*, 11 U.C. DAVIS J. JUV. L. & POL'Y 161, 166–67 (2007) (describing a "dis-awareness" spectrum that spans from ignorant to hyper-aware).

53. In 2005, the Council of Parent Attorneys and Advocates (COPAA) received funding from the U.S. Department of Education, under a joint application with the University of Southern California University Center for Excellence in Developmental Disabilities, to create a standardized training curriculum and materials, and to develop guidelines and protocols for class instruction for lay advocate trainees. COPAA, Training Calendar, http://www.copaa.org/seat/index.html (last visited Feb. 18, 2008). The Special Education Advocacy Training (SEAT) Project trainees are required to complete approximately 115 hours of coursework, as well as a six-month practicum with an experienced special education attorney or advocate. *See id.* I served as an expert reviewer to the Project's curriculum advisory committee.

54. Martin A. Kotler, *The Individuals with Disabilities Education Act: A Parent's Perspective and Proposal for Change*, 27 U. MICH. J. L. REFORM 331, 364 (1994) (internal punctuation and citation omitted); *see also* Steven Marchese, *Putting Square Pegs into Round Holes: Mediation and the Rights of Children with Disabilities Under the IDEA*, 53 RUTGERS L. REV. 333, 351 (2001) (remarking that parents must face school officials "often speaking to each other in technical terms").

55. David M. Engel, *Law, Culture, and Children with Disabilities: Educational Rights and the Construction of Difference*, 1991 DUKE L.J. 166, 189 (1991). The IDEA provides an elaborate scheme of non-compliance complaints, mediation, "resolution sessions," and due process administrative hearing procedures. 20 U.S.C. § 1415(e)–(j) (Supp. 2006); 34 C.F.R. §§ 300.506–.518 (2006). Other characteristics that may interfere with parents' ability to advocate for an appropriate education for their children include the following: fear of retaliation against the student, a desire to maintain good relations with the school, cultural norms that place educators in positions of unquestioned authority, feelings of shame about having a child with a disability, and a sense of powerlessness. Rosenbaum, *When It's Not Apparent*, *supra* note 39, at 166, 176–81.

resolution (ADR), mediation, and due process hearings.[56] These parents also are far less likely to have sufficient resources to pay a traditional attorney to guide them through these mechanisms to secure an appropriate education for their children, but a lay advocate may be more within their means.

In each state, Parent Training and Information Centers and Community Parent Resource Centers train parents of disabled children and professionals who work with children. These centers often employ paralegal advocates who are themselves parents of children with disabilities. This assistance helps parents participate more effectively with professionals to meet the educational needs of children with disabilities.[57] However, this assistance is usually limited to *group* training or *individualized* information and referral, as opposed to direct representation.

Unfortunately, free legal service providers,[58] who are usually lawyers, are limited by staff capacity, case priority, and service guidelines. As the result of understaffing, most low-income and middle-income families cannot realistically secure representation under the current system. Therefore, cost is an obvious plus factor for paralegal participation at the IEP meeting or other pre-administrative hearing stages.[59] Lawyers' time should be preserved for more complex due process hearings[60] and for litigation, while paralegals can provide assistance throughout the process.[61]

---

56. *See* Massey & Rosenbaum, *supra* note 14, at 281–82; Rosenbaum, *Aligning or Maligning?*, *supra* note 37, at 11–12.

57. *See* Technical Assistance Alliance for Parent Centers, http://www.taalliance.org. These parent centers are funded by the U.S. Department of Education through discretionary grants authorized under IDEA. 20 U.S.C. §§ 1472–73 (2006).

58. The protection and advocacy systems throughout the states receive federal funding to represent individuals with disabilities, including special education students, to obtain their service, legal and human rights. 29 U.S.C. § 794e (2000); 42 U.S.C. §§ 10803 et seq., 15041 et seq. (2000). *But see infra* note 60 (describing possible limits on representation).

59. One commentator on special education representation writes, "On a more global level, people of low or moderate means often do not have access to the judicial system. Attorney fees are so extravagant that most of the populace cannot afford an attorney's hourly rates." Flynn, *supra* note 32, at 901–02 (citations omitted).

60. The special education legal community does not uniformly support advocacy by paralegals outside of IEP consultation and informal negotiations, notwithstanding the broad language of IDEA. *See* 20 U.S.C. § 1415(h)(1) (Supp. 2006) (stating that a party has the "right to be accompanied and advised by counsel and by individuals with special knowledge or training with respect to problems of children with disabilities"). At least one non-attorney parent has been prosecuted for unauthorized practice. *See In re* Arons, 756 A.2d 867 (Del. 2000) (upholding disciplinary counsel's ruling that IDEA does not authorize due process representation by nonlawyer advocates, and parent information centers can constitute unauthorized practice of law). This interpretation is not uniform in all states. *See, e.g.*, Mootz, *supra* note 30, at 196 (stating that District of Columbia permits a "representative of person's choosing" to appear at hearing).

61. *See* Rosenbaum, *When It's Not Apparent*, *supra* note 39, at 167–71 (explaining the "upsides and downsides" of litigation under the IDEA); *see also* Dean Hill Rivkin, *Legal*

Recognizing impediments to enforcement, the National Council on Disability made a number of recommendations to increase the availability of attorneys, technical assistance, and self-advocacy services.[62] The Council called on the Department of Education to fund a greater number of lawyers to counsel clients and to set up a national back-up center, along with self-advocacy training programs, for students with disabilities and their parents.[63]

Because Congress failed to authorize sufficient funds for more lawyers, this presents an opportunity to enlarge and train a corps of specialized paralegals as a less costly alternative. Also, in many instances, lay advocates better relate to their clients with respect to class, ethnicity, language, and parental status. They can assist parents at earlier stages in advising, negotiating, or informal decision-making. The contributions of these paraprofessionals would facilitate the entry of lawyers for more complex transactions, such as due process hearings or appeals.

Beyond their role in individualized educational planning, lay advocates can help instigate systemic change for disabled students. Organized parents have played a significant role in the enactment and reauthorization of special education laws. They have served as catalysts initiating change in the way that schools address the needs of students with disabilities.[64] Group advocacy can include anything from serving on an advisory committee or a consultative council[65] to joining statewide coalitions and *ad hoc* mass actions to forming

---

Advocacy and Education Reform, *supra* note 44, at 277–82 (discussing the mix of litigation and extrajudicial advocacy strategies necessary to enforce educational rights).

62. NCD, BACK TO SCHOOL, *supra* note 41, at 217–18; *see also* IMOBERSTEG, *supra* note 47, at 8–9 (reporting stakeholders agreement on recommendations for the special education mediation and hearing systems in California).

63. NCD, BACK TO SCHOOL, *supra* note 41, at 217–18 (Recommendation VII.7). Given the particular need of poor and underserved families, the Council specifically recommended that a lawyer be available at each parent center. *Id.* at 217. The community resource centers were specifically created to give training and information to the underserved parents of children with disabilities, including those who are low-income, have limited English proficiency, or are themselves disabled. 20 U.S.C. §§ 1472–73 (Supp. 2006). One member of the presidentially-appointed Council had recommended that public funds be used to train more lay advocates to help youngsters and their parents navigate the special education system, and that law schools give students more exposure to education disability law. *See* Lilliam Rangel-Díaz, *Ensuring Access to the Legal System for Children and Youth with Disabilities in Special Education Disputes*, 27 HUM. RTS. 17, 21 (2000); *see also supra* note 53 (describing the SEAT Project, a non-degree program for lay advocates jointly administered by a non-profit legal services organization and a university applied-research center).

64. Kotler, *supra* note 54, at 361–62; Rosenbaum, *Aligning or Maligning?*, *supra* note 37, at 30–37 (describing the need for "macro-advocacy" on behalf of classes of (disabled) students, as well as "micro-advocacy" in individualized IEP process). Lay advocates may support parents successfully at advisory committee sessions, parent strategy meetings, and at *tête-à-têtes* with school authorities.

65. Such councils are set up under the No Child Left Behind Act. 20 U.S.C. §§ 6316–7941 (2003).

parent-led organizations.[66] While working with the Harvard Family Research Project, Dr. M. Elena López observed that "[t]hrough one-on-one conversations, group dialogue, and reflection, parents and other residents develop a strong sense of community, and learn how to use their collective power to advocate for school change."[67]

Special education attorneys sometimes teach classes for parents or distribute self-help literature.[68] Some even have information links on their web sites that encourage parents to contact support groups during or after representation.[69] This partly acknowledges that parents continue advocating long after a single dispute had been resolved. These parents need not hire an attorney every time a disagreement with a school district arises. Client training has increased the number of parents and other nonlawyers who can serve as effective advocates at IEP meetings, mediation, or other ADR venues.[70] Paralegals can often substitute for or assist attorneys in supporting these parents, which would result in a great social benefit.

IV. A BLUEPRINT FOR EXPANSION

There's a sense of elitism and entitlement in law schools. . . . We are so self-contained in our own buildings and social activities . . . .[71]

Regardless of whether this new paraprofessional program is housed completely within the law school, the curriculum should not simply mirror a

---

66. ERIC ZACHARY & SHOLA OLATOYE, N.Y. INSTIT. FOR EDUC. & SOC. POL'Y, A CASE STUDY: COMMUNITY ORGANIZING FOR SCHOOL IMPROVEMENT IN THE SOUTH BRONX 6 (2001); *see also* M. ELENA LÓPEZ, HARVARD FAMILY RESEARCH PROJECT, TRANSFORMING SCHOOLS THROUGH COMMUNITY ORGANIZING: A RESEARCH REVIEW 6–8 (2003), *available at* http://www.gse.harvard.edu/hfrp/content/projects/fine/resources/research/lopez.pdf (discussing various forms of schools-based organizing).

67. LÓPEZ, supra note 66, at 1; *see also* Lyn Slater et al., *Report of the Parent Self-Advocacy Working Group*, 70 FORDHAM L. REV. 405, 408–09 (2001) [hereinafter *Self-Advocacy Working Group*] (stating that through value-based and skills-based training, professionals learn ways to empower parents to be strong and effective self-advocates); Rosenbaum, *When It's Not Apparent*, *supra* note 39, at 193–94 (giving examples of alliance building in special education context).

68. Massey & Rosenbaum, *supra* note 14, at 315 (citing informal conversation with attorneys at national conference of student-parent bar and distribution of popular informational DVD).

69. *See, e.g.*, Wrights Law, http://www.wrightslaw.com (providing informational resources regarding special education law and advocacy).

70. In anticipation of the 2004 IDEA amendments, a Bay Area group of lawyers and self-help providers, BASE-A, recommended the distribution to parents of "advo-kits" as part of a grassroots campaign to increase self-advocacy. Rosenbaum, *Aligning or Maligning?*, *supra* note 37, at 10.

71. SULLIVAN ET AL., *supra* note 9, at 150 (comments from law student focus group).

conventional paralegal certification program.[72] In addition to standard doctrinal courses, educational programs should focus on skills and traits such as those discussed above. Educators can also look to demand in the private sector for one good indicator of what these skills might be. Moreover, a paraprofessional program should foster "integrated" education by requiring joint classroom and clinical participation with traditional law students. While the paraprofessional graduates will probably have a shorter tenure at law school than would-be attorneys—and will generally command less pay for their work—they must nonetheless be viewed by J.D. candidates as colleagues with courses and resources in common. Paralegal education should not be merely a training program for junior lawyers or a fancy trade school for legal technicians and administrative assistants. This Article now turns to further discuss some of the core components of a paralegal program under the auspices of the law school faculty.

A. Professionalization

As already noted, law schools teach students professional values. Carnegie Foundation evaluators praised this part of the current *formation* effort.[73] The academy has a similar duty to instill these values in paralegals. Also, the professional benefits of a paralegal program are just as great for conventional law students as for the paralegals themselves. Lawyers-in-training build better relationships with paralegals, who are a type of worker that likely will be a daily part of their future career. Educators should continually challenge law students about their perceptions of co-workers, clients, and the communities in which they live.[74] Thus, integrating paralegal programs into law schools would further universities' goals of becoming public and democratizing institutions.

---

72. *See, e.g.*, Monke, *supra* note 13, at 23–24 (describing curriculum topics required for standard and advanced certification by a national paralegal association). One paralegal institute coordinator observes: "Even among paralegal educators, there's a great deal of debate about just exactly what we should be teaching paralegal students . . . ." Lori Tripoli, *How To Find and Groom the Practiced—and Practical—Paralegal*, 3 OF COUNSEL 12, 12 (2007) (quoting Pam Bailey, program coordinator of the Duquesne University Paralegal Institute).

73. SULLIVAN ET AL., *supra* note 9, at 185–86 (noting that socialization, professionalism, and career forming functions are among the few positive attributes of contemporary law school education). *But see* STUCKEY ET AL., *supra* note 11, at 100 ("Law schools do not currently foster professional conduct; just the opposite"). Professor Stuckey urges that professionalism be taught "pervasively and continuously" throughout a student's law school tenure, in both doctrinal and experiential courses and in the conduct of faculty and administrators. *Id.* at 100–04, 129, 170. He credits Professor Rhode for promoting the pervasive teaching of professional responsibility. *Id.* at 102. Under the rubric of "professionalism" Stuckey includes "appropriate behaviors and integrity in a range of situations . . . ." *Id.* at 79.

74. Opening the academy's doors to this class of students, and the accompanying curriculum, can help instill some of the insurgency that may be fading in the clinics. *See* Rivkin, *Clinical Education: Reflections*, *supra* note 12, at 340–41; Wizner & Aiken, *supra* note 12, at 998.

Law schools can build on the successes of existing programs that rely on advocacy without a law degree. One of the most prominent of these is the Court Appointed Special Advocate (CASA) program, which currently operates in every state in some form. Through CASA, volunteer participants represent children in the juvenile and family court system.[75] These advocates typically handle only one case at a time, and they are often motivated and well trained. CASA representatives have proven effective, especially in the tasks of investigation and monitoring.[76]

A paralegal program built upon a model code of professional conduct would serve clients well. It would stress adequate investigation, development of relationships with clients, monitoring of caseloads, and generally performing professional responsibilities in an ethical manner. Upon graduation, these advocates will need clear guidance on their roles and abilities to deliver quality representation.[77] Preparation for the workforce may require that faculty and students look beyond the law school walls to obtain necessary experience and knowledge.[78]

B. Legal Commodification

Law schools that admit students on a paralegal track would acknowledge the changing nature of legal practice and the changing role of lawyers

---

75. Randi Mandelbaum, *Revisiting the Question of Whether Young Children in Child Protection Proceedings Should Be Represented by Lawyers*, 32 LOY. U. CHI. L.J. 1, 23–24 (2000). CASAs are authorized by juvenile courts to advocate for the best interests of dependent and delinquent youth. 42 U.S.C. §§ 13011–13 (2005 & Supp. 2007).

76. Mandelbaum, *supra* note 75, at 24; *see also* Sparer et al., *supra* note 21, at 498–99 (discussing longstanding experience of non-attorney advocates in workers' compensation and unemployment benefits proceedings).

77. Mandelbaum, *supra* note 75, at 26. With reference to child dependency proceedings, Professor Mandelbaum writes:

> In those states where a representative is appointed, the qualifications, training, and support of the representatives vary greatly from state to state .... [O]nly about half of the states mandate that all children receive representation by attorneys. Where representation is not required to be by attorneys, it may be provided by paid or volunteer lay advocates ....

Id. at 23 (citations omitted).

78. Melissa Breger, Suellyn Scarnecchia, Frank Vandervort & Naomi Woloshin, *Building Pediatric Law Careers: The University Of Michigan Law School Experience*, 34 FAM. L.Q. 531, 532 (2000). Increasingly, lawyers in all practice settings are working with professionals across disciplines to resolve problems "in a more holistic, efficient, comprehensive and cost-effective fashion." V. Pualani Enos & Lois H. Kanter, *Who's Listening? Introducing Students to Client-Centered, Client-Empowering, and Multidisciplinary Problem-Solving in a Clinical Setting*, 9 CLINICAL L. REV. 83, 88 (2002) (citation omitted); *see also* Wettach, *supra* note 23, at 305–06 (discussing the growing trend of collaboration between legal and medical specialists); *infra* text accompanying notes 82–90 (arguing for interdisciplinary instruction and training).

themselves. The trend toward "unbundled services"[79] shows that certain legal cases can be disassembled and simplified. With proper training, lay practitioners can play a key role in this unbundling. After all, skilled paralegals specialize in routinizing legal output, and law schools can train them to perform these tasks even better by introducing fundamental legal concepts into their studies. Therefore, paralegal education must include some of the traditional courses in black letter law, basic research, and procedures. It should also include simulated exercises and clinical experience in tasks like drafting documents.

Both paralegals and lawyers would benefit from reform in the curriculum. Critics fault the legal academy for disjoining the teaching of substantive law and practical application of the law to standard legal instruments.[80] Supplementing the law school curriculum with more worldly experience in drafting and procedure will benefit *all* students.[81]

C. Sensitivity to Human Factors

The curriculum for paralegals should include courses that emphasize the development of intrinsic values, motivations, and problem-solving skills. A number of commentators have written about the lack of humanity in the typical law school diet. The widely acclaimed MacCrate Report encouraged legal educators to teach overlooked skills and values, but it was criticized for failing to address students' sensitivity to human factors.[82]

---

79. *See* Rochelle Klempner, *Unbundled Legal Services in New York State Litigated Matters: A Proposal to Test the Efficacy Through Law School Clinics*, 30 N.Y.U. REV. L. & SOC. CHANGE 653 (2006). Ms. Klempner states that unbundled legal services, also described as "discrete task representation" or "limited scope legal assistance," is a practice in which the lawyer and client agree that the lawyer will provide some, but not all, of the work involved in traditional full-service representation. *Id.* at 654. Simply put, the lawyers perform only the agreed-upon tasks, rather than the whole "bundle," and the clients perform the remaining tasks on their own. *Id.* Unbundled services can take countless forms, including providing advice and information, "coaching," drafting court papers, and making limited court appearances. *Id.*

80. In a recent reiteration of this criticism, the Carnegie Foundation team wrote that law school reinforces "the habits of thinking like a student rather than an apprentice practitioner, thus conveying the impression that lawyers are more like competitive scholars than attorneys engaged with problems of clients." SULLIVAN ET AL., *supra* note 9, at 188.

81. Law firm consultant and professor James Fanto writes:

> A common task of beginning lawyers is to add value quickly by doing something that is relatively routine: generating a first draft of a transaction agreement. Law schools . . . generally have not prepared their students to undertake this task. Students have not been trained to see the connection between the transaction agreements and the business law that they have learned . . . .

James A. Fanto, *When Those Who Do Teach: The Consequences of Law Firm Education for Business Law Education*, 34 GA. L. REV. 839, 844 (2000). The same critique can be applicable to other substantive law fields.

82. Professor Carrie Menkel-Meadow was among those who took the task force to task for

This concept dates back to pioneer clinician John Bradway, who wrote that "the clinical student has the opportunity to study the client as a whole in relation to [society] as a whole."[83] This feature distinguishes clinical education from older practices like apprenticeship. Familiarity with the client's human side requires exposure to other disciplines "such as medicine, social work, or religion, or . . . a combination of several of the social or physical sciences."[84] Bradway's protégé, Professor Charles Miller, also tried to integrate other disciplines into the legal clinic he founded at the University of Tennessee. He did this by establishing a relationship between the clinic and the College of Social Work.[85]

According to contemporary clinical dogma, "interdisciplinary, collaborative and real world experiences within a clinical setting" encourage sensitivity to human factors.[86] Clinical courses allow law students to develop "their subjective well-being, life satisfaction, and self esteem."[87] Just as the lawyer of tomorrow must "have a broader, more multi-dimensional and more

---

its failure to address the human sensitivity factor in law school training. *See* Carrie Menkel-Meadow, *Narrowing the Gap by Narrowing the Field: What's Missing From the MacCrate Report—Of Skills, Legal Science and Being a Human Being*, 69 WASH. L. REV. 593, 595–96 (1994); *see also* Janet Weinstein, *Coming of Age: Recognizing the Importance of Interdisciplinary Education in Law Practice*, 74 WASH. L. REV. 319, 346–47 (1999) (arguing that the MacCrate Report relied on a narrow, common-law approach to problem solving and did not pay enough attention to "the more humanistic roles of values, interests, problem prevention, interdisciplinary analysis, creative thinking and self-reflection"). One of Professor Stuckey's principles for best educational practices is that law schools help students "deal sensitively and effectively with diverse clients and colleagues," STUCKEY ET AL., *supra* note 11, at 77, 79, particularly in the area of cross-cultural competence. *Id.* at 88–89 (crediting the lawyering model developed by Professors Susan Bryant and Jean Koh Peters). For a thorough exploration of this model, see generally Susan Bryant, *The Five Habits: Building Cross-Cultural Competence in Lawyers*, 8 CLINICAL L. REV. 33 (2001).

83. John S. Bradway, *Some Distinctive Features of a Legal Aid Clinic Course*, 1 U. CHI. L. REV. 469, 470 (1934). Professor Bradway explained that "viewing the law as one of the social sciences . . . introduces a distinctly new element—the human equation. Not only the *legal* problems of the client but all his problems—social, economic and otherwise—should pass in review." *Id.*

84. *Id.* at 471.

85. Douglas A. Blaze, *Déjà Vu All Over Again: Reflections on Fifty Years of Clinical Education*, 64 Tenn. L. Rev. 939, 955 (1997). Ultimately, the clinic had "a social worker on staff and serv[ed] as a field placement for masters-level social work students." *Id.* Professor Sparer and associates also commented on the value of social workers operating as advisors and informal representatives in legal and quasi-legal contexts. Sparer et al., *supra* note 21, at 499–500; *see also infra* text accompanying notes 108–111 (arguing that law faculty should look to their colleagues in social work as a resource).

86. Janet Weinstein & Linda Morton, *Interdisciplinary Problem Solving Courses as a Context for Nurturing Intrinsic Values*, 13 CLINICAL L. REV. 839, 848 (2007) (describing courses at California Western School of Law).

87. *Id.* at 842.

interdisciplinary outlook on issues, as well as a more balanced life,"[88] the same applies to her paraprofessional peer. To that end, lawyers and paralegals should train together, study in the same institutions, and collaborate in clinics.

Cultural competency[89] also has been accepted as a core component of legal education. It should certainly be part of the paralegal studies curriculum. Many applicants to paraprofessional programs will be members of cultural minorities[90] and will further develop their cultural identities during their law school tenure.

D. Collaboration

The art and value of teaching collaboration may be more elusive than other topics. Increasing emphasis on teamwork and the growing diversity in the legal profession highlight the importance of collaboration.[91] Currently, relationships between attorneys and other legal staffers are frequently less than collegial.[92]

---

88. *Id.* at 840–43.

89. Among the classics on cultural competence, see Bryant, *supra* note 82, at 38 n.11 (2001); and Angela McCaffrey, *Hamline University School of Law Clinics: Teaching Students to Become Ethical and Competent Lawyers for Twenty-Five Years*, 24 HAMLINE J. PUB. L. & POL'Y 1, 57–59 (2002) (teaching cultural competence and identifying bias in the judicial system); and for other excellent sources discussing cultural competence, see also Stacy L. Brustin, *Bias in the Legal Profession*, *in* J. P. OGILVY, LEAH WORTHAM & LISA G. LERMAN, LEARNING FROM PRACTICE 346–53 (2d ed. 2007) [hereinafter LEARNING FROM PRACTICE] (helping students recognize cultural lenses and develop multicultural competence); Carolyn Copps Hartley & Carrie J. Petrucci, *Practicing Culturally Competent Therapeutic Jurisprudence: A Collaboration Between Social Work and Law*, 14 WASH. U. J. L. & POL'Y 133, 170–80 (2004) (asserting that educational models for developing cultural competence should be infused throughout law school curriculum, with attention to issues of power and oppression).

90. Cultural competence embraces disability as well as ethnicity, race, and language. *See, e.g.*, Massey & Rosenbaum, *supra* note 14, at 285–94 (discussing dis-awareness); STUCKEY ET AL., *supra* note 11, at 79 (discussing the capacity to "relate appropriately" to issues of culture and disability as well as "deal[ing] sensitively and effectively" with those from a range of social, economic and ethnic backgrounds).

91. Susan Bryant, *Collaboration in Law Practice: A Satisfying and Productive Process for a Diverse Profession*, 17 VT. L. REV. 459, 459–60 (1993). Fostering professional peer relationships while in law school is a core concept in clinical education. University of Tennessee Clinic Founder Charles Miller used his clinic, in part, "to help students establish professional relationships with lawyers in the community in which they intended to practice." Blaze, *supra* note 85, at 954; *see also* Munneke, *supra* note 8, at 146 (stating human relations are fundamental part of practice and lawyers "need to possess skills the necessary to work with others"). This Article proposes extending this concept to attorney-*paralegal* relationships.

92. T. Michael Mather, *Twelve Most Common Mistakes by Beginning Attorneys*, 26 TEMP. J. SCI., TECH. & ENVTL. L. 43, 47–48 (2007) (asking "[w]hat makes people think, that because they graduated from law school, they have a license to be abusive to secretaries, paralegals, mailroom personnel, information technology people, and so on?"). In contrast, one large firm, O'Melveny & Myers, "reminds its professionals to, well, behave professionally" through a specific firm-wide initiative, which includes joint attorney-paralegal training. Tripoli, *supra*

The typical law school agenda does not include a course about collaboration with paralegals.[93] Indeed, the legal education community has only recently recognized that teamwork and collaboration between fellow attorneys are skills worth cultivating.[94]

To the extent that scholars and teachers have addressed the subject of hierarchical relationships, such studies mainly have focused on the relationships between senior lawyers and new associates[95] or lawyers working in teams with other lawyers.[96] Scholars also should examine the relationships between lawyers and paraprofessionals. One challenge for new paralegals, as well as for new attorneys, "is to maintain a sense of professional identity and some autonomous control over professional development, while working successfully within the realities of law practice collaborations."[97]

Ideally, future lawyers and future paralegals will collaborate as peers in law school. Lawyers and "para-lawyers" labor differently, however, and their jobs demand different skills, attributes, and preparation. Thus, expecting their ultimate relationship to be nonhierarchical may be unrealistic. Even among J.D. candidates working in a clinical setting, "student collaborations may reflect subtle hierarchies."[98] This does not invalidate the need to instill collaborative, egalitarian values and habits. However, as Professor Catherine Gage O'Grady observes, the "subtle power differentials" should be accounted for and utilized

---

note 72, at 14. The firm's paralegal director notes, "That's a nice and unusual partnership [that] doesn't typically happen at other firms." *Id.*

93. One paralegal coordinator for a firm in San Francisco acknowledged that "[a]ttorneys don't learn how to use paralegals in law school" and are ill-equipped to know "how best to use them to their advantage." Tripoli, *supra* note 72, at 14.

94. *See, e.g.*, Bryant, *supra* note 91, at 459–61; David F. Chavkin, *Matchmaker, Matchmaker: Student Collaboration in Clinical Programs*, 1 CLINICAL L. REV. 199 (1994); Catherine Gage O'Grady, *Preparing Students for the Profession: Clinical Education, Collaborative Pedagogy, and the Realities of Practice for the New Lawyer*, 4 CLINICAL L. REV. 485 (1998); Lucia Ann Silecchia, *Management Skills*, *in* LEARNING FROM PRACTICE, *supra* note 89, at 319–23 (explaining how to work collaboratively with colleagues and professionals). Professor Stuckey endorses the principle of "collaborative learning" in part to help prepare students for their future roles in being accountable to law firm partners, supervisors, and other third parties. STUCKEY ET AL., *supra* note 11, at 120 (citing David Dominguez, *Principle 2: Good Practice Encourages Cooperation Among Students*, 49 J. LEGAL EDUC. 386, 387 (1999)).

95. *See, e.g.*, O'Grady, *supra* note 94, at 505–12 (discussing demands on new lawyer in working relationship with senior partner).

96. *See, e.g.*, Bryant, *supra* note 91, at 467–68 (examining hierarchy and bureaucracy in law firm culture).

97. O'Grady, *supra* note 94, at 512; *see also* Bryant, *supra* note 91, at 460 (noting importance of "structuring joint decision making in a non-hierarchical fashion"); Sparer et al., *supra* note 21, at 514 (discussing ways that lawyers relate to lay advocates).

98. O'Grady, *supra* note 94, at 521. Professor O'Grady cautions that "if one student is perceived as an 'expert' in an area, or is a year ahead in law school, or is a clinic 'veteran' returning for a second semester in the clinic, the other team member(s) may feel intimidated by such apparent expertise." *Id.* at 521–22.

"to teach practical lessons on maintaining autonomy while working within a collaboration."[99]

E. Alternative Dispute Resolution

The resolution of legal disputes through alternative forms and forums has significantly changed modern legal practice.[100] Its impact, as Professor Okianer Dark has pointed out, "can be seen in the law school curriculum, . . . in the publication of casebooks and other materials[,] . . . and in the development of [Alternative Dispute Resolution] centers or institutes at law schools."[101] The alternative dispute approaches are more amenable to long-term, non-adversarial relations than are adjudication and investigation. Some of the creative work in ADR is occurring in the field of special education.[102]

Lay advocates are well suited to engage in an ADR practice and have done so successfully.[103] The cost and time savings that come with mediation, conciliation, and other forms of informal dispute resolution are accomplished in large part by nonlawyers. A paralegal concentration should give high priority to this subject, and traditional law students should be encouraged to enroll in

---

99. *Id.* at 522. Clinicians are faced with a choice of eliminating "all hierarchy from student work teams and then guid[ing] the students in their efforts to engage in true collaborative decision-making" or accepting some degree of student-on-student hierarchy, as is inherent in supervisor-student collaborations, and use it as a teaching tool. *Id.* In recounting the early approach to supervision adopted by pioneer clinician John Bradway, Professor Blaze points out that there are "rich educational opportunities" to be found in working collaboratively on a matter with a supervisor by modeling, as well as experiential learning, and that one should not sacrifice the former. Blaze, *supra* note 85, at 956.

100. Okianer Christian Dark, *Transitioning from Law Teaching to Practice and Back Again: Proposals for Developing Lawyers within the Law School Program*, 28 J. LEGAL PROF. 17, 23–24 (2004) (citations omitted).

101. *Id.*; *see also* Black & Wirtz, *supra* note 2, at 1012–15 (describing field of concentration in advocacy and dispute resolution); Weinstein, *supra* note 82, at 324 (noting growing trend toward ADR, such as mediation, not only in practice but also in law school training). Pepperdine University School of Law, for example, offers a certificate in dispute resolution for those who hold a bachelor's degree, with a wide range of theoretical and practice courses in mediation, arbitration, negotiation, and conflict resolution. *See* Pepperdine Univ. Sch. of Law, Certificate Program, http://www.law.pepperdine.edu/straus/opportunities/certificate.html (last visited Feb. 18, 2008). *But see* RHODE, *supra* note 22, at 132–35 (advocating for a broader range of procedural choices and more information about ADR effectiveness).

102. For example, the Consortium for Appropriate Dispute Resolution in Special Education (CADRE) issues publications, sponsors symposia, and maintains a comprehensive website on alternative dispute resolution in the special education context. *See* CADRE, http://www.directionservice.org/cadre; *see also* Massey & Rosenbaum, *supra* note 14, at 308 (noting that other commentators have said that a mediation clinic can encourage "party empowerment and self-help" even more than a litigation clinic).

103. *See* Sparer et al., *supra* note 21, at 502.

doctrinal and clinical classes in this field as well. This would provide an appropriate area for law students and paralegals to collaborate.

F. Organizing

Commentators have explored the ambiguous and overused meaning of "organizing" in the legal context.[104] While the role of fostering client autonomy and empowerment, as well as working with established or loosely organized groups, has been urged upon community-based lawyers,[105] it is also well within the competency of lay advocates. Law schools can enhance the natural organizer traits that many paralegals bring to this movement. Prospective lawyers would benefit from the training as well.[106] However, teaching organizational skills demands a disciplinary perspective absent from most law faculties.[107] To fill the holes, we must turn to social workers[108] and professional organizers to teach "organization building, mobilization, education, consciousness raising, and legislative advocacy."[109] This will require law schools to call on colleagues in social work, urban or regional planning, education, or other departments to augment the curricular offerings through co-teaching, co-managed clinics,[110] joint appointments, and

---

104. *See, e.g.*, Scott L. Cummings & Ingrid V. Eagly, *A Critical Reflection on Law and Organizing*, 48 UCLA L. REV. 443, 460–69 (2001).

105. *See, e.g., id.* at 493–95, 500–01; Michael Diamond, *Community Lawyering: Revisiting the Old Neighborhood*, 32 COLUM. HUM. RTS. L. REV. 67, 123–26 (2000).

106. *See* Massey & Rosenbaum, *supra* note 14, at 311–15.

107. Professor Shin Imai is one of the exceptions. By undergoing a series of core communication and collaboration skills, Professor Shin's students are trained for work mainly in indigenous communities and other Canadian communities of color. Shin Imai, *A Counter-Pedagogy for Social Justice: Core Skills for Community-Based Lawyering*, 9 CLINICAL L. REV. 195, 201–25 (2002) (stating that skills include practicing "plain English" and emotional engagement). Professor Katherine Kruse writes about teaching skills for problem-solving for a client community, not just for individual clients. Katherine R. Kruse, *Biting Off What They Can Chew: Strategies for Involving Students in Problem-Solving Beyond Individual Client Representation*, 8 CLINICAL L. REV. 405, 408–09 (2002).

108. Professor Paula Galowitz catalogues the possible roles played by social workers, including teaching lawyers about working with community groups and community analysis. Paula Galowitz, *Collaboration between Lawyers and Social Workers: Re-Examining the Nature and Potential of the Relationship*, 67 FORDHAM L. REV. 2123, 2131–32 (1999) (citing Heather B. Craige & William G. Saur, *The Contribution of Social Workers to Legal Services Programs*, 14 CLEARINGHOUSE REV. 1267, 1268–71 (1981)); *see also* Aiken & Wizner, *supra* note 50, at 65–66 (arguing that empowerment of groups and communities, pursuit of social and economic justice, and reform are central to social workers' professional obligations).

109. Cummings & Eagly, *supra* note 104, at 481–84.

110. *See* Enos & Kanter, *supra* note 78, at 100 (pointing out that an increasing number of legal programs use multidisciplinary approach to service delivery by forming partnerships with other professionals).

interdisciplinary workshops or courses. This also may give law students and paralegals the opportunity to work in culturally diverse communities.[111]

G. Self-Advocacy

In addition to educating legal paraprofessionals, law schools should open their doors to members of the general public for workshops, abbreviated courses, and seminars. These programs should be aimed at improving the advocacy skills and legal literacy of persons who constantly encounter the same bureaucratic and quasi-legal procedures as lawyers and paralegals. This audience would comprise people from the community who pursue knowledge for its own sake, generally with the aim of helping themselves or a family member. Rather than a formal degree, these short-term adult learners might take home a hand-lettered certificate of attendance. The skills that these non-traditional students would gain could be transferred to numerous informal and administrative forums.

In the education context, legal training for former clients and community members can enhance parental skills and the capacity of school-based constituencies. Special education legal clinics, in particular, have embraced this approach as a way to serve the client community and provide unique learning opportunities for law students. One law school offers in-depth parent training through a lay advocate certification program that enables former client parents to help other parents become more effective advocates for their children.[112] Another school offers training as a component of its services.[113] Yet another law school runs advice clinics within a larger live-client framework. These special education legal clinics provide training, information, and self-help strategies to those whose cases are not selected for direct representation.[114]

---

111. *See* Rosenbaum, *Aligning or Maligning?*, *supra* note 37, at 10 (discussing the need for more intensive and nontraditional outreach); *see also* Aiken & Wizner, *supra* note 50, at 65–66 (stating that social workers learn skills, including participation in decision-making, cross-cultural awareness, and consideration of "the 'system' within which the client exists").

112. Massey & Rosenbaum, *supra* note 14, at 316–17 (stating that law school at State University of New York at Buffalo offers lay advocacy training and certification); *see also Self-Advocacy Working Group*, *supra* note 67, at 408 (stating that law schools and other institutions of professional education should be "targets of parents' advocacy efforts").

113. Massey & Rosenbaum, *supra* note 14, at 316 (referring to the Disability Rights Legal Center, located at Loyola of Los Angeles Law School). Clinics may also want to conduct periodic training in designated advocacy skills for case workers, educators, therapists, probation officers, and other professionals.

114. *Id.* at 317 (noting the practice of the University of San Diego's law school special education clinic). Interestingly, the drive for, and desire to improve, legal literacy seems to be at the same level in developing countries. *See, e.g.*, Stapleton, *supra* note 22, at 22 (African regional conference calls for lay advocate training and legal literacy programs); AFRICAN COMM'N ON HUMAN AND PEOPLES' RIGHTS, THE LILONGWE DECLARATION ON ACCESSING LEGAL AID IN THE CRIMINAL JUSTICE SYSTEM IN AFRICA para. 10 (adopted at

H. Program Structure

The most obvious mechanism for overseeing paraprofessional education would be establishment of a degree-granting program at the law school. This would mean offering doctrinal and skills courses, clinical experience, and field placements. Presumably, law schools would offer joint black-letter-law classes for law students and paralegal candidates, as well as specialized legal curriculum for the latter. When it comes to designing specific courses for paralegal students, law schools should review course descriptions and syllabi available from those schools and other public institutions that offer, or offered, paralegal studies programs.[115]

In addition, these students could take courses in other disciplines, such as social work, public policy, or planning. These "non-legal" courses might be offered by other departments on campus, sister schools, or by jointly appointed faculty. Hopefully, there will be some interdisciplinary courses, which would meet at the law school and also be available to traditional law students.[116]

In most instances, accreditation will not be an obstacle. Creating a new program should not jeopardize a school's standing with the American Bar Association (ABA) or American Association of Law Schools (AALS). For ABA certification—in the form of "acquiescence"—the additional degree program must not detract from the school's ability to maintain a J.D. degree program that satisfies the ABA Standards.[117] At the very least, ABA rules are not preclusive because some accredited law schools already offer paralegal programs.[118] Accreditation is also provided by regional agencies not affiliated with the ABA.[119]

ACHPR/Res.100(XXXX)(06) (Nov. 15–29, 2006)), *reprinted in* ACCESS TO JUSTICE, *supra* note 22, at 39, 44 (encouraging legal literacy).

115. *See supra* text accompanying notes 17–18.

116. *See supra* text accompanying notes 84–85.

117. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Standard 308 (2007), *available at* http://www.abanet.org/legaled/standards/standards.html (follow "Chapter 3" hyperlink). Acquiescence may be withheld for "lack of sufficient full-time faculty," for "lack of adequate physical facilities," or for "lack of an adequate law library to support both a J.D. and an advanced degree program . . . ." *Id.* at Interpretation 308-1. A Juris Doctor degree curriculum "lacking sufficient diversity and richness in course offerings" is also grounds for withholding acquiescence. *Id.* Regrettably, this is not an endorsement of multiple degree offerings: "Acquiescence in a degree program other than the first degree in law is not an approval of the program itself, and, therefore, a school may not announce that the program is approved by the [ABA]." *Id.* at Interpretation 308-2.

118. At least one ABA-accredited law school offers a paralegal training program: Capital University Law School. *See* Capital Univ. Law Sch., Paralegal Programs, www.law.capital.edu/Paralegal (last visited Feb. 18, 2008). This certified Legal Assistant Program, endorsed by the Columbus, Ohio Bar Association, was inaugurated in 1972 as the "first of-its-kind in the nation to offer paralegal studies for post-baccalaureate students." *See* Capital Univ. Law Sch., Our History, www.law.capital.edu/About/OurHistory (last visited Feb.

For AALS membership, so long as the J.D. program is not "impaired," a school may create a second educational program.[120] However, the association takes a more active approach than the ABA in its review of significant changes in operation. A major programmatic change must be reported to the AALS Executive Committee and reviewed by the committee before it is implemented.[121]

In-house and community-based clinics and practice settings also must be an integral part of the paralegal curriculum. These clinics would involve candidates for the J.D. and the paralegal degree working in teams on litigation, policy, negotiation, or organizing campaigns. Clinics would provide opportunities for conscious collaboration, mindful mentoring, and serious supervision. Also, law students and future lay advocates could share responsibilities in training members of the general public in short-term classes and occasional workshops.[122]

Some challenges will be determining the criteria for student admissions, academic standing, and costs. These will not necessarily mirror those already in place for traditional law students. Paralegal degree applicants may not be required to have a bachelor of arts or sciences, but perhaps an associate of arts or some other certificate of postsecondary study might suffice. The examination, grading, or other evaluative processes also will need to be

---

18, 2008); *see also* Theodore P. Seto, *Understanding the* U.S. News *Law School Rankings*, 60 SMU L. REV. 493, 535 (2007) (using a definition of "student" that includes LL.M.s, S.J.D.s, MBTs and *paralegals*). The ABA has approved over 250 paralegal and legal assistant training programs that meet their voluntary guidelines, almost all of which are situated outside of law schools. *See* Justice Schools, *supra* note 19. Online programs in paralegal studies have also been accredited by the ABA. Nick Dranias, *Past the Pall of Orthodoxy: Why the First Amendment Virtually Guarantees Online Law School Graduates Will Breach the ABA Accreditation Barrier*, 111 PENN. ST. L. REV. 863, 868 & n.24 (2007). On the benefits of distance learning for law students and ABA endorsement of same, see Michael L. Perlin, *An Internet-Based Mental Disability Law Program: Implications For Social Change in Nations With Developing Economies*, 30 FORDHAM INT'L L.J. 435. 439–42 (2007).

119. *See* Justice Schools, *supra* note 19. In 1976, the National Association of Legal Assistants initiated a certification, which includes a paralegal specialty credential. Monke, *supra* note 13, at 22, 24.

120. ASS'N OF AM. LAW SCHS. HANDBOOK, EXECUTIVE COMMITTEE REGULATIONS, § 6-7.6 (May 2005), *available at* http://www.aals.org/about_handbook_regulations.php.

121. *Id.* at § 8.2. After a member school "report[s] fully" its proposed change, the Executive Committee makes a determination of membership compliance, with an inspection if necessary. The Committee may consider such elements as changes in student recruitment and enrollment patterns, faculty hiring, teaching assignments and participation in governance, as well as relationships between the dean, faculty, administrators, staff and students. *Id.* at § 8.2(d).

122. The various special education or child advocacy clinic prototypes are ideal for this kind of lawyer or lay advocate training. *See* Massey & Rosenbaum, *supra* note 14, at 294–330, 333–34 app. 2 (reviewing skills, structure, and client caseload, as well as showing a chart of law school clinics).

reviewed. Finally, law schools should implement procedures that allow paralegal students to transfer into the J.D. program and vice-versa.

The tenure of lay paraprofessional students undoubtedly will be less than three years—perhaps one or two.[123] This should result in lower tuition or fees. The additional revenues that law schools will receive from tuition paid by paralegal students likely will be offset by additional costs of personnel and infrastructure. Also, because even reduced fees are likely to be a burden for many paralegal applicants, law schools should consider providing scholarships, financial aid, and loan forgiveness programs. Targeted recruitment of non-traditional law students also will be necessary.

Still, distinct criteria for a *para* doctoral and *juris* doctoral degree may pose less of a problem administratively than philosophically or fiscally. The main reasons for co-educating lawyers and paralegals in one building are fostering future collaboration and de-emphasizing hierarchical relationships among lawyers and paralegals. Separate standards in admissions and graduation could undermine these goals.

Faculty and staff must encourage an environment where a diverse law student body can navigate its way through the curriculum with equitable learning opportunities and mutual appreciation. Tensions inevitably will arise due to differences in students' academic or intellectual orientation, socio-economic status, and diverging career paths. Temptations to track students and segregate more than integrate may arise. Administrators and faculty also may oppose the co-educational scheme.[124] However, the latest Carnegie study reminds us that "in all movements for innovation, champions and leaders are essential factors in determining whether or not a possibility becomes realized."[125]

---

123. The most recent Carnegie Foundation study recommended that the third year of law school be a year of specialization for the J.D. candidate. SULLIVAN ET AL., *supra* note 9, at 195. By implication, the first two years are adequate for a basic foundation in the law. *See* RHODE, *supra* note 22, at 190 (questioning the necessity and adequacy of the three-year program).

124. There will likely be "[r]esistance to change in a largely successful and comfortable academic enterprise . . . ." SULLIVAN ET AL., *supra* note 9, at 202; *see also* STUCKEY ET AL., *supra* note 11, at 283–85 (commenting on the many reasons for the legal academy's "well-entrenched" resistance to change). The faculty divisions may not necessarily be along doctrinal or clinical lines. My dinner partner at the banquet culminating the September 15, 2007 University of Tennessee symposium on clinical legal education speculated that many law professors would have a hard time embracing my paraprofessional degree proposal. I fear their opposition may be founded more on elitism than principle, otherwise known as the "I Don't Do Paralegal Teaching Syndrome."

125. SULLIVAN ET AL., *supra* note 9, at 202.

V. FROM GOLD TO DIAMONDS TO GOLD—AGAIN

There seemed to be more green leaves down here.
The sun helped to make the place beautiful.[126]

To heed Professors Wirtz and Black, the desired outcome of a paraprofessional degree program should be conscientious advocacy with heightened sensitivity, at less cost, and with less contentiousness. For many paralegals, like their J.D. peers, their education will be enhanced by their own cultural backgrounds and workplace experiences. Ultimately, these students will enter the world of practice with rigorous training under their belts and professional principles on their minds. Adopting a paralegal program also presents an opportunity to demystify the law, democratize the law school, and deemphasize professional elitism. The legal community often aspires to these goals but only occasionally attains them.[127]

At the last decadal celebration of the nation's oldest continuous law school clinic, University of Tennessee Clinical Programs Director Doug Blaze reminded golden anniversary attendees of the need for ongoing discussion about the mission and methodology of clinical programs.[128] Professor John Elson urged them not to accept the status quo in American legal education nor to expect voluntary reform from law school administrators or the American Bar Association.[129] This advice is still relevant today: Neither be complacent about change nor build only on what has gone on before, lest it be "*déjà vu* all over again."[130] Now should be a time to reflect on ways to open the door to previously overlooked students in the legal academy, but reflection alone is insufficient. We also must move to restructure campus classrooms and law offices to prepare practitioners to meet the changing needs of the future.

126. Personal Journal Entry, *supra* note 1 (Feb. 23, 1973).

127. Bellow, *supra* note 22 at 376; Rivkin, *Clinical Education: Reflections*, *supra* note 12, at 340–41; Sparer et al., *supra* note 21, at 494; Wizner & Aiken, *supra* note 12, at 998.

128. Blaze, *supra* note 85, at 962.

129. Elson, *supra* note 11, at 1135.

130. Blaze, *supra* note 85, at 939.

A LEXICAL EXAMINATION AND (UNSCIENTIFIC) SURVEY OF EXPANDED CLINICAL EXPERIENCES IN U.S. LAW SCHOOLS

BECKY L. JACOBS*

I. INTRODUCTION

The future ain't what it used to be.
Yogi Berra[1]

In September 2007, the University of Tennessee College of Law's Legal Clinic celebrated its sixtieth year of continuous operation. To mark this significant milestone, the College of Law hosted a Symposium that explored the future of clinical legal education—"Looking Forward: The Next Sixty Years of Clinical Legal Education." An impressive and diverse array of clinical scholars attended the event, many of whom participated on panels organized to highlight emerging issues for clinical programs.

One of these panels addressed the topic of "Expanding Clinical Experiences," and I was honored, and not a little intimidated, to join the impressive scholars who participated on this panel.[2] As I began preparing my remarks for the event, I was challenged by a common problem, a problem about which Symposium attendees debated and upon which this Essay will focus. That problem is one of definitions: How does one (and who should) define a clinical experience, how does one (and who should) organize and label clinical offerings, and how does one (and who should) define "clinician"?

---

* Associate Professor of Law, University of Tennessee College of Law.

1. I have attempted to follow the venerable Doug Blaze tradition of referring to Yogi Berra. *See, e.g.*, Douglas A. Blaze, *Déjà Vu All Over Again: Reflections on Fifty Years of Clinical Education*, 64 TENN. L. REV. 939, 939 (1997).

2. *See* The University of Tennessee, Mediasite Presentations Catalog, http://mediabeast.ites.utk.edu/mediasite4/Catalog/ (follow "Charles Miller Legal Clinic—60th Anniversary Celebration" hyperlink in sidebar) (last visited Feb. 12, 2008). My co-panelists were Kim Diana Connolly, Associate Professor of Law and Director of the Environmental Law Clinic, University of South Carolina School of Law; Carl Pierce, W. Allen Separk Distinguished Professor of Law, The University of Tennessee College of Law; and Susan Deller Ross, Professor of Law and Director of the International Women's Human Rights Clinic, Georgetown University Law Center. *Id.*

II. DEFINITIONS AND LEXICAL NUANCES; OR, A CLINIC OR CLINICIAN BY ANY OTHER NAME

Don't get me right, I'm just asking.
Yogi Berra

Definitional challenges pertaining to law school clinical programs arise in several contexts. In this Essay, I will focus on three particular lexical obstacles: (1) the precise characterization of the history of the University of Tennessee's (UT's) Legal Clinic, (2) the dichotomous and rather inexplicable nature of the relevant "specialty" rankings published by *U.S. News & World Report*, and (3) the clinical offerings available to students at UT and other U.S. law schools. I also will mention the difficulties associated with identifying members of a law school's clinical faculty and why definitional issues may contribute to this difficulty.

A. UT Legal Clinic's Historical Pedigree

I wish I had an answer to that, because I'm tired of answering that question.
Yogi Berra

I stumbled onto the first definitional challenge when reviewing the venerable history of UT's Legal Clinic. This Symposium commemorated the Clinic's sixtieth anniversary. We here at UT are almost annoyingly, but justifiably, proud to say that our Clinic is one of the country's oldest[3] and most successful programs of its kind. Definitional challenges, however, require precision when claiming and describing our historical pedigree.

Program characteristics and nuances account for this need for linguistic precision. Since the late 1800s, law schools have flirted with experiential learning programs.[4] Indeed, most historians have identified the "legal dispensary" operated by students at the University of Pennsylvania Law School in 1893 as the first law school clinical program.[5] A number of law schools established similar programs over the next couple of decades.[6] These primarily extracurricular, non-credit programs were run by students and were voluntary.[7] UT students were among this vanguard of the non-credit clinical movement.

---

3. Blaze, *supra* note 1, at 940 n.3.

4. *See* William P. Quigley, *Introduction to Clinical Teaching for the New Clinical Law Professor: A View from the First Floor*, 28 AKRON L. REV. 463, 467 (1995).

5. *See* Robert MacCrate, *Educating a Changing Profession: From Clinic to Continuum*, 64 TENN. L. REV. 1099, 1102–03 (1997); Quigley, *supra* note 4, at 467.

6. Law schools with projects similar to Penn's "legal dispensary" included those at Cincinnati, Denver, George Washington, Harvard, Northwestern, Tennessee, and Yale Universities. *See* MacCrate, *supra* note 5, at 1103; Quigley, *supra* note 4, at 467.

7. *See* Blaze, *supra* note 1, at 940.

First-year students established the Free Legal Aid Bureau in 1915 and pledged to "'spend a certain amount of time each week in the assistance of the poor and needy citizens of Knoxville, whose wrongs would otherwise go without righting.'"[8]

The University of Southern California experimented with a for-credit law school clinical program in the late 1920s.[9] In this program, which lasted only six weeks, students earned credit for work at the Los Angeles Legal Aid Foundation.[10] (Stay with me here; I have almost reached the basis for UT's boast.) Duke University established the first for-credit, in-house legal clinic in 1931[11] but eliminated it twenty-eight years later.[12] It was in 1947 that UT created its for-credit in-house clinic, a clinic that, as commemorated by this Symposium, still is going strong today.[13] These sixty years of operation give UT's Clinic bragging rights as the oldest continuously operating legal clinic in the nation.[14]

Since the creation of the UT Clinic in 1947, many have reported on the development of clinical legal education programs in U.S. law schools.[15] Clinics now have become an integral part of the curriculum at nearly every law school in the nation.[16] The recent report on legal education by the Carnegie Foundation for the Advancement of Teaching recognized "the potential of clinical-legal education for bringing together the multiple aspects of legal knowledge, skill, and purpose."[17]

Indeed, all ABA-accredited law schools must "offer substantial opportunities for . . . live-client or other real-life practice experiences, appropriately supervised and designed to encourage reflection by students on their experiences and on the values and responsibilities of the legal profession,

---

8. MacCrate, *supra* note 5, at 1103 (quoting COLLEGE OF LAW, THE UNIV. OF TENN., DEDICATION 9 (1950)).

9. *Id.* at 1103–04, 1103 n.32.

10. *Id.* at 1103 n.32.

11. Margaret Martin Barry, Jon C. Dubin & Peter A. Joy, *Clinical Education for This Millennium: The Third Wave*, 7 CLINICAL L. REV. 1, 8 n.23 (2000).

12. Blaze, *supra* note 1, at 940 n.3.

13. *See id.* at 939.

14. *Id.* at 940 n.3.

15. Barry et al., *supra* note 11, at 3 n.6 ("There are numerous books, symposia, articles, and reports devoted to recounting, discussing, and examining the history of clinical legal education."); *see, e.g.*, Blaze, *supra* note 1, at 939–942 (describing the development of the first law school clinical programs); MacCrate, *supra* note 5, at 1102–05 (outlining the progression of legal clinics prior to World War II); *Report of the Committee on the Future of the In-House Clinic*, 42 J. LEGAL EDUC. 508, 511 (1992) (describing "the goals and teaching methods that many clinical teachers employ in their in-house, live-client clinics"); *Report of Committee on Legal Aid Clinics*, 1959 ASS'N AM. L. SCHS. 121 (reporting on the integration of clinical work into the law school curriculum).

16. *See* Barry et al., *supra* note 11, at 30.

17. WILLIAM M. SULLIVAN, ANNE COLBY, JUDITH WELCH WEGNER, LLOYD BOND & LEE S. SHULMAN, EDUCATING LAWYERS: PREPARATION FOR THE PROFESSION OF LAW (2007).

and the development of one's ability to assess his or her performance and level of competence."[18] Law schools have responded to this requirement by creating a dizzying array of clinical offerings with subject matters encompassing the full panoply of legal practice areas and with names that appear to have been assigned without reference to any consistent coding convention or organizing principles.

The fine distinction with which we at UT refer to the pedigree of our Clinic does not in any way diminish the significant accomplishment that these historical facts represent. It does, however, illustrate the nuances of nomenclature that abound in the literature reporting on and describing law school legal clinics in general and clinical offerings in particular. While clinicians ostensibly are armed with a "common vocabulary,"[19] that vocabulary is rich and textured, replete with subtle synonyms confusing to those not steeped in the parlance of clinical practitioners. The next sections explore in more detail the challenges of mastering this vocabulary.

B. U.S. News Law School Specialty Rankings

I knew exactly where it was, I just couldn't find it.
Yogi Berra

A second definitional issue arose when I consulted the much reviled, yet feared, *U.S. News & World Report* rankings for clinical programs.[20] *U.S. News* ranks "Clinical Training" programs at "America's Best" law schools, a list on which UT ranks a respectable number 16.[21] There is, however, a separate ranking for "Dispute Resolution" programs.[22]

---

18. SECTION OF LEGAL EDUC. AND ADMISSIONS TO THE BAR, AM. BAR ASS'N, STANDARDS FOR APPROVAL OF LAW SCHOOLS, Standard 302(b)(1) (2007), *available at* http://www.abanet.org/legaled/standards/20072008StandardsWebContent/2007-08%20Standards%20book.pdf [hereinafter ABA STANDARDS].

19. Barry et al., *supra* note 11, at 18.

20. U.S. News and World Report, America's Best Graduate Schools 2008, Law Specialties: Clinical Training, http://www.usnews.com/usnews/edu/grad/rankings/law/lawindex.php (follow "Clinical Training" hyperlink) (last visited Mar. 31, 2008) [hereinafter Clinical Training] (on file with the Tennessee Law Review).

21. *Id.* This rank is particularly impressive given the size of the community in which UT is situated. While the Knoxville population is estimated to be 182,337, U.S. CENSUS BUREAU, http://factfinder.census.gov/ (follow "Population Finder" hyperlink; then search "Knoxville, Tennessee") (last visited Jan. 31, 2008), the majority of the other schools appearing on the U.S. News Clinical Training ranking are located in much larger metropolitan areas with a concomitant increase in opportunities for clinical training fora and externship placements.

22. U.S. News & World Report, America's Best Graduate Schools 2008, Law Specialties: Dispute Resolution, http://www.usnews.com/usnews/edu/grad/rankings/law/lawindex.php (follow "Dispute Resolution" hyperlink) (last visited Nov. 9, 2007) [hereinafter Dispute Resolution] (on file with the Tennessee Law Review). *U.S. News* also ranked "Trial Advocacy" as a specialty for the first time in 2007. *See* U.S. News & World Report, America's Best

I preface all that follows by declaring my sincere regard for all of the law schools that appear on both the Clinical Training and Dispute Resolution lists. Regardless of the precise criteria by which schools are judged, academics would likely agree that the ranked schools do indeed merit their inclusion on a list of "America's Best" in their respective specialties. (Well, except for the travesty that UT is not ranked in the Dispute Resolution category!)

Thus, while I believe that I instinctively "know it when I see it"[23] and understand the programmatic distinction between these two specialties, the lexical characterization somewhat escapes me, and for the uninitiated, it might be even more inexplicable. Why, a neophyte might ask, is Dispute Resolution (DR) a separate specialty? Are clinics not the crucible where students learn dispute resolution in all of its forms?

This blurring of distinctions is apparent if one adopts Marc Galanter's conception of "litigotiation," his neologism for "a single process of strategic maneuver and bargaining in the (actual or threatened) presence of courts."[24] As Professor Galanter reminds us, most cases do not proceed to a full-blown adjudicative proceeding; "[s]ettlement is not an 'alternative' process, separate from adjudication, but is intimately and inseparably entwined with it."[25] If this is so, clinics are DR labs that offer students the opportunity to experience multiple stages of the single "litigotiation" process.

This conceptualization appears to comport with the way many academics who research and write on DR topics view the relationship of traditional models of adjudication, such as litigation, to so-called "alternative" DR processes.[26] For example, Professor Leonard Riskin and his co-authors illustrate "The Conflict Resolution Continuum" in their casebook, *Dispute*

Graduate Schools 2008, Law Specialties: Trial Advocacy, http://www.usnews.com/usnews/edu/grad/rankings/law/lawindex.php (follow "Trial Advocacy" hyperlink) (last visited Nov. 9, 2007) [hereinafter Trial Advocacy] (on file with the Tennessee Law Review). Yet more blurring of distinctions? *See also infra* note 30 and accompanying text.

23. Jacobellis v. Ohio, 378 U.S. 184, 197 (1964) (Stewart, J., concurring). In borrowing Justice Stewart's oft-quoted phrase, I certainly do not intend to compare legal clinics or DR programs to *Les Amants*. *See id.* at 186 (majority opinion). However, one might argue that, like the claimed theme of that film, many clinics are devoted to the cause of freedom, and the work of most clinical programs is infused with uncertainty.

24. Marc S. Galanter, *The Federal Rules and the Quality of Settlements: A Comment on Rosenberg's, The Federal Rules of Civil Procedure in Action*, 137 U. PA. L. REV. 2231, 2232–33 (1989). Professor Galanter first coined "litigotiation" in Marc Galanter, *Worlds of Deals: Using Negotiation to Teach About Legal Process*, 34 J. LEGAL EDUC. 268, 268 (1984).

25. Marc Galanter, *The Quality of Settlements*, 1988 J. DISP. RESOL. 55, 82.

26. *But see* Carrie Menkel-Meadow, *Pursuing Settlement in an Adversary Culture: A Tale of Innovation Co-opted or "The Law of ADR,"* 19 FLA. ST. U. L. REV. 1, 44 (1991). Rather than a process involving both cooperative and adversarial maneuvers, Professor Carrie Menkel-Meadow appears to view "litigotiation" as more of an adversarial process, one in which lawyers may use DR procedures such as negotiation and mediation as, for example, extra-procedural discovery mechanisms. *See id.* at 34–36, 33 n.167.

Resolution and Lawyers.[27] At the far left of this Continuum are the "Consensual Processes," beginning with negotiation. Mediation appears just to the right of this.[28] Moving further to the right, the Continuum identifies the "Adjudicatory Processes" of arbitration and, finally, trial.[29]

In a clinical setting, students encounter these integrated approaches to DR routinely; therefore, the *U.S. News* Dispute Resolution specialty ranking could be considered redundant to the Clinical Training category. Of course, the Clinical Training category may not take into account a law school's non-clinical offerings that pertain to DR—consistent with an "I know it when I see it" approach. It is this, I presume, that distinguishes the two specialties in the minds of those who rank.[30] Arguably, and I know that I am spouting heretical crazy talk here, if a law school does not have strong curricular offerings in alternative dispute resolution (ADR) topics as well as in traditional litigation-related subjects, the educational value of a clinical experience to students might be somewhat diminished.

It is interesting to compare the schools that appear on the *U.S. News* Clinical Training list with those ranked for Dispute Resolution and to note that there is overlap. For example, in 2007, six of the fifteen schools ranked in the Dispute Resolution category also were ranked on the Clinical Training list: Harvard University, Yeshiva University (Cardozo), Fordham University, Georgetown University, University of Nevada–Las Vegas (Boyd), and Northwestern University.[31] Even if one does not agree that the specialties are

---

27. LEONARD L. RISKIN ET AL., DISPUTE RESOLUTION AND LAWYERS 12 (3d ed. 2005).

28. *Id.*

29. *Id.*

30. Because I am morbidly curious, I went straight to the source and e-mailed *U.S. News & World Report* to discover what it believed was the distinction between a Clinical Training specialty and a Dispute Resolution specialty. They responded quite promptly, with this note: "[B]oth are separate course areas, clinical training is required as part [of] the law school curriculum[.] [D]ispute resolution is [a] separate area that deals with a narrower part [of] the law." E-mail from Bob Morse, U.S. News and World Report, to author (Sept. 11, 2007, 11:57:23 EDT) (on file with the Tennessee Law Review). I leave the reader to ponder that.

31. The fifteen schools noted for their Dispute Resolution specialty in 2007 are, in rank order: Pepperdine University, University of Missouri - Columbia, Hamline University, Harvard University, Ohio State University (Moritz), Marquette University, Yeshiva University (Cardozo), Pennsylvania State University (Dickinson), University of Oregon, Fordham University, Georgetown University, University of Nevada - Las Vegas (Boyd), Willamette University (Collins), Northwestern University, and Quinnipiac University. Dispute Resolution, *supra* note 22. The top Clinical Training programs are: Georgetown University, American University (Washington), New York University, Washington University in St. Louis, University of Maryland, University of New Mexico, CUNY - Queens College, Yale University, University of Michigan - Ann Arbor, Northwestern University, Catholic University of America (Columbus), Columbia University, Harvard University, University of California - Los Angeles, Fordham University, University of California - Berkeley, Boston College, Seattle University, University of Tennessee - Knoxville, Northeastern University, Stanford University, University of Nevada - Las Vegas (Boyd), George Washington University, Yeshiva University (Cardozo),

redundant, this overlap suggests at a minimum that there are definite synergies between these two specialties that improve the overall quality of both programs.

It follows, then, that those schools that appeared on the Dispute Resolution list but not on the Clinical Training list must have some characteristic that distinguishes them from the many other schools with excellent and strong DR curricular offerings and faculty.[32] Several factors may be relevant, such as the presence of a DR Institute or Center at the school;[33] the school's publication of a DR journal;[34] a school's LLM, certificate, or concentration in DR;[35] or the reputation of a school's DR faculty.[36]

University of Chicago, University of the District of Columbia (Clarke), University of Baltimore, University of California (Hastings), Brooklyn Law School, Rutgers - Newark, Tulane University, University of Wisconsin - Madison, and William Mitchell College of Law. Clinical Training, *supra* note 20.

32. UT, for example.

33. Thirteen of the top fifteen DR schools have an institute, center, or identified program pertaining to dispute or conflict resolution. *See* Pepperdine University School of Law, Straus Institute for Dispute Resolution, http://law.pepperdine.edu/straus/ (last visited Feb. 12, 2008); University of Missouri School of Law, Center for Dispute Resolution, http://law.missouri.edu/csdr/ (last visited Feb. 12, 2008); Hamline University School of Law, Dispute Resolution Institute, http://law.hamline.edu/adr/dispute-resolution-institute-hamline.html (last visited Feb. 12, 2008); Harvard Law School Program on Negotiation, http://www.pon.harvard.edu/ (last visited Feb. 12, 2008); Moritz College of Law, Alternative Dispute Resolution, http://moritzlaw.osu.edu/programs/adr/ (last visited Feb. 12, 2008); Benjamin N. Cardozo School of Law, Kukin Program for Conflict Resolution, http://www.cardozo.yu.edu/directory.aspx?page=3 (follow "Kukin Program" hyperlink) (last visited Feb. 12, 2008); Penn State Dickinson School of Law, Institute of Arbitration Law and Practice, http://www.dsl.psu.edu/academics/arbitration.cfm (last visited Feb. 12, 2008); University of Oregon, Appropriate Resolution Center, http://www.law.uoregon.edu/org/adr/ (last visited Feb. 12, 2008); Fordham Law, Feerick Center, http://law.fordham.edu/feerickcenter.htm (last visited Feb. 12, 2008); Georgetown Law, Georgetown-Hewlett Program in Conflict Resolution and Legal Problem Solving, http://www.law.georgetown.edu/hewlett/ (last visited Feb. 12, 2008); Willamette University College of Law, Center for Dispute Resolution, http://www.willamette.edu/wucl/cdr/ (last visited Feb. 12, 2008); Northwestern University School of Law, Program on Negotiation and Mediation, http://www.law.northwestern.edu/legalclinic/simulation/negotiations/ (last visited Feb. 12, 2008); Quinnipiac University School of Law, Center on Dispute Resolution, http://law.quinnipiac.edu/x127.xml (last visited Feb. 12, 2008).

34. Seven of the fifteen top DR schools publish a DR-related specialty journal. *See* Pepperdine University School of Law, Dispute Resolution Law Journal, http://law.pepperdine.edu/organizations/dispute_resolution_law_journal/ (last visited Feb. 12, 2008); University of Missouri School of Law, Center of Dispute Resolution Journal, http://www.law.missouri.edu/csdr/journal/ (last visited Feb. 12, 2008); Harvard Law School Program on Negotiation, Harvard Negotiation Law Review, http://www.pon.harvard.edu/publications/hnlr.php (last visited Feb. 12, 2008); Ohio State Journal on Dispute Resolution, http://moritzlaw.osu.edu/jdr/ (last visited Feb. 12, 2008); Cardozo Journal of Conflict Resolution, http://www.cojcr.org/ (last visited Feb. 12, 2008); Penn State Dickinson School of Law, World Arbitration and Mediation Review, http://www.dsl.psu.edu/publications/

The influence of these factors and of non-clinical courses related to clinical offerings on the Dispute Resolution specialty rankings raises yet another issue that is embedded in this query: What is the very basic definition of a "clinic"? The ABA Standards for Approval of Law Schools do not define the term. As previously mentioned, Standard 302(b)(1) requires that all ABA-accredited law schools offer "substantial opportunities for . . . live-client or other real-life practice experiences, appropriately supervised and designed to encourage reflection."[37] Interpretation 302-5 of that Standard notes that law schools might fulfill this requirement through "clinics or field placements."[38] Except for rather unhelpfully acknowledging that there is a distinction between a clinic and a field placement, this Interpretation leaves the ultimate definitional issue unresolved.

The Association of American Law Schools (AALS) definition seems to be the most commonly cited. In its 1992 report on future of in-house clinics, it stated that

> [c]linical education is first and foremost a method of teaching . . . [by which] students are confronted with problem situations of the sort that lawyers confront in practice; the students deal with the problem in role; the students are required to interact with others in attempts to identify and solve the

---

worldarbitration/index.cfm (last visited Feb. 12, 2008); Willamette Journal of International Law and Dispute Resolution, http://www.willamette.edu/wucl/journals/wjildr/ (last visited Feb. 12, 2008).

35. Ten of the top fifteen DR schools offer a certificate or Master's degree in DR. *See* Pepperdine University School of Law, JD/MDR, http://law.pepperdine.edu/academics/joint_degree_programs/jdmdr.html (last visited Feb. 12, 2008); University of Missouri School of Law, Master of Law in Dispute Resolution, http://www.law.missouri.edu/llm/ (last visited Feb. 12, 2008); Hamline University School of Law, Certificate Program in Dispute Resolution, http://law.hamline.edu/llm/dr-certificate.html (last visited Feb. 12, 2008); Moritz College of Law, Certificate in Dispute Resolution, http://moritzlaw.osu.edu/programs/adr/certificate.php (last visited Feb. 12, 2008); Marquette University, Center for Dispute Resolution, Law School Joint Program, http://www.marquette.edu/disputeres/programs/joint.shtml (last visitied Mar. 27, 2008); Kukin Program for Conflict Resolution, Certificate, http://www.cardozo.yu.edu/directory.aspx?page=3 (follow "Kukin Program" hyperlink, then follow "Certificate" hyperlink) (last visited Feb. 12, 2008); Penn State Dickinson School of Law, Certificate in Dispute Resolution and Advocacy, http://www.dsl.psu.edu/academics/certificate.cfm (last visited Feb. 12, 2008); University of Oregon, Conflict and Dispute Resolution Program, http://conflict.uoregon.edu/dual.html (last visited Feb. 12, 2008); Willamette University College of Law, Certificate Program in Dispute Resolution, http://www.willamette.edu/wucl/cdr/certificate/ (last visited Feb. 12, 2008); Quinnipiac University School of Law, Civil Advocacy and Dispute Resolution, http://law.quinnipiac.edu/x89.xml (last visited Feb. 12, 2008).

36. Virtually all of these schools have one of more faculty members with national reputations in the field.

37. ABA STANDARDS, *supra* note 18, Standard 302(b)(1).

38. *Id.* at Interpretation 302-5.

> problem; and, perhaps most critically, the student performance is subjected to intensive critical review.[39]

Although oft-cited, this description focuses more on method than form and fails to clarify the precise boundaries of a "clinic." This formulation appears to encompass the "three different branches of clinical education in the United States: in-house live-client clinics, externship programs, and simulation courses."[40] Other commentators view clinical education more narrowly and would refine the definition componentially:

> [A] law school clinical program would have six components. First, it is created through a law school with the intent that the program be integrally linked to the academic program of the institution. Second, law students, usually in their final years of law school, learn experientially by providing legal services or advice to real clients who qualify for representation by the law school's clinic. Third, those students are closely supervised by an attorney admitted to practice in the relevant jurisdiction, preferably by a member of the law school faculty or a private practitioner, who shares the pedagogical objectives of the clinical experience. Fourth, the clients served by the clinical program generally are not able to afford the cost of hiring private counsel, and they usually come from traditionally disadvantaged, underserved or marginal sectors of the community. Fifth, supervised case representation by students is preceded or accompanied by a pedagogical program that prepares students in what might be called theories of the practice of law. This would include components of substantive doctrine, skills, ethics, and values of law practice, and would be taught by a professor who knows the students' cases well enough to integrate that experience into the clinic classroom. Sixth, the students would receive academic credit toward graduation, hopefully for both the case and class-work they undertake as part of their participation in a clinic.[41]

---

39. *Report of the Committee on the Future of the In-House Clinic*, *supra* note 15, at 511.

40. Elliott S. Milstein, *Clinical Legal Education in the United States: In-House Clinics, Externships, and Simulations*, 51 J. LEGAL EDUC. 375, 376 (2001). Professor Milstein defines these three branches as follows:

> *In-house live-client clinics* are built around an actual law office, usually located in the law school, that exists for the purpose of providing students with a faculty-supervised setting within which to practice law and learn from the experience. Students learning in *externship programs* are placed in professional settings external to the law school, including law offices within governmental agencies and nongovernmental organizations. Law schools use the students' experience in those offices as the basis for teaching and learning. *Simulation* is a teaching method in which students are put into simulated lawyer roles to perform some aspect of the lawyering process in a controlled setting. Each of these uses the students' experiences as the subject matter for analysis, both within and outside the classroom.

Id. (footnotes omitted).

41. Richard J. Wilson, *Training for Justice: The Global Reach of Clinical Legal*

This debate,[42] while certainly of interest to academics, is not merely academic, and it has repercussions for a law school's curricular choices and categorical decisions. UT, for example, offers a number of "courses" that incorporate a clinical component, although they are not listed as part of our clinical program. Before the remarkable Fran Ansley[43] retired, she taught several courses at UT in which students collaborated with individuals in underrepresented communities to explore various dimensions of law and the legal system through research, education, or participation in mounting justice claims.[44] My ethics guru colleague Carl Pierce also has offered a course in which students supported the work of a Tennessee Bar Association committee charged with improving underrepresented parties' access to justice. Based upon the AALS definition of the term, all of these courses legitimately could be considered "clinics," yet UT did not categorize them as such.

Other law schools, however, may decide that similar curricular selections are more appropriately included on the roster of their clinical programs, and each law school likely would assign a unique descriptive label to similar course selections. While clinicians claim to be "[e]quipped with a 'common vocabulary' and a generally accepted definition of a methodology,"[45] a quick glance at the plethora of "clinics" offered by law schools in the United States reveals that the definitional/categorical issue is still largely unresolved, the topic addressed in the following section.

Education, 22 PENN ST. INT'L L. REV. 421, 423 (2004).

42. I deliberately have not addressed the debate that still rages about the role of service in clinical programs. Some, while acknowledging persistent arguments in favor of public service work, have nevertheless pronounced that debate settled. *See* Frederick M. Hart & J. Michael Norwood, *Key Parameters of the Clinical Method of Study*, *in* PROFESSIONAL EDUCATION IN THE UNITED STATES: EXPERIENTIAL LEARNING, ISSUES, AND PROSPECTS 86, 90–92 (Solomon Hoberman & Sidney Mailick eds., 1994). Clearly they are not reading the same material or attending the same conferences as am I.

43. I am not alone in my admiration for Fran; she was recently honored with the Society of American Law Teachers (SALT) Great Teacher Award. University of Tennessee College of Law, News and Events, http://www.law.utk.edu/news/AnsleySALT.htm (last visited Feb. 10, 2008).

44. Fran and our colleague Cathy Cochran of the UT Law Library faculty created a great website to exhibit a permanent collection of selected student projects from Fran's community-based field work courses. *See* University of Tennessee College of Law Student Field Projects in Community Law, http://www.law.utk.edu/Library/teachinglearning/default.html (last visited Feb. 12, 2008).

45. Barry et al., *supra* note 11, at 18.

C. Law School Clinical Offerings

If you don't know where you are going, you will wind up somewhere else.
Yogi Berra

The third lexical conundrum that I[46] encountered concerns the diversity of clinical offerings available to students at UT and other U.S. law schools. To collect the data reported in this Essay, I identified all of the clinical offerings listed on the websites of each of the top 100 U.S. law schools, as ranked by *U.S. News*. My goal was to catalogue the data to determine what types of clinics were widely available, what offerings were new and interesting, and what trends might appear.

What I discovered is that clinicians are a very creative and energetic bunch and that U.S. law students have access to numerous, richly diverse clinical opportunities. However, because there was no consistency to the way in which law schools refer to or denominate their clinical offerings, I was left to struggle to create somewhat arbitrary categories into which to fit each of the incredibly diverse programs that I found. This categorization was further complicated by each school's treatment of its externship and field placement programs.[47] Again, there did not appear to be an established methodology to differentiate between clinics and externships, and the ABA Standards provided little guidance.

---

46. "I" in this context is a truly misleading euphemism and used in the Royal "We" sense. Without the able assistance of my amazing administrative assistant, the poet Monica Miller, these data would still appear in raw form on each law school's individual website.

47. I often found it difficult to distinguish between "in-house clinics" and externships based upon a review of course names alone. However, the common conception of these terms is as follows. The term "in-house clinic" typically refers to programs in which students are certified by a state or federal court to act as lawyers for real clients with real legal problems under the close supervision of licensed attorneys, who act as counsel of record. *See* Milstein, *supra* note 40, at 18. The supervising attorneys may be faculty members or local lawyers who are adjuncts to the faculty. *See* Philip G. Schrag, *Constructing a Clinic*, 3 CLINICAL L. REV. 175, 186 (1996). The term "externship," on the other hand, refers to a program in which students perform legal work in various capacities in a governmental or nonprofit agency or office. *See* Milstein, *supra* note 40, at 380. For example, an externship student may work in a judge's chambers or a prosecutor's office. These students are supervised by lawyers regularly employed in the office. *See id.* Law school faculty maintain a role in externships, but typically are not directly involved in representing externship clients. *See id.* Clinics and externships offer students different experiences and opportunities. *See* J.P. Ogilvy, *Guidelines with Commentary for the Evaluation of Legal Externship Programs*, 38 GONZ. L. REV. 155, 159–60 (2003). "In-house clinics are usually organized to provide students with primary responsibility for a case, while externship students have that responsibility less often, depending on the nature of the placement." Harriet N. Katz, *Reconsidering Collaboration and Modeling: Enriching Clinical Pedagogy*, 41 GONZ. L. REV. 315, 318 (2006).

With those caveats in mind, I will attempt to report on the collected data.[48] First, depending upon how one counts, there are upwards of 523 separate in-house, live-client clinical offerings at the top 100 U.S. law schools. The most common clinics (ninety or so) involve some sort of civil practice, such as landlord-tenant and domestic relations work. Other civil offering designations include "Civil Litigation" clinics,[49] "Small Claims" clinics,[50] and "General Practice" clinics.[51]

Various types of criminal clinics are also common; there are between sixty and seventy-five, again depending upon how one counts. These clinics cover the criminal law waterfront, engaging in both prosecution and defense work, trial and appellate. Students can work on death penalty cases,[52] innocence projects,[53] inmate and family matters,[54] and parole issues,[55] among other topics.

Many schools—approximately thirty-six—list juvenile clinics as a separate offering. The subject matters vary, including abuse and neglect, delinquency, termination of parental rights, and criminal/juvenile justice.[56] Other clinical programs conduct legal work in support of children, including children's rights clinics[57] and special education advocacy.[58]

---

48. Unless otherwise noted, all comments in this Essay about clinical offerings at the top law schools are based on my analysis of the data I gathered through an extensive review of the schools' web sites. The list of the top 100 law schools is available at U.S. News & World Report, America's Best Graduate Schools 2008, Top Law Schools, http://www.usnews.com/usnews/edu/grad/rankings/law/lawindex.php (follow "Top Law Schools" hyperlink) (last visited Feb. 12, 2008) (on file with the Tennessee Law Review). My data are on file with the Tennessee Law Review.

49. *See, e.g.*, Boston University School of Law, Civil Litigation Program, http://www.bu.edu/law/prospective/jd/clinics/civil.html (last visited Feb. 12, 2008).

50. *See, e.g.*, University of San Diego School of Law, Small Claims Clinic, http://www.sandiego.edu/usdlaw/about/legalassist/clinics/studentinfo/smallclaims.php (last visited Feb. 12, 2008).

51. *See, e.g.*, CUA Columbus School of Law, General Practice Clinic, http://law.cua.edu/clinics/cle/clinics_general.cfm (last visited Feb. 12, 2008).

52. *See, e.g*, Duke University Law School, Death Penalty Clinic, http://www.law.duke.edu/deathpenalty/ (last visited Feb. 12, 2008).

53. *See, e.g.*, William and Mary School of Law, Innocence Project Clinic, http://www.wm.edu/law/academicprograms/curriculum/experiences/law747-01.shtml (last visited Feb. 12, 2008).

54. *See, e.g.*, Indiana University School of Law, Inmate Legal Assistance Project, http://www.law.indiana.edu/students/groups/ilap/index.shtml (last visited Feb. 12, 2008); Indiana University School of Law, Community Legal Clinic, http://www.law.indiana.edu/curriculum/programs/clinics/community_legal.shtml (last visited Feb. 12, 2008).

55. *See, e.g.*, McGeorge School of Law, On-Campus Clinics, http://www.mcgeorge.edu/x691.xml (last visited Feb. 12, 2008).

56. *See, e.g.*, University of Michigan Law School, Child Advocacy Law Clinic, http://www.law.umich.edu/centersandprograms/clinical/calc/Pages/default.aspx (last visited Feb. 12, 2008).

57. *See, e.g.*, University of Texas School of Law, Children's Rights Clinic, http://www.utexas.edu/law/academics/clinics/childrens/ (last visited Feb. 12, 2008).

Clinics related to immigration and refugee law; to community and economic development, nonprofits, and business planning; to ADR-related legal work; and to environmental law also are popular. Over thirty clinics pertain to immigration, refugee matters, or both,[59] and about forty clinics engage in some form or combination of nonprofit, economic development, or small business planning.[60] Roughly thirty-one clinics focus on negotiation, mediation, or ADR[61] and over twenty specialize in environmental work.[62]

The variety of available specialty clinics is inspiring, and I will just mention a few that caught my eye. Seizing an opportunity, several Iraqi Tribunal Clinics[63] and hurricane relief clinics[64] have been formed. Several gay, lesbian, bisexual, and transgender (GLBT) clinical programs,[65] an HIV/AIDS clinic,[66] and a vaccine injury clinic[67] also exist, as well as offerings devoted to

---

58. *See, e.g.*, Pepperdine University School of Law, Special Education Advocacy Clinic, http://law.pepperdine.edu/clinical/special_education_advocacy_clinic/ (last visited Feb. 12, 2008).

59. *See, e.g.*, University of Virginia School of Law, Clinics, http://www.law.virginia.edu/html/academics/clinics.htm#immigration (last visited Feb. 12, 2008).

60. *See, e.g.*, University of California Hastings College of the Law, Civil Justice Clinic, http://www.uchastings.edu/?pid=127 (last visited Feb. 12, 2008).

61. *See, e.g.*, University of Toledo College of Law, Dispute Resolution Clinic, http://www.utlaw.edu/students/clinics/disputeresolution.htm (last visited Feb. 12, 2008).

62. *See, e.g.*, University of South Carolina School of Law, Environmental Law Program, http://www.law.sc.edu/environmental/ (last visited Feb. 12, 2008).

63. *See, e.g.*, William and Mary School of Law, International Law Clinic: Iraqi Special Tribunal, http://www.wm.edu/law/academicprograms/curriculum/experiences/law748-01.shtml (last visited Feb. 12, 2008); *see also* University of Virginia School of Law, News and Events, Law Students Contribute to Iraqi Tribunals, www.law.virginia.edu/html/news/2006_fall/iraqiclinic.htm (last visited Mar. 31, 2008) (describing a clinical experience offered in the fall of 2006 and taught by a visiting professor).

64. *See, e.g.*, Boalt Hall School of Law, Promoting Human Rights Within the United States, http://www.law.berkeley.edu/clinics/ihrlc/rights_in_us.html (last visited Feb. 12, 2008); Yale Law School, Hurricane Relief Project, http://www.law.yale.edu/academics/1210.asp (last visited Feb. 12, 2008).

65. *See, e.g.*, Columbia Law: Clinics, http://www.law.columbia.edu/llm_jsd/grad_studies/courses/clinics/ (last visited Feb. 12, 2008); Legal Services of Harvard Law School: Gay, Lesbian, Bisexual and Transgender (GLBT) Law Clinic, http://www.law.harvard.edu/academics/clinical/lsc/clinics/gay.htm (last visited Feb. 12, 2008).

66. While not among the top 100 law schools in the *U.S. News* rankings, the University of the District of Columbia David A. Clarke School of Law is listed as having one of the top clinical programs. *See* Clinical Training, *supra* note 20. One of its clinics focuses on "provid[ing] comprehensive, holistic legal services to families with AIDS." University of the District of Columbia, David A. Clarke School of Law, HIV/AIDS Legal Clinic, http://www.law.udc.edu/programs/hiv/index.html (last visited Feb. 12, 2008).

67. *See* The George Washington University Law School, Vaccine Injury, http://www.law.gwu.edu/Academics/Clinical+Programs/Vaccine+Injury+Clinic.htm (last visited Feb. 12, 2008).

tobacco control,[68] investor justice,[69] the arts,[70] and sexual violence.[71] Other, more common specialty clinics include those devoted to community lawyering,[72] taxpayers,[73] health law,[74] international human rights,[75] legislative advocacy,[76] and administrative/government benefits representation, including disability,[77] workers' compensation, and welfare.

Of course, all of my musings are dependent upon my interpretive decision regarding how best to classify each law school's clinical offerings. It may be that I make much of a "unique" clinic that, in fact, is replicated at other law schools, lacking only an interesting, eye-catching name. This lack of a common lexicon for clinical offerings could be problematic in a number of contexts, not the least of which is the challenge that it presents to those humble souls collecting empirical data.

Another more serious concern is the lack of transparency for potential applicants. Law students are consumers and, like all consumers, may be susceptible to attractively packaged and cleverly marketed products.[78] While

---

68. *See, e.g.*, University of Maryland School of Law, Tobacco Control Clinic, http://www.law.umaryland.edu/course_info.asp?coursenum=534D (last visited Feb. 12, 2008).

69. *See, e.g.*, University of San Francisco, USF Law Clinics, http://www.usfca.edu/law/academics/shared-content/In-HouseClinics.html (last visited Feb. 12, 2008); Pace Law School: Investor Rights Clinic, http://www.pace.edu/page.cfm?doc_id=23714 (last visited Feb. 12, 2008).

70. *See, e.g.*, Seattle University School of Law, Clinical Law Courses, http://www.law.seattleu.edu/clinic/courses?mode=standard#artslegal (last visited Feb. 12, 2008).

71. *See* University of Washington School of Law, Sexual Violence and the Law Clinic, http://www.law.washington.edu/clinics/SexualViolence.html (last visited Feb. 12, 2008).

72. I identified seventeen such clinics. *See, e.g.*, Rutgers School of Law, Community Law Clinic, http://law.newark.rutgers.edu/clinics_community.html (last visited Feb. 12, 2008).

73. There are fifteen tax-related clinics. *See, e.g.*, American University Washington College of Law, Janet R. Spragens Federal Tax Clinic, http://www.wcl.american.edu/clinical/federal.cfm (last visited Feb. 12, 2008).

74. Fifteen clinics focus on health law. *See, e.g.*, University of San Diego School of Law, Mental Health Clinic, http://www.sandiego.edu/usdlaw/about/legalassist/clinics/studentinfo/mentalhealth.php (last visited Feb. 12, 2008).

75. There are fifteen clinics devoted to this subject. *See, e.g.*, Georgetown Law, The International Women's Human Rights Clinic (IWHRC), http://www.law.georgetown.edu/clinics/iwhrc/ (last visited Feb. 12, 2008). At UT's Symposium, Professor Susan Deller Ross described the amazing clinic she directs at Georgetown, the International Women's Human Rights Clinic. While a number of schools offer clinics involving international human rights, Professor Ross's particular focus on women's human rights appears unique.

76. I identified fourteen legislative advocacy clinics. *See, e.g.*, Yale Law School: Legislative Advocacy Clinic, http://www.law.yale.edu/academics/1217.asp (last visited Feb. 12, 2008).

77. Thirteen clinics focus on administrative and government benefits. *See, e.g.*, Penn State Dickinson School of Law: Disability Law Clinic, http://www.dsl.psu.edu/clinic/disability.cfm (last visited Feb. 12, 2008).

78. *See generally* Peter H. Bloch, *Seeking the Ideal Form: Product Design and Consumer Response*, J. MARKETING, July 1995, at 16 (1995) (discussing consumer response to attractive

law schools justifiably take pride in the breadth and depth of their clinical programs, it behooves neither law schools nor potential applicants to encourage enrollment based upon an uninformed besottedness with the name of a clinic. Appealing apperception does not necessarily convey the subject matter coverage accurately, nor does it always commensurate with clinical quality, particularly to the uninitiated. As a clinic's title may potentially be misleading, so too can the titles of its directors or instructors, the topic of the next section.

D. Law School Clinical "Faculty"

In theory there is no difference between theory and practice.
In practice there is.
Yogi Berra

My final thoughts pertain to the complexities associated with identifying the members of a law school's clinical faculty and to the reasons why definitional issues may contribute to this difficulty. One would assume that it would be a relatively easy task to locate the members of a law school's clinical faculty. Yet, for a variety of reasons, this is not so. For example, many law schools do not list their clinical faculty separately on faculty rosters, or they may not name a clinic director or instructors with the description of a clinic course in order to provide flexibility in staffing.

UT is a case in point, as the description of UT's Mediation Clinic does not identify me as the instructor.[79] Nor does UT's website distinguish in any way its full-time clinical faculty, all of whom are either tenured or tenure-track, from the remaining members of the full-time faculty. This, however, is not true at every law school, a fact that may complicate the nomenclature issue.

The debate continues over ABA Standard 405(c), which states that:

> A law school shall afford to full-time clinical faculty members a form of security of position reasonably similar to tenure, and non-compensatory perquisites reasonably similar to those provided other full-time faculty members. A law school may require these faculty members to meet standards

---

product form).

79. University of Tennessee College of Law, Clinical Programs: Mediation Clinic, http://www.law.utk.edu/departments/CLINIC/clinicmediation.htm (last visited Feb. 12, 2008). While I have directed this clinic for several years, it was created by the much honored Grayfred Gray and has attracted several other excellent directors during its existence. Grayfred is a legend in mediation circles in Tennessee and beyond. He recently received the first annual "Grayfred Gray Public Service Mediation Award," an award named in honor of "his original and lasting contributions to mediation awareness in Tennessee." Lipscomb University, Mayor Celebrates Mediation Day with ICM on Campus, http://news.lipscomb.edu/filter.asp?SID=14&fi_key=724&co_key=12551 (last visited Feb. 12, 2008). He retired from the UT faculty in 2001 and currently serves as the Training Director of the Lancaster Mediation Center. *See* Lancaster Mediation Center, Who We Are, http://www.lancmed.org/staff.html (last visited Feb. 12, 2008).

> and obligations reasonably similar to those required of other full-time faculty members. However, this Standard does not preclude a limited number of fixed, short-term appointments in a clinical program predominantly staffed by full-time faculty members, or in an experimental program of limited duration.[80]

Kim Diane Connolly raised this issue in her thought-provoking discussion at the Symposium,[81] reminding us that some law schools still maintain some form of differentiated tenure for clinical faculty or offer clinic instructors contracts that comply with the last sentence of ABA Standard 405(c).[82] If a law school lists only full-time, tenure-track faculty on its website or other routes of public access, clinic faculty may be impossible to identify.

Another complication may arise when law school doctrinal or classroom faculty[83] also teach in the clinic. I am one such faculty member. While I direct UT's superb Mediation Clinic, I do not teach full-time in the clinic, nor did I

---

80. ABA STANDARDS, *supra* note 18, Standard 405(c).

81. Video recording: Charles Miller Legal Clinic 60th Anniversary Celebration, held by the University of Tennessee College of Law, Session IV: Expanding Clinical Experiences (Sept. 15, 2007), *available at* http://mediabeast.ites.utk.edu/mediasite4/Catalog/ (follow "Charles Miller Legal Clinic—60th Anniversary Celebration" hyperlink in sidebar, then follow "Session IV" hyperlink) [hereinafter Video recording, Session IV].

82. Professor Barry and her colleagues discuss this issue in more detail:

> Presently, there is tenure or contract status information on 789 clinicians, with 134 out of 183 schools reporting at least one clinician who was tenured or on tenure-track; thirty-one schools reporting at least one clinician who was clinical tenured or on the clinical tenure track; seventy-one schools reporting at least one clinician who was on a long-term contract; and 112 schools reporting at least one clinician who was on a short-term contract.
>
> With respect to tenure, 245 clinicians reported that they had tenure, and ninety-three clinicians reported that they were tenure-track but had not yet attained tenure. In addition, twenty-nine clinicians reported that they had clinical tenure and twenty-five clinicians reported that they were clinical tenure-track but had not yet attained clinical tenure. In terms of those clinical faculty with contract rather than tenure status, 161 clinicians reported that they were on long-term contracts of three years or more, and 236 clinicians reported that they were on short-term contracts.

Barry et al., *supra* note 11, at 31 (footnotes omitted). *See generally* Peter A. Joy & Robert R. Kuehn, *The Evolution of ABA Standards for Clinical Faculty*, 75 TENN. L. REV. 183 (2008) (providing a detailed history of the evolution of the ABA Standards related to clinical faculty tenure and participation in law school governance).

83. Professor Laura Rovner chose the term "classroom faculty" to describe faculty who do not teach in the clinic, explicitly acknowledging that the term "may not accurately reflect how some such faculty view themselves and their teaching." Laura L. Rovner, *The Unforeseen Ethical Ramifications of Classroom Faculty Participation in Law School Clinics*, 75 U. CIN. L. REV. 1113, 1114 n.1 (2007). In so identifying this group of law faculty, Professor Rovner confronted her own lexical challenges as she considered alternative labels, including "stand-up," "podium," and "doctrinal"—all of which she decided were problematic in one way or another, for example, because these terms imply "that clinical teaching is not theoretical, academic, doctrinal, etc." *Id.*

specialize in trial work in law school or in practice. Because of my law school experience and my professional conditioning, I somehow feel that I have not earned the right to have the title of "clinician" bestowed upon me, and I still pinch myself with glee every time that the Clinic faculty here claim me as their own. And I am not alone among many traditional "classroom instructors" joining the clinical fray. Indeed, there appears to be an increasing and varied level of participation by classroom faculty in the work of law school clinical programs,[84] a trend for which several causes have been ascribed and of which there may be various consequences.

ABA Standard 302(b)(1)[85] has been identified as one possible cause of the increasing involvement of classroom instructors in law school clinics. The ABA Standard's mandate that law schools "offer substantial opportunities for . . . live-client or other real-life practice experiences"[86] has placed enormous demands not only upon law school clinical faculty, but also upon administrations and budgets. Compliance with the Standard requires creative solutions, one of which may be to draw upon all available resources to do so, including classroom instructors. My colleague Carl Pierce discussed this concept during his Symposium presentation, in which he urged the expansion of collaborative clinical offerings between classroom and clinical faculty.[87] Carl mentioned a number of interesting possibilities, including a clinical collaboration between a family law instructor and a mediation clinical supervisor in which students would offer mediation services to unrepresented divorcing parties and provide "limited scope" representation to assist them in filing any resulting settlement agreements.[88]

Collaborative endeavors such as these would be a wonderful, enriching experience for students. However, scholars have warned of serious consequences from classroom faculties' participation in law school clinical programs. For example, this collaboration could "raise[] important . . . ethical issues that may significantly affect faculty, students and clients."[89]

One such ethical issue is the possibility that classroom faculty could be civilly or criminally liable for the unauthorized practice of law.[90] Classroom

84. *Id.* at 1114.

85. ABA STANDARDS, *supra* note 18, Standard 302(b)(1).

86. *Id.*

87. Video recording, Session IV, *supra* note 81.

88. *Id.*

89. Rovner, *supra* note 83, at 1114. Furthermore, Professor Rovner explores another definitional conundrum in her article: whether a law school clinic constitutes a law firm as described by Model Rule of Professional Conduct 1.0(c). *See* Rovner, *supra* note 83, at 1122–23. *See generally* MODEL RULES OF PROF'L CONDUCT R. 1.0(c) (6th ed. 2007) (defining "firm" and "law firm"). This question is relevant to her exploration of the ethical issues inherent in involving classroom instructors in clinical programs. *See generally* Rovner, *supra* note 83, at 1143–69 (discussing the ethical ramifications of classroom faculty participation in clinical courses). As it exceeds the scope of my Essay, I commend Professor Rovner's Article to the interested reader's attention.

90. *See* Rovner, *supra* note 83, at 1143–50.

faculty who are not currently licensed to practice in the state in which their school is located could potentially face criminal charges for violating the state's unauthorized-practice-of-law statute.[91] Classroom faculty also could be subjected to disciplinary charges, both in the state in which their school is located and the state or states in which they are licensed.[92] Further, clinical faculty colleagues also could be penalized, as ethical rules forbid licensed attorneys from assisting a person in the unauthorized practice of law.[93]

Additionally, the involvement of classroom faculty in a law school clinic could result in breaches of client confidentiality or waiver of privileges.[94] Depending upon the specific role that classroom teachers play in a clinic—whether they act as practicing attorneys or as consultants or experts—the clinic and the classroom teachers must remain mindful not only of the obligations of confidentiality,[95] but also of attorney-client[96] and work product[97] privileges.

Conflicts of interest also may be a concern when classroom faculty participate in clinical programs, sometimes even requiring disqualification of the clinic as counsel for the client.[98] The particular conflicts that may arise as a result of collaborations between clinical and classroom faculty reflect the atypical structure of the law clinic as compared to other law firms[99] and make it very difficult to provide meaningful procedures for conflict checks. It is easy to imagine a classroom instructor being approached by a student with a work-related question that involves a matter to which the clinic might be adverse, particularly in smaller communities.[100]

Further, while classroom faculty bring a wealth of substantive legal expertise and instruction experience to their clinical teaching, they often have little or no experience in litigation or case management. This poses a number of problems for clinical administrators, students, and clients, not to mention for transitioning classroom instructors. First, "[n]o clinician wants clients to suffer or students to be embarrassed because the supervisors as well as the student are utter novices in the clinic's area of practice."[101] To avoid this situation, new clinical teachers should receive training in the subject matter of the clinic as

---

91. *Id.* at 1144.

92. *Id.* at 1144–45.

93. *Id.* Commentators have noted the irony of this result when one considers the high regard in which law professors are often held within the legal profession. *Id.* at 1145 (quoting Jett Hanna, *Moonlighting Law Professors: Identifying and Minimizing the Professional Liability Risk*, 42 S. TEX. L. REV. 421, 433–34 (2001)).

94. *See* Rovner, *supra* note 83, at 1150–65.

95. *See id.* at 1150–56.

96. *See id.* at 1156–62.

97. *See id.* at 1162–65.

98. *See id.* at 1165–69.

99. *Id.* at 1166.

100. *See id.* at 1166–68.

101. Schrag, *supra* note 47, at 211.

well as in the practice norms and rules of the courts in which they will be practicing and supervising students.[102]

Also, as one clinical scholar describes, "the process of teaching litigation (or other skills) is rather different from handling cases."[103] Thus, to guide and supervise students effectively, new clinicians should receive some training in clinical teaching.[104] Fortunately, this type of training is readily available. For example, the AALS offers training courses in clinical teaching.[105] Additionally, "[s]hort courses in the practice of nearly every kind of law are offered frequently by local bar associations, non-profit legal aid and other advocacy organizations, and specialized training groups such as the National Institute of Trial Advocacy."[106] It also may be feasible to allow transitioning faculty to observe proceedings of the type that they will encounter or to actually handle one or two cases, alone or with a more experienced practitioner.[107] Unlike my experience at UT, not all law schools offer their clinical faculty this type of preparation. Students at those schools may not find their clinical experiences as rewarding as do students at schools where clinical faculty have extensive experience with the subject matter and clinical teaching.

These concerns apply equally to the tireless and talented adjuncts upon whom schools often rely to provide support and supervision in our clinical programs. Returning to this Essay's definitional theme: How does one classify these talented adjuncts with regard to their faculty status? The ABA Standards would appear to allow adjunct faculty to supervise clinic students,[108] but most schools do not list adjuncts (or at least do not identify them as such) on their clinic websites, nor do they involve them in regularly scheduled faculty meetings or planning sessions.[109]

Most commentators acknowledge that "no resource is as critical [to a clinical program] as the teaching and support staff."[110] Accordingly, while my onomastic obsession may appear to be, well, obsessive, these lexical ambiguities are significant.

---

102. *See id.* at 211–12.

103. *Id.* at 187.

104. *Id.*

105. *Id.*

106. *Id.* at 212.

107. *Id.*

108. ABA STANDARDS, *supra* note 18, Standard 405(c).

109. *Cf.* Schrag, *supra* note 47, at 186, 188–89 (noting the possibility of adjunct involvement and discussing "two possible competing models for relations among clinic staff[:] the hierarchical model and the collaborative model"); ABA STANDARDS, *supra* note 18, Interpretation 405-8 ("A law school shall afford to full-time clinical faculty members participation in faculty meetings, committees, and other aspects of law school governance in a manner reasonably similar to other full-time faculty members. This Interpretation does not apply to those persons referred to in the last sentence of Standard 405(c) [i.e., short-term, non-faculty clinic staff].").

110. Schrag, *supra* note 47, at 186.

III. CONCLUSION

It ain't over 'til it's over.
Yogi Berra

While I learned much during the UT Legal Clinic's sixtieth anniversary Symposium, I still have not resolved in my mind the question of how one defines and labels a clinical experience—or who that "one" should be. As I hope that I have suggested, the delicately nuanced taxonomy of clinic labels and the identification of clinical faculty are not merely "dancing on the head of a pin"[111] exercises and have implications beyond those academic.

That being said, what is most significant to me is the number and variety of clinics devoted to serving the underserved: the poor, the elderly, vulnerable youth, those in need of mental health services, farm workers, Native Americans, inmates, and ex-offenders, to name just a few. And, while there are those who claim that the AALS and ABA Guidelines for Clinical Legal Education settled the "service" versus "educational objective" issue,[112] clinicians not only teach students knowledge and skill, they also integrate valuable ethical and social concerns into the clinical experience. Bridget M. McCormack raised a similar point in her Symposium presentation;[113] participation in a law school clinic instills a sense of professionalism in students that cannot be learned or experienced in a classroom environment or simulated setting. The recent Carnegie Report also addresses the critical role that clinical education plays in teaching the ethical demands of practice and the virtues of a socially responsible practice of law, noting that "[c]linics can be a key setting for integrating all the elements of legal education, as students draw on and develop their doctrinal reasoning, lawyering skills, and ethical engagement, extending to contextual issues such as the policy environment."[114]

Clinicians are educators and public servants in the very real sense of those words. Their work impacts a broad spectrum of national and international issues as well as the lives and professional development of their students.

---

111. For those interested in esoterica, this reference, too, suffers from some interpretive re-categorization. Most of us have heard the phrase "angels dancing on the head of a pin." However, the actual quote appears to be, "[S]ome who are far from Atheists, may make themselves merry, with that *Conceit of Thousands of Spirits, dancing at once upon a Needle[']s Point.*" RALPH CUDWORTH, THE TRUE INTELLECTUAL SYSTEM OF THE UNIVERSE 778 (Garland Publishing 1978) (1678).

112. *See, e.g.*, discussion *supra* note 42 and accompanying text.

113. Video recording: Charles Miller Legal Clinic 60th Anniversary Celebration, held by the University of Tennessee College of Law, Session V: The Future of Clinics and the Law School Curricula (Sept. 15, 2007), *available at* http://mediabeast.ites.utk.edu/mediasite4/Catalog/ (follow "Charles Miller Legal Clinic—60th Anniversary Celebration" hyperlink in sidebar, then follow "Session V" hyperlink).

114. SULLIVAN ET AL., *supra* note 17, at 121, 132.

However unworthy I feel to be included as part of UT's clinical faculty, I am honored and proud to be a part of this amazing group, however lexically categorized. I also am proud to have been a part of UT's sixtieth anniversary celebration and hope to be around for at least its seventy-fifth. Go Vols!

CONTRACTING OUT OF PROCESS, CONTRACTING OUT OF CORPORATE ACCOUNTABILITY: AN ARGUMENT AGAINST ENFORCEMENT OF PRE-DISPUTE LIMITS ON PROCESS

MEREDITH R. MILLER*

"CORPORATION, *n.* An ingenious device for obtaining individual profit without individual responsibility."[1]

"ACCOUNTABILITY, *n.* The mother of caution."[2]

INTRODUCTION

Should a cable television company be accountable to millions of customers for automatically charging unexplained fees for unnecessary services? Should an employer be liable for discriminating against an employee on the basis of her national origin? Should a nursing home corporation be responsible when an employee's negligence injures an elderly tenant? Should a fast food corporation be required to answer to hundreds of defrauded investors who had dreams of setting up their own restaurant franchises? The resounding answer to these rhetorical questions should be "yes." However, the law has elevated a mythical notion of contractual autonomy at the expense of corporate social accountability. With the law of corporations and of civil procedure both deferring to contract law, the arbitration trend has invited corporations to contract around process and, thus, accountability.

In the field of corporate law, the "nexus of contract" model is the dominant theoretical explanation of the law concerning the management of corporations.[3]

---

* Assistant Professor of Law, Touro College, Jacob D. Fuchsberg Law Center. My gratitude extends to Dean Larry Raful, who provided generous summer support for this project. Thanks are also due to Marjorie Silver, Jack Graves, Sheila Scheuerman, Katherine Traylor Schaffzin and Kerri Lynn Stone for their insightful comments on earlier drafts. And, thanks to Roy Sturgeon, James Dougherty and Isaac Samuels for their tireless research assistance. Any errors are (of course) my own.

1. AMBROSE BIERCE, THE DEVIL'S DICTIONARY 28 (Hill & Wang 1957) (1911), *available at* http://www.alcyone.com/max/lit/devils. Reference to these definitions is not intended as an endorsement of certain other definitions in *The Devil's Dictionary* that, written at the turn of the century, are racist, sexist and, in certain cases, just plain offensive.

2. *Id.* at 5.

3. *See* STEPHEN M. BAINBRIDGE, CORPORATION LAW AND ECONOMICS 199–200 (2002) ("The dominant model of the corporation in legal scholarship is the so-called nexus of contracts

Under this view, corporations are nothing more than a network of contracts between voluntary, private actors.[4] Moreover, under the traditional, theoretical description of corporate law management, corporate managers are only accountable to the firm's shareholders to maximize the return on their investments.[5] Accordingly, corporations[6] (really, their managers) owe no general responsibility to other constituencies affected by the firm's activities.[7] These constituencies, often referred to by corporate law theorists as "stakeholders," include the firm's employees, consumers, suppliers, and communities generally.[8] Proponents of stakeholder theory argue that corporate management should have at least some generalized accountability to constituencies other than shareholders.[9] However, proponents of the prevailing, traditional view of shareholder wealth maximization and the contractarian theorists of the firm, argue that, absent legislation, a corporation's general obligations to stakeholders are defined only by its contracts with these constituencies.[10]

Further, in the field of civil procedure, arbitration has taken a strong foothold. Since Congress enacted the Federal Arbitration Act[11] (FAA) in 1925, and the Supreme Court's subsequent, repeated pronouncement of a "liberal federal policy favoring arbitration agreements,"[12] corporations have used arbitration with increasing frequency.[13] Under the FAA, arbitration agreements "shall be valid, irrevocable, and enforceable, save upon such grounds as exist at law or in equity for the revocation of any contract."[14] Here, again, a body of law defers to contract. Indeed, Professor Judith Resnick notes that the trend in

---

theory."); KENT GREENFIELD, THE FAILURE OF CORPORATE LAW 17 (2006) (describing the traditional view that corporations are seen as "an intricate network of contracts").

4. *See* GREENFIELD, *supra* note 3, at 16–17.

5. *See* Tara J. Radin, *Stakeholders and Sustainability: An Argument for Responsible Corporate Decision-Making*, 31 WM. & MARY ENVTL. L. & POL'Y REV. 363, 375–88 (2007) ("The prevalent view of the firm characterizes it as merely a vehicle for profit maximization.").

6. The term "corporation" throughout this Article may, at times, loosely refer to any form of incorporated or uninicorporated business entity, such as a limited liability company or partnership.

7. *See* Radin, *supra* note 5, at 375–77.

8. *See id.* at 381–84 (discussing the stakeholder view of a corporation).

9. *Id.*

10. *See* GREENFIELD, *supra* note 3, at 16–18. This model is a normative, theoretical construct; however, it has significantly affected the law's evolution.

11. Federal Arbitration Act, 9 U.S.C. §§ 1–16 (2000 & Supp. V 2007).

12. Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 25 (1991); Moses H. Cone Mem'l Hosp. v. Mercury Constr. Corp., 460 U.S. 1, 24 (1983).

13. Indeed, pre-dispute arbitration has become so widely used that some scholars have warned of an "end of law." Rex R. Perschbacher & Debra Lyn Bassett, *The End of Law*, 84 B.U. L. REV 1, 28–32 (2004). *See generally* Charles L. Knapp, *Taking Contracts Private: The Quiet Revolution in Contract Law*, 71 FORDHAM L. REV. 761, 765 (2002) (discussing the "privatization of American contract law").

14. Federal Arbitration Act, 9 U.S.C. § 2 (2000).

civil procedure has been "the wholesale application of extant principles of contract law."[15] It has even been suggested that pre-dispute arbitration clauses have been elevated to a status of "super contract" because of their "near-automatic enforcement by means of specific performance."[16]

With procedural and corporate law so readily deferring to contract, it seems appropriate to consider what these contracts actually look like. This examination paints a troubling picture—one of corporations using the pre-dispute arbitration regime to contract around process in an attempt to insulate themselves from any potential responsibility they might otherwise take on by virtue of their contractual relationships with stakeholder constituencies.[17] As an outgrowth of the trend toward arbitration, corporations have increasingly used standardized forms with provisions that expressly limit[18] significant procedural rights of the other contracting parties—often, for example, franchisees, borrowers, consumers, employees, and insureds. This Article will focus on the express, pre-dispute procedural limitations of "collective action waivers,"[19]

---

15. *See generally* Judith Resnik, *Procedure as Contract*, 80 NOTRE DAME L. REV. 593, 598–99 (2005) (describing how changes in adjudicatory practice are shifting the focus of civil procedure from "due process procedure" to "contract procedure").

16. David H. Taylor & Sarah M. Cliffe, *Civil Procedure By Contract: A Convoluted Confluence of Private Contract and Public Procedure in Need of Congressional Control*, 35 U. RICH. L. REV. 1085, 1088 (2002).

17. In this Article, "stakeholder" refers generally to those individuals or groups affected by the operations of a corporation, including, a corporation's employees, consumers, suppliers, franchisees, insureds, debtors, creditors, and communities generally. *See* Radin, *supra* note 5, at 381–88. This Article examines corporations' arbitration agreements with these constituencies. In the arbitration context, other scholars have used the term "consumer" liberally in much the same way. *See* Richard E. Speidel, *Consumer Arbitration of Statutory Claims: Has Pre-Dispute [Mandatory] Arbitration Outlived Its Welcome?*, 40 ARIZ. L. REV. 1069, 1072–74 (1998) (describing the "consumerization" of arbitration). A stakeholder could also be another business entity, or even a competitor. However, recent empirical studies show that corporations use mandatory arbitration provisions in less than 10% of their material nonconsumer and nonemployment contracts, compared to the use of mandatory arbitration in over 75% of their consumer contracts. Theodore Eisenberg et al., *Arbitration's Summer Soldiers: An Empirical Study of Arbitration Clauses in Consumer and Nonconsumer Contracts* 15 (Cornell Law School Legal Studies Research Paper Series), *available at* http://ssrn.com/abstract=1076968 (last visited Apr. 19, 2008).

18. The word "waiver" in this context is a misnomer. A waiver is a "voluntary [and] knowing relinquishment of a right." Green v. United States, 355 U.S. 184, 191 (1957) ("'Waiver' is a vague term used for a great variety of purposes . . . . In any normal sense, however, it connotes some kind of voluntary knowing relinquishment of a right."). Because the law has imported contractual standards of assent to arbitration clauses, a voluntary and knowing relinquishment is not required but, rather, some action (or inaction) that can be interpreted objectively as a manifestation of assent. *See* Taylor & Cliffe, *supra* note 16, at 1104 (discussing the term "waiver" as a misnomer in this context); discussion *infra* Part III.

19. The term is interchangeably described in case law and scholarship as "no class action waivers," "class action waivers," "collective action waivers," and "class action preclusion clauses." This Article borrows the term "collective action waiver" from Professor Myriam

discovery limitations, and shortened statutes of limitation. These express, pre-dispute limitations[20] may, in effect, work to create a barrier to the enforcement of substantive laws concerning, for instance, consumer and employee protection, civil rights, and common law negligence.[21] Thus, these contractual limitations have been aptly analogized to exculpatory clauses.[22] Further, to the extent that pre-dispute limitations are inserted in arbitration clauses that are used by various industries in standard form agreements, they have effectively become the legislation governing contractual relationships of corporations. These pre-dispute procedural limitations have, in essence, provided an opportunity for corporations to flout legislative and social policy and "deregulate themselves"[23] through contract.

Because the FAA defers to existing contract defenses, the enforceability of arbitration terms has largely been determined under the doctrine of

---

Gilles because, as she explains, these clauses waive any right to bring a class action or class arbitration or otherwise proceed collectively. Myriam Gilles, *Opting Out of Liability: The Forthcoming, Near-Total Demise of the Modern Class Action*, 104 MICH. L. REV. 373, 376 n.15 (2005).

20. This Article addresses the contractual terms that determine which characteristics of civil adjudication will or will not comprise part of the arbitration proceeding. Forum selection and choice of law clauses are generally beyond the scope of this discussion, but they are certainly at its periphery. *See infra* notes 267, 281–87 and accompanying text. These clauses are as problematic as pre-dispute arbitration agreements and, likewise, can have the effect of denying a party's vindication of substantive rights. *See generally* Linda S. Mullenix, *Another Easy Case, Some More Bad Law: Carnival Cruise Lines and Contractual Personal Jurisdiction*, 27 TEX. INT'L L. J. 323, 325–27 (1992) (arguing that forum-selection clauses in consumer adhesion contracts are unconscionable); Linda S. Mullenix, *Another Choice of Forum, Another Choice of Law: Consensual Adjudicatory Procedure in Federal Court*, 57 FORDHAM L. REV. 291, 295 (1988) (explaining how the doctrine of consensual adjudicatory procedure advances "purely prudential considerations" at the expense of "substantial litigation rights").

21. *See generally* Paul D. Carrington, *The Dark Side of Contract Law*, 36 TRIAL 73, 73 (2000) ("The contemporary fashion is not to require the weaker party to surrender substantive rights, but to require him or her to surrender procedural rights needed if the substantive rights are to retain their value . . . .").

22. *See* Discover Bank v. Superior Court, 113 P.3d 1100, 1109 (Cal. 2005); Scott v. Cingular Wireless, 161 P.3d 1000, 1006 (Wash. 2007) (en banc); *see also* Samuel Issacharoff & Erin F. Delaney, *Credit Card Accountability*, 73 U. CHI. L. REV. 157, 17–82 (2006) (discussing the decision to hold a credit card agreement's class action waiver as an exculpatory clause violative of public policy). "Exculpatory clause" is defined and used here to describe "[a] contractual provision [prospectively] relieving a party from liability resulting from a negligent or wrongful act." BLACK'S LAW DICTIONARY 608 (8th ed. 2004).

23. David S. Schwartz, *Enforcing Small Print to Protect Big Business: Employee and Consumer Rights Claims in an Age of Compelled Arbitration*, 1997 WIS. L. REV. 33, 37 (1997) ("The enforcement of adhesive arbitration clauses allows firms to lessen the regulatory impact of statutory claims—in short, to deregulate themselves."). Corporations seek to externalize risks and maximize profits for shareholders. Naturally, then, they readily use liability-limiting contract provisions. Therefore, a clear line should be drawn to articulate which terms in pre-dispute arbitration clauses will and will not be enforced.

unconscionability.[24] This case-by-case treatment of problematic clauses has, as Professor Arthur Leff predicted forty years ago, "substituted a highly abstract word 'unconscionable' for the possibility of more concrete and particularized thinking about particular problems of social policy."[25] Indeed, rather than stating a policy against such express, pre-dispute limitations, the use of the unconscionability doctrine has lead to a patchwork of irreconcilable decisions and unpredictability in contract drafting.[26]

Many well-articulated and convincing critiques have been aimed at "mandatory" arbitration,[27] and some equally strong counterarguments have also been made.[28] Moreover, some scholars have criticized arbitration itself as an implicit "waiver" of procedural rights such as the right to have a dispute heard by a jury and the right to an appeal.[29] Indeed, presently before Congress is

---

24. *See* discussion *infra* Part I.B.

25. Arthur Allen Leff, *Unconscionability and the Code—The Emperor's New Clause*, 115 U. PA. L. REV. 485, 515 (1967); *see also* Knapp, *supra* note 13, at 797 (citing Leff, *supra*); Jeffrey W. Stempel, *Arbitration, Unconscionability, and Equilibrium: The Return of Unconscionability Analysis as a Counterweight to Arbitration Formalism*, 19 OHIO ST. J. ON DISP. RESOL. 757, 763 n.21 (2004) (citing Leff, *supra*).

26. *See* discussion *infra* Part III.

27. For example, compelling arguments have been made that the FAA was never intended to apply to consumer and employee contracts but, rather, was intended only to govern commercial relationships between business entities. *See, e.g.*, Schwartz, *supra* note 23, at 75–81 (arguing that Congress did not intend the FAA to be given such broad interpretation by the courts). Furthermore, strong arguments have been made that the Supreme Court has misinterpreted the FAA as a statement of a liberal public policy favoring arbitration. *See id.* at 81–109; Jean R. Sternlight, *Panacea or Corporate Tool?: Debunking the Supreme Court's Preference for Binding Arbitration*, 74 WASH. U. L.Q. 637, 644–674 (1996). In the consumer and employment contexts, there have been numerous calls for reform. *See generally, e.g.*, Richard M. Alderman, *Pre-Dispute Mandatory Arbitration in Consumer Contracts: A Call for Reform*, 38 HOUS. L. REV. 1237 (2001) (focusing on the shortcomings of arbitration in consumer transactions); Senator Russell D. Feingold, *Mandatory Arbitration: What Process is Due?*, 39 HARV. J. ON LEGIS. 281, 281–82 (2002) (examining how "thousands of people . . . are being deprived of their rights to go to court by mandatory, binding arbitration clauses" within employment contracts, franchise agreements, and consumer credit agreements); Speidel, *supra* note 17 (discussing whether mandatory arbitration has outlived its usefulness in the areas of consumer transactions and employment contracts); Jean R. Sternlight, *Mandatory Pre-Dispute Arbitration: Steps Need to be Taken to Prevent Unfairness to Employees, Consumers*, DISP. RESOL. MAG., Fall 1998, at 5 (observing that mandatory binding arbitration clauses "may permit a knowledgeable and powerful entity to trick or coerce individuals into effectively waiving their rights under federal or state law").

28. *See generally* Christopher R. Drahozal, *"Unfair" Arbitration Clauses*, 2001 U. ILL. L. REV. 695, 698 (2001) (questioning anecdotal criticisms of "unfair" arbitration clauses and arguing that "fully informed individuals [may] benefit by agreeing to arbitration clauses that appear unfair"); Eric J. Mogilnicki & Kirk D. Jensen, *Arbitration and Unconscionability*, 19 GA. ST. U. L. REV. 761 (2003) (contending that "arbitration is fair to individuals and provides benefits unavailable in traditional litigation").

29. *See* Carole J. Buckner, *Due Process in Class Arbitration*, 58 FLA. L. REV. 185, 216

proposed legislation titled "the Arbitration Fairness Act of 2007," that would bar pre-dispute arbitration clauses in the consumer, franchise, and employment contexts.[30] Maligned as the plaintiff bar's "pro-lawsuit legislation,"[31] however, the Arbitration Fairness Act is not predicted to pass. Consequently, across varying industries, the pre-dispute arbitration regime endures unheedingly. Thus, this Article sets aside the arguments aimed generally at pre-dispute arbitration clauses and, instead, sets its sights on some of the terms that arise in such clauses. More specifically, this Article focuses on the appended, additional procedural limitations often contained[32] within arbitration clauses.[33]

(2006) ("[C]ourts addressing this issue hold that, by agreeing to arbitration, parties effectively waive the right to insist upon procedural due process and other constitutional rights that would be required if a state actor were involved.") (citing Edward Brunet, *Arbitration and Constitutional Rights*, 71 N.C. L. REV. 81, 102 (1992) ("'The orthodox view holds that parties who consent by contract to arbitration expressly waive their constitutional rights.'")); Paul H. Dawes, *Alternative Dispute Resolution, in* SECURITIES LITIGATION 1999, at 599, 603 (PLI Corp. Law & Practice Course, Handbook Series No. B0-00DM, 1999) ("The risks [of arbitration], broadly speaking, can be grouped into three major concerns: lack of appeal rights, waiver of other procedural and substantive rights and, ironically, a perception that like jurors, arbitrators can be unpredictable, under-qualified and swayed by emotion."); Stephen J. Ware, *Domain-Name Arbitration in the Arbitration-Law Context: Consent to, and Fairness in, the UDRP*, 6 J. SMALL & EMERGING BUS. L. 129, 153 (2002) ("An arbitration agreement . . . is a waiver of many of the procedural rights guaranteed in litigation."); *see also* Flores v. Evergreen at San Diego, LLC, 55 Cal. Rptr. 3d 823, 832 (Ct. App. 2007) ("[A]rbitration agreements waive important legal rights . . . .").

30. Arbitration Fairness Act of 2007, S. 1782, 110th Cong. § 3 (1st Sess. 2007) (proposing to amend the Federal Arbitration Act to invalidate pre-dispute agreements to arbitrate franchise, consumer and employment disputes); Arbitration Fairness Act of 2007, H.R. 3010, 110th Cong. § 3 (1st Sess. 2007) (proposing identical amendments as S. 1782).

31. Editorial, *Party at Ralph's*, WALL ST. J., Nov. 7, 2007, at A22; *see also* Joan Claybrook, Editorial, *Party at Joan's*, WALL ST. J., Nov. 17, 2007, at A9 (clarifying that the consumer group Public Citizen opposes *mandatory*, not voluntary, arbitration); Editorial, *No Lawyers, Please*, WALL ST. J., April 5, 2008, at A8 (arguing against enactment of Arbitration Fairness Act).

32. This Article is not based on an empirical analysis concerning the frequency with which pre-dispute limitations are included or enforced. The plethora of cases that have arisen in the past decade, however, suggests that the use of express pre-dispute limitations, or at least the litigation of such clauses, is a growing trend.

33. Moreover, this discussion encapsulates procedural limitations contained outside of arbitration clauses—though, in reality, they are overwhelmingly used in the context of arbitration clauses. Drawing on the developments in arbitration practice, recent scholarship has imagined a system of contractually modified litigation. *See* Elizabeth Thornburg, *Designer Trials*, 2006 J. DISP. RESOL. 181, 211 (2006). *See generally* Michael L. Moffitt, *Customized Litigation: The Case for Making Civil Procedure Negotiable*, 75 GEO. WASH. L. REV. 461, 461 (arguing that customized litigation advances justice, promotes efficiency, and increases public accessibility to civil trials to a greater degree than do current procedural rules); Henry S. Noyes, *If You (Re)Build It, They Will Come: Contracts to Remake the Rules of Litigation in Arbitration's Image*, 30 HARV. J.L. & PUB. POL'Y 579, 583–4 (2007) ("[M]odified litigation has

Part I of this Article provides a brief background of the rise of pre-dispute arbitration. Part II discusses the pre-dispute limitation devices of collective action waivers, discovery limitations, and shortened statutes of limitation. Part III explores some of the concerns raised by express, pre-dispute limitations on procedural rights. The use of the unconscionability doctrine to police these terms is discussed in each of the first three Parts. Finally, by analogy to the treatment of exculpatory clauses and to section 195 of the Restatement (Second) of Contracts, Part IV argues that federal legislative reform should specify that certain express limits on procedural rights contained in standardized form agreements are per se invalid. While perhaps facing an uphill political battle, the simplest way to accomplish this reform is by amending the FAA.

The ability of autonomous, private individuals and business entities to enter into contractual arrangements is a cornerstone of a stable and efficient market economy. Thus, contract law aims to foster the ability of parties to make arrangements for the future and to assess and allocate the risk of doing business.[34] However, at the point where the exercise of private contractual self-regulation meets with abuses, the autonomy and efficiencies championed by the contractarian theorists of the corporation and the proponents of arbitration must give way to some palpable measure of public regulation.[35] The general policy favoring pre-dispute arbitration agreements has invited corporate abuse in the form of additional, pre-dispute limitations on the procedural rights of the stakeholder constituencies with whom they contract. A per se ban on certain pre-dispute limitations respects the idea of party autonomy to enter into arbitration agreements and, at the same time, allows a better balance of private rights, corporate social accountability, and fundamental procedural fairness.[36]

---

significant advantages over arbitration: it is cheaper than arbitration; it includes a meaningful right to appellate review; it guarantees the appointment of a neutral, independent decision-maker; and it avoids problems with handling certain types of disputes . . . that may not be easily amenable to arbitration."). In light of these recent articles, the position of this Article might seem to be a step backwards. However, the idea of using pre-dispute agreements to modify the Federal Rules of Civil Procedure proceeds upon a premise with which this Article disagrees: Modified procedures are valid in arbitration, so they are (or should be) valid in litigation as well.

34. *See* E. ALLEN FARNSWORTH, CONTRACTS § 1.3, at 8 (4th ed. 2004) (noting function of contract law from parties' perspective is "planning for the future"); ANTHONY T. KRONMAN & RICHARD A. POSNER, *Introduction: Economic Theory and Contract Law*, *in* THE ECONOMICS OF CONTRACT LAW 1, 4 (1979) (asserting that a basic function of contract law is to enforce the "agreed-upon allocation of risk" between parties).

35. *See* FARNSWORTH, *supra* note 34, § 5.1, at 313 (stating parties have freedom to contract until countervailing public interest outweighs contract enforcement).

36. Taylor & Cliffe, *supra* note 16, at 1087 ("[A]ny decision to enforce a [pre-litigation agreement] must balance private contractual autonomy and the attendant efficiencies of [pre-litigation agreements] against the desire to maintain an aura of fairness, which by necessity must be the hallmark of a system of public dispute resolution.").

I. THE RISE OF PRE-DISPUTE ARBITRATION AGREEMENTS AND THE EMERGENCE OF UNCONSCIONABILITY

The history of pre-dispute arbitration is well-documented in the scholarship.[37] At common law, a pre-dispute agreement to arbitrate was not enforceable because of concern that private contractual arrangements would be able to "oust the courts of jurisdiction conferred by law."[38] In 1925, as a response to this common law hostility,[39] Congress passed the FAA,[40] which statutorily recognized the enforcement of these agreements. Since the FAA's enactment, the Supreme Court has consistently declared an "emphatic federal policy in favor of arbitral dispute resolution,"[41] and arbitration is now ubiquitous across various industries and contractual relationships.[42]

The FAA applies rather broadly to any transaction involving interstate commerce and preempts any state law that would contradict a policy favoring arbitration.[43] Section 2, the chief substantive provision of the FAA, provides:

> A written provision in any . . . contract evidencing a transaction involving commerce to settle by arbitration a controversy thereafter arising out of such contract . . . or an agreement in writing to submit to arbitration an existing controversy arising out of such contract . . . shall be valid, irrevocable, and enforceable, save upon such grounds as exist at law or in equity for the revocation of any contract.[44]

Arbitration is, thus, a creature of contract, and courts have interpreted the FAA as putting arbitration agreements on "equal footing" with other contracts.[45]

Some of the perceived benefits of arbitration include its simplicity, informality and expedience relative to civil litigation.[46] In light of these

---

37. *See, e.g.*, Drahozal, *supra* note 28, at 700–05; Schwartz, *supra* note 23, at 81–109; Stempel, *supra* note 25, at 768–92.

38. Taylor & Cliffe, *supra* note 16, at 1092–93 (quoting Home Ins. Co. of N.Y. v. Morse, 87 U.S. 445, 451 (1874)).

39. *See* Schwartz, *supra* note 23, at 75 ("Dissatisfaction with these anti-arbitration doctrines among the business community, bench and bar led to a reform movement at the turn of the century, which in turn led to the adoption of the FAA.").

40. Federal Arbitration Act, 9 U.S.C. §§ 1–16 (2000 & Supp. V 2007).

41. Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., 473 U.S. 614, 631 (1985).

42. *See* discussion *infra* Part II.

43. *See* Allied-Bruce Terminix Cos. v. Dobson, 513 U.S. 265, 269, 272–73 (1995) (holding that Alabama statute invalidating pre-dispute arbitration agreements is preempted by FAA); Southland Corp. v. Keating, 465 U.S. 1, 10, 16 (1984) (holding that California franchise statute requiring claims be brought in court notwithstanding arbitration agreements is preempted by FAA).

44. Federal Arbitration Act, 9 U.S.C. § 2 (2000).

45. *See* Cap Gemini Ernst & Young, U.S., L.L.C. v. Nackel, 346 F.3d 360, 364 (2d Cir. 2003).

46. Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 31 ("[B]y agreeing to arbitrate,

benefits, corporations often place pre-dispute agreements to arbitrate in their standardized form contracts.[47] This has led many critics of pre-dispute arbitration agreements to describe them as "mandatory" or "compelled," because a contracting party is presented with a form agreement on a take-it-or-leave-it basis and likely does not read or consider its terms.[48] Then, once a dispute arises, that party is compelled to arbitrate the dispute even though she would prefer to litigate in court.[49]

Pre-dispute arbitration clauses[50] have been used with increasing frequency in any situation where the parties' relationship involves a contract—whether between two corporations, a corporation and a consumer, or a corporation and an employee.[51] As explained in the following discussion, corporations have also attempted to use form arbitration clauses to exact limits on the other party's procedural rights. Because the FAA defers to extant contract law, the policing of these terms is largely left to the unconscionability doctrine.

A. With the Rise of Arbitration, the Rise of Express Pre-Dispute Procedural Limitations

Arbitration has been described as an implicit waiver of rights.[52] By agreeing to arbitrate, a party agrees to forego the judicial forum and, with that, the formal rules of evidence and procedure, the right to a jury trial, if applicable, and the right to take an appeal from the award.[53] It is debatable,

---

a party 'trades the procedures and an opportunity for review of the courtroom for the simplicity, informality, and expedition of arbitration.'" (quoting *Mitsubishi Motors Corp.*, 473 U.S. at 628)). However, some have questioned whether arbitration actually produces these benefits. *See, e.g.*, Noyes, *supra* note 33, at 584–91 (arguing that arbitration is not necessarily faster or cheaper than litigation). *See generally* Bruce A. Rubin & Jennifer J. Roof, *A Contrarian's Checklist to Arbitration Clauses*, 74 DEF. COUNS. J. 242 (2007) (delineating "myths" of arbitration).

47. *See* examples *infra* Part II.

48. *See, e.g.*, Drahozal, *supra* note 28, at 706–08; Schwartz, *supra* note 23, at 37 n.10.

49. *See, e.g.*, Drahozal, *supra* note 28, at 707.

50. Federal Arbitration Act, 9 U.S.C. § 2 (2000) enforces both pre-dispute agreements to arbitrate as well as "submission" agreements, which are agreements to arbitrate once a dispute has arisen. As submission agreements are relatively uncontroversial, this Article is concerned only with express, procedural limitations contained within pre-dispute agreements, or "ex ante" agreements—that is, agreements to arbitrate that are formed prior to any dispute. Some had argued that the jurisprudence interpreting the FAA has failed to recognize important distinctions between pre-dispute agreements and submission agreements. *See, e.g.*, Schwartz, *supra* note 23, at 104–05; Taylor & Cliffe, *supra* note 16, at 1085 n.2.

51. *See* examples discussed *infra* Part III.

52. *See* sources cited *supra* note 29.

53. *See* UNIF. ARBITRATION ACT §6, 7 U.L.A. 28 cmt. 7 (amended 2000) ("[A]n arbitration agreement effectively waives a party's right to a jury trial . . . ."); Mark E. Budnitz, *Arbitration of Disputes Between Consumers and Financial Institutions: A Serious Threat to Consumer Protection*, 10 OHIO ST. J. ON DISP. RESOL. 267, 283 (1995) ("Obviously, in

however, whether a general agreement to arbitrate equates to an implicit abdication of certain procedural mechanisms that are traits of civil adjudication. By broadly agreeing to arbitrate, does a party implicitly waive the right to bring a class action or class arbitration? Does that party implicitly waive the right to take depositions? Increasingly, companies are reluctant to leave these questions open to an arbitrator's interpretation and have instead expressly added procedural limitations to their arbitration clauses.

After *Green Tree Financial Corp. v. Bazzle*,[54] corporations have strong incentive to add express procedural limitations to their arbitration clauses. In *Bazzle*, a group of homeowners each signed mortgage contracts with the defendant mortgage company that contained a substantially similar, general clause agreeing to arbitrate any disputes arising out of the loan contracts.[55] The homeowners brought a putative class arbitration against the lender for consumer protection violations.[56] Although the arbitration clause was silent concerning the homeowners' right to bring a class arbitration, the lender argued that, by generally agreeing to arbitration, the homeowners implicitly waived the right to seek class relief.[57] The Supreme Court addressed the issue of *who* should decide whether the arbitration clause permitted the homeowners to proceed on a class basis—the arbitrator or the court?[58] The Court held that, when the contract contains a broad clause generally agreeing to arbitrate, the

---

arbitration, the parties waive their rights to factfinding by a jury of their peers . . . . to a trial presided over by a judge who is an elected or appointed public official . . . . [and] to full-blown discovery.") (footnotes omitted); Ryan Griffitts, *Steering Clear of the Runaway Jury*, 68 TEX. B.J. 320, 320 (2005) ("By executing arbitration agreements, the parties waive their right to have their case decided by a judge, and, more important, a jury."); Jean R. Sternlight, *The Rise and Spread of Mandatory Arbitration as a Substitute for the Jury Trial*, 38 U.S.F. L. REV. 17, 19 (2003) ("[C]ritics attack mandatory arbitration on a variety of grounds, including not only its elimination of access to courts and juries, but also its actual or potential lack of neutrality, high cost, diminution of claimants' remedies, elimination of class actions, and curtailment of discovery.") (footnotes omitted); Steven C. Bennett, *Institution Versus Individual: The Arbitration Alternative to Litigation*, METROPOLITAN CORP. COUNS., Aug. 2005, at 9. Mr. Bennet, a trial lawyer practicing in the field of commercial arbitration, describes the practical consequences of arbitration agreements:

> "[A]rbitration agreements and rules rarely require strict adherence to rules of evidence. Thus, hearsay and other forms of suspect evidence are often admitted in arbitration proceedings. Many arbitrators will take all evidence offered "for what it's worth," placing principal emphasis on the weight of the evidence, rather than its admissibility. Only limited rights of appeal exist. Even where an arbitrator may have committed an error of law, courts generally will not upset an arbitration award."

Bennett, *supra*, at 9.

54. 539 U.S. 444 (2003) (plurality opinion).

55. *Id.* at 448–49.

56. *Id.*

57. *Id.* at 449–50.

58. *Id.* at 451–53.

arbitrator (and not the court) should interpret whether the clause permits the arbitration to proceed on a classwide basis.[59]

In the wake of *Bazzle*, businesses have not left these questions to implication by simply using broad agreements to arbitrate.[60] Rather, with increasing frequency, corporations have expressly limited the right to proceed on a classwide basis.[61] Additionally, they have attempted to use pre-dispute arbitration clauses to expressly limit the availability of discovery mechanisms and even to shorten the statute of limitations on potential claims.[62]

B. The Use of Unconscionability to Police Pre-Dispute Arbitration Agreements

Because arbitration is a creature of contract, the first questions a court or arbitrator must ask are whether the parties agreed to arbitrate and whether the dispute is within the scope of that agreement.[63] The question of whether the parties agreed to arbitrate is governed by general principles of contract interpretation, which look to the parties' apparent intentions.[64] Whether the dispute is within the scope of arbitration is decided "by applying the 'federal substantive law of arbitrability, applicable to any arbitration agreement [under the FAA].'"[65] The Supreme Court has interpreted the FAA as "establish[ing] that, as a matter of federal law, any doubts concerning the scope of arbitrable issues should be resolved in favor of arbitration."[66] *Who* decides the threshold issue of arbitrability (the arbitrator or the court) depends upon whether there is evidence that the parties "clearly and unmistakably" agreed to submit the question of arbitrability itself to arbitration.[67] If no such evidence exists, a

---

59. *Id.* at 453.

60. *See* Gilles, *supra* note 19, at 376–78; *infra* Part II.A. Indeed, this was Justice Stevens' prediction at the *Bazzle* oral argument when he asked, "Does this case have any future significance, because isn't it fairly clear that all the arbitration agreements in the future will prohibit class actions?" Transcript of Oral Argument at 55, *Bazzle*, 539 U.S. 444 (No. 02-634).

61. *See* Gilles, *supra* note 19, at 376–78; *infra* Part II.

62. *See infra* Part II (providing examples of express, pre-dispute limitations); *see also* Hans Smit, *Class Actions and Their Waiver in Arbitration*, 15 AM. REV. INT'L ARB. 199, 200 (2004) ("[L]awyers [have] . . . turned their attention to the arbitration agreement and to the great freedom in the making of contracts traditionally fostered by the common law.").

63. *See* Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., 473 U.S. 614, 626 (1985).

64. 7 SAMUEL WILLISTON & RICHARD A. LORD, A TREATISE ON THE LAW OF CONTRACTS § 15:11 (4th ed. 1997).

65. *Mitsubishi Motors Corp.*, 473 U.S. at 626 (quoting Moses H. Cone Mem'l Hosp. v. Mercury Constr. Corp., 460 U.S. 1, 24 (1983)).

66. *Id.* (citing *Moses H. Cone Mem'l Hosp.*, 460 U.S. at 24–25).

67. First Options of Chi., Inc. v. Kaplan, 514 U.S. 938, 944 (1995) (alteration in original) (citations omitted). For a brief but thorough discussion of who determines threshold issues of arbitrability, see June Lerhman, *On the Threshold of Arbitration*, L.A. LAW., December 2003, at 20.

presumption arises that the court (not the arbitrator) should decide arbitrability.[68]

Assuming there is an arbitrable dispute, there still exists a question of clause construction: Did the parties provide any detail in their agreement concerning how the dispute would proceed? For example, after *Bazzle*, if the parties' contract contains only a general arbitration clause while remaining silent as to class relief, the arbitrator (and not the court) decides whether the clause permits arbitration to proceed on a classwide basis.[69]

As discussed, however, rather than leaving these limitations to implication, companies are expressly stating such procedural limitations in their arbitration clauses.[70] What happens, then, when an arbitration clause contains a procedural limitation, for example, expressly waiving the right to class arbitration or to take depositions? These additional procedural limitations within arbitration clauses are presumptively enforceable.[71] They are only unenforceable "upon such grounds as exist at law or inequity for the revocation of any contract."[72] In other words, the FAA determines the validity of arbitration provisions and the terms contained therein under extant, state contract law.

Therefore, an arbitration clause that is not tainted with fraud or duress is enforceable unless it is unconscionable or against public policy.[73] Not surprisingly then, contracting parties claim unconscionability as a defense with increasing frequency in the context of arbitration.[74] Litigants have aimed challenges either at the arbitration clause in its entirety,[75] or more specifically at the express, procedural limitations contained within the clause.[76]

---

68. *First Options of Chi., Inc.*, 514 U.S. at 944–45.

69. *See* Fin. Corp. v. Bazzle, 539 U.S. 444, 452–53 (2003) (plurality opinion); *see also* AM. ARBITRATION ASSOC., SUPPLEMENTARY RULES FOR CLASS ARBITRATIONS R. 3 (2003), *available at* http://www.adr.org/sp.asp?id=21936&printable=true ("[T]he arbitrator shall determine as a threshold matter . . . whether the applicable arbitration clause permits the arbitration to proceed on behalf of or against a class . . . .").

70. *See supra* note 60 and accompanying text.

71. *See generally, e.g.*, Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 31 (1991) (discovery limitations); Discover Bank v. Superior Court, 113 P.3d 1100, 1110 (2005) (class arbitration waivers).

72. *See* Federal Arbitration Act, 9 U.S.C. § 2 (2000)

73. Doctor's Assocs., Inc. v. Casarotto, 517 U.S. 681, 687 (1996).

74. *See* Susan Randall, *Judicial Attitudes Toward Arbitration and the Resurgence of Unconscionability*, 52 BUFF. L. REV. 185, 194 (2004) ("[A]s the use of arbitration agreements has increased, claims of unconscionability have also increased . . . ."); *see also* Stempel, *supra* note 25, at 761–62 ("Beginning in the 1990s, . . . courts[] began to take a harder look at arbitration agreements and their enforcement. Several courts began invoking concepts related to unconscionability in order to refuse enforcement of arbitration clauses. The phenomenon accelerated in the late 1990s.").

75. *See* Davis v. O'Melveny & Myers, 485 F.3d 1066, 1084 (9th Cir. 2007) (finding where the offending provisions are not merely ancillary to the agreement, but rather go to its heart, the agreement as a whole is unenforceable); Alexander v. Anthony Int'l, L.P., 341 F.3d

Such challenges raise another question: Should the court or arbitrator decide the merits of an unconscionability defense? Under the doctrine of separability, the answer depends upon whether the challenge is addressed to the contract in general or, more specifically, to the arbitration provision. Under *Prima Paint Corp. v. Flood & Conklin Manufacturing Co.*,[77] and more recently, according to *Buckeye Check Cashing, Inc. v. Cardegna*,[78] when the challenge is directed to the validity of the entire contract, an arbitrator decides the contract's enforceability. [79] However, if the challenge is directed specifically to the arbitration clause, but not the contract as a whole, the court determines the validity of the arbitration clause.[80]

Another threshold inquiry concerning the validity of the terms of an arbitration clauses is which state's unconscionability law applies.[81] This choice of law analysis is critical because the elements necessary to prove unconscionability tend to vary from state to state.[82] Some states require a showing of both procedural *and* substantive unconscionability, [83] while others

---

256, 271 (3d Cir. 2003) ("The cumulative effect of so much illegality prevents us from enforcing the arbitration agreement. Because the sickness has infected the trunk, we must cut down the entire tree.").

76. *See* Gannon v. Circuit City Stores, Inc., 262 F.3d 677, 681 (8th Cir. 2001) (finding inclusion of unconscionable damages-limitation clause does not require the invalidation of the arbitration agreement as a whole); Muhammad v. County Bank of Rehoboth Beach, 912 A.2d 88, 103 (N.J. 2006) (enforcing the remaining provisions of an arbitration agreement after severing an unconscionable class action waiver provision).

77. 388 U.S. 395 (1967)

78. 546 U.S. 440 (2006).

79. *See Buckeye Check Cashing, Inc.*, 546 U.S. at 445–46; *Prima Paint Corp.*, 388 U.S. at 404.

80. *See Buckeye Check Cashing, Inc.*, 546 U.S. at 445–46; *Prima Paint Corp.*, 388 U.S. at 403–04.

81. *See, e.g.*, Coady v. Cross County Bank, 729 N.W.2d 732, 737 (Wis. Ct. App. 2007) (finding unconscionability of arbitration provision and, thus enforceability, presents a preliminary inquiry into which state's unconscionability law applies).

82. *See* Bennett, *supra* note 53, at 9 ("The precise law of unconscionability varies from state to state. The mere fact that the individual must 'take it or leave it' with regard to a contract does not automatically invalidate the contract. The individual typically has at least the 'leave it' choice in responding to the proffered contract.").

83. *See* Fotomat Corp. v. Chanda, 464 So.2d 626, 629–30 (Fla. 1985) ("Most courts take a 'balancing approach' to the unconscionability question, and to tip the scales in favor of unconscionability, most courts seem to require a certain quantum of procedural plus a certain quantum of substantive unconscionability.") (citations omitted); Martin v. Sheffer, 403 S.E.2d 555, 557 (N.C. Ct. App. 1991) (requiring both procedural and substantive unconscionability); Constr. Assocs., Inc. v. Fargo Water Equip. Co., 446 N.W.2d 237, 241–42 (N.D. 1989) (describing a "two-pronged framework" of procedural and substantive unconscionability); 8 SAMUEL WILLISTON & RICHARD A. LORD, A TREATISE ON THE LAW OF CONTRACTS § 18:10, at 62 (4th ed. 1998) ("It has often been suggested that a finding of a procedural abuse, inherent in the formation process, must be coupled as well with a substantive abuse, such as an unfair or unreasonably harsh contractual term which benefits the drafting party at the other party's

require a strong showing of only one *or* the other to invalidate a contract.[84] In this regard, a choice of law clause can be extremely significant in determining the burden of proof to invalidate the arbitration clause or certain limitations contained therein—the choice of law may affect whether the party challenging the clause must show both procedural and substantive unconscionability.[85] Nevertheless, a court may decline to enforce the parties' choice of law clause if applying another state's substantive law would violate the public policy of the state where the court is located.[86] A court in State X could, for example, refuse to apply State Z's unconscionability law on the theory that such application would allow a collective action waiver to be upheld in a situation that would undermine State X's public policy favoring class relief.[87]

The importance of the choice of law determination is more evident when considering the differences between procedural and substantive unconscionability. A showing of procedural unconscionability considers the

---

expense.") (footnote omitted).

84. *See* Maxwell v. Fidelity Fin. Servs., Inc., 907 P.2d 51, 59 (Ariz. 1995) (allowing claim of unconscionability based on substantive unconscionability alone); Gillman v. Chase Manhattan Bank, N.A., 534 N.E.2d 824, 829 (N.Y. 1988) ("While determinations of unconscionability are ordinarily based on the court's conclusion that both the procedural and substantive components are present, there have been exceptional cases where a provision of the contract is so outrageous as to warrant holding it unenforceable on the ground of substantive unconscionability alone.") (citations omitted); 8 WILLISTON & LORD, *supra* note 83, § 18:10, at 64–66. Professor Lord shrewdly questions the coupling of procedural and substantive unconscionability:

> The distinction between procedural and substantive abuses, however, may become quite blurred; overwhelming bargaining strength or use of fine print or incomprehensible legalese may reflect procedural unfairness in that it takes advantage of or surprises the victim of the clause, yet the terms contained in the resulting contract—whether in fine print or legal "gobbledygook"—would hardly be of concern unless they were substantively harmful to the nondrafting party as well. Thus, the regularity of the bargaining procedure may be of less importance if it results in harsh or unreasonable substantive terms, or substantive unconscionability may be sufficient in itself even though procedural unconscionability is not.

8 WILLISTON & LORD, *supra* note 83, § 18:10, at 64-66 (footnotes omitted).

85. The party challenging the arbitration clause (usually, the plaintiff) has the burden of showing that a provision is unconscionable and, thus, unenforceable. Scott v. Cingular Wireless, 161 P.3d 1000, 1005 (Wash. 2007).

86. *See, e.g.*, Discover Bank v. Superior Court, 113 P.3d 1100, 1117 (Cal. 2005); Am. Online, Inc. v. Superior Court, 108 Cal. Rptr. 2d 699, 708 (Ct. App. 2001).

87. *See, e.g.*, *Discover Bank*, 113 P.3d at 1117; *Am. Online*, 108 Cal. Rptr. at 708. Moreover, the law remains unsettled concerning the effect of a choice of law clause on the rules applying to arbitration, adding yet another layer of complexity. For example, it is unclear whether a general choice of law clause should invoke state arbitration law and thereby "opt out" of the FAA. *See* Jennifer Trieshmann, *Horizontal Uniformity and Vertical Chaos: State Choice of Law Clauses and Preemption under the Federal Arbitration Act*, 2005 J. DISP. RESOL. 161, 169 (2005) (describing the majority and minority views by courts).

manner in which the parties entered into the contract.[88] This analysis looks at "whether the imposed-upon party had meaningful choice about whether and how to enter into the transaction."[89] Thus, procedural unconscionability can exist where the agreement is a contract of adhesion—that is, a contract in which the party with superior bargaining power presents the weaker party with a non-negotiable ("take it or leave it") contract on a pre-printed, standardized form.[90] Typical examples of adhesion contracts include cell phone service contracts and credit card agreements because there is no deal if the consumer does not agree to the terms of the standardized form. Despite their one-sidedness, contracts of adhesion are not per se unconscionable.[91]

Substantive unconscionability focuses on the terms of the agreement and whether they favor the party with the superior bargaining position.[92] When assessing whether an arbitration clause is substantively unconscionable, a good number of courts consider the "mutuality" of the provision—i.e., whether the burdens of the terms either expressly or effectively weigh on both parties equally.[93] If there is a lack of mutuality, these courts have held that the terms are unconscionable.[94]

Assuming there is a successful showing of unconscionability, there does not appear to be a clear trend in case law concerning the available remedy. A

---

88. *See* 8 WILLISTON & LORD, *supra* note 83, § 18.10, at 57–64.

89. *See id.* at 57.

90. For a discussion defining the term "contract of adhesion," see Todd D. Rakoff, *Contracts of Adhesion: An Essay in Reconstruction*, 96 HARV. L. REV. 1173, 1176–80 (1983).

91. 8 WILLISTON & LORD, *supra* note 83, § 18.10, at 62–64.

92. *Id.* at 57, 65–66.

93. *See* Christopher R. Drahozal, *Nonmutual Agreements to Arbitrate*, 27 J. CORP. L. 537, 547–52 (2002) (discussing jurisdictions that use a mutuality analysis).

94. For example, in *Lytle v. CitiFinancial Services, Inc.*, 810 A.2d 643 (Pa. Super. Ct. 2002), the Pennsylvania Superior Court determined the enforceability of an arbitration provision that required the mortgagors to arbitrate all issues involving more than $15,000, while the mortgagee retained the right to enforce its repayment rights or commence foreclosure in the courts. *Id.* at 660. The court held that, "under Pennsylvania law, the reservation by CitiFinancial of access to the courts for itself to the exclusion of the consumer creates a presumption of unconscionability." *Id.* at 665. However, the Superior Court recognized that this holding conflicted with *Harris v. Green Tree Financial Corp.*, 183 F.3d 173 (3d Cir. 1999). *Id.* at 665 n.13. In *Harris*, the Third Circuit assessed a similar arbitration provision and found that "the mere fact that [the mortagee] retains the option to litigate some issues in court, while the [mortgagors] must arbitrate all claims does not make the arbitration agreement unenforceable." *Harris*, 183 F.3d at 177–78, 183. In light of conflicting decisions applying Pennsylvania law, the Third Circuit certified to the Pennsylvania Supreme Court the question of "whether an arbitration agreement, consummated in connection with a residential mortgage loan, which reserves judicial remedies related to foreclosure is presumptively unconscionable." Salley v. Option One Mortgage Corp., 925 A.2d 115, 116 (Pa. 2007). The Pennsylvania Supreme Court answered negatively, holding that "the exception from mandatory arbitration for foreclosure contained within the [lender's] arbitration agreement, in and of itself, does not render the agreement presumptively unconscionable under Pennsylvania law." *Id.* at 129.

court may prohibit the entire arbitration process or, instead, sever the unconscionable procedural limitations from the arbitration clause and otherwise allow the dispute to proceed in arbitration.[95] This decision may depend, in part, upon the court's determination of whether the unenforceable procedural limitations are severable. Severability may, in turn, depend upon whether the parties' agreement contains a valid severability clause or if public policy supports severing the offending clause from the agreement. Thus, for example, a court could hold a collective action waiver unenforceable, sever it, and then allow the dispute to proceed in class arbitration.[96] Alternatively, the same court could hold that, because the collective action waiver is unenforceable, the arbitration clause is likewise unenforceable, and the dispute should therefore proceed as a class action in court.[97] As for other procedural limitations, it is certainly much simpler for a court to sever a discovery waiver or a shortened statute of limitations provision from an agreement, and then allow arbitration to otherwise proceed.[98] However, when such clauses are present, courts have in some instances refused entirely to compel arbitration.[99]

II. EXAMPLES OF PRE-DISPUTE PROCEDURAL LIMITATIONS AND THE APPLICATION OF THE UNCONSCIONABILITY DOCTRINE TO POLICE THEM

While arbitration shifts the proceeding to a private forum, the Supreme Court has time and again stated that "[b]y agreeing to arbitrate a statutory claim, a party does not forgo the substantive rights afforded by the statute; it only submits to their resolution in an arbitral, rather than a judicial, forum."[100] The Supreme Court has further stated that "so long as the prospective litigant effectively may vindicate [his or her] statutory cause of action in the arbitral

95. *See, e.g.*, Booker v. Robert Half Int'l, Inc., 315 F. Supp. 2d 94, 109 (D.C. 2004) (severing limitation on punitive damages from arbitration clause and otherwise allowing arbitration to proceed).

96. *See* Skirchak v. Dynamics Research Corp., 508 F.3d 49, 63 (1st Cir. 2007) (striking collective action waiver from employment agreement and otherwise compelling arbitration); Kinkel v. Cingular Wireless LLC, 857 N.E.2d 250, 278 (Ill. 2006) (finding class action waiver within an arbitration agreement both unconscionable and severable); Muhammad v. County Bank of Rehoboth Beach, 912 A.2d 88, 103 (N.J. 2006).

97. *See, e.g.*, Riensche v. Cingular Wireless LLC, No. C06-1325Z, 2006 WL 3827477, at *13 (W.D. Wash. Dec. 27, 2006) (refusing to enforce entire arbitration clause because class action waiver was found to be unconscionable and because parties' agreement provided that, if class action waiver is found to be unenforceable, the entire arbitration clause is null and void); Martinez v. Master Prot. Corp., 12 Cal. Rptr. 3d 663, 675 (Ct. App. 2004) (vacating, inter alia, order compelling arbitration and restoring case to litigation calendar); Scott v. Cingular Wireless, 161 P.3d 1000, 1009 (Wash. 2007) (finding entire arbitration agreement unenforceable due to the unconscionability and contractual inseverability of class action waiver provision).

98. *See, e.g.*, *Booker*, 315 F. Supp. 2d at 109.

99. *See, e.g.*, *Martinez*, 12 Cal. Rptr. 3d at 675.

100. Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., 473 U.S. 614, 628 (1985).

forum, the statute will continue to serve both its remedial and deterrent function."[101] An evident premise of the Supreme Court's jurisprudence is that moving the litigation to an arbitral forum is not intended to equate to an exculpation of the defendant from substantive liability. Nevertheless, after reviewing the cases that address express pre-dispute procedural limitations, a theme emerges: the potential to effect an exculpation of a corporation from substantive liability.

This Part considers three types of pre-dispute procedural limits: collective action waivers, limitations on discovery, and shortened statutes of limitation. For example, a clause that limits discovery might provide that the parties have the right to depose experts only. Alternatively, a provision requiring that an employee file a notice of arbitration within thirty days might effectively truncate a statute of limitations. Notably, these pre-dispute limitations might not be mentioned in the agreement but can nevertheless be incorporated by express reference to institutional rules that contain such limitations. This Part will address these examples in turn.[102]

A. "Collective Action Waivers"

The collective action waiver first appeared in the late 1990s, when trade magazines advised franchisors and other business entities to add express limitations on class actions to their form agreements.[103] Afterwards,

---

101. *Id.* at 637.

102. While this Part discusses each of these express limitations separately, it is important to note that they are not often used in isolation, as corporations may use them simultaneously. For example, an employment contract might contain both limitations on discovery and a shortened statute of limitations. *See, e.g.*, Martinez v. Master Prot. Corp., 12 Cal. Rptr. 3d 663, 668–75 (Ct. App. 2004). When used together, a court might find the arbitration clause substantively unconscionable based on the totality of the circumstances—the restrictions, taken together, evince the one-sided nature of the arbitration clause. *Id.* at 673. Moreover, in addition to the limitations highlighted in this Article, a review of the case law reveals that numerous other restrictions have arisen in arbitration clauses. Common restrictions that are factored into the substantive unconscionability analysis include: prohibitively high arbitration costs to be paid by one party, limitations on remedies, curtailed judicial review, and excessive confidentiality. *See, e.g.*, Davis v. O'Melveny & Myers, 485 F.3d 1066, 1078 (9th Cir. 2007) (excessive confidentiality); Spinetti v. Serv. Corp. Int'l, 324 F.3d 212, 216–17 (3d Cir. 2003) (prohibitively high arbitration costs for one party); Morrison v. Circuit City Stores, Inc., 317 F.3d 646, 670–71 (6th Cir. 2003) (limitations on remedies); Hooters of Am., Inc. v. Phillips, 39 F. Supp. 2d 582, 614 (D.S.C. 1998) (curtailed judicial review). By adding numerous restrictions, the arbitration clause is more likely to be found substantively unconscionable based on its one-sidedness.

103. *See* Gilles, *supra* note 19, at 408–13. The express collective action waiver grew in popularity after the Supreme Court's *Bazzle* decision in 2003. *Id.* at 410 (citing Green Tree Fin. Corp. v. Bazzle, 539 U.S. 444 (2003) (plurality opinion)). The Court remanded the case to an arbitrator to determine whether an arbitration clause, under which the parties agreed to submit to arbitration "'*[a]ll* disputes, claims, or controversies arising from or relating to this contract or

corporations across many industries have attempted to use arbitration clauses to have contracting parties expressly limit the right to bring a class action or class arbitration.[104] In fact, in any place where a business has a contract, it can insert a collective action waiver into its arbitration clause.

For example, the franchise agreement of Quizno's Corporation, a sandwich chain franchisor, contains the following clause under its "Limitation of Claims" section: "Franchisor and Franchisee agree that *any proceeding will be conducted on an individual, not a class-wide, basis*, and that a proceeding between Franchisor and Franchisee . . . may not be consolidated with another proceeding between Franchisor and any other person or entity."[105] Similarly, the standard form employment agreements of U-Haul Co., the moving van company, contain the following limitations within the arbitration clause:

> This mutual obligation to arbitrate means that both you and [defendant] are bound to use the [U-Haul Arbitration Policy] as the only means of resolving any employment related disputes. This mutual obligation to arbitrate claims also means that both you and [defendant] forego any right either may have to a jury trial on claims relating in any way to your employment, and *both you and [defendant] forego and waive any right to join or consolidate claims in arbitration with others or to make claims in arbitration as a representative or as a member of a class or in a private attorney general capacity, unless such procedures are agreed to by both you and [defendant].*[106]

In addition to franchise and employment contracts, other examples of collective action waivers abound in consumer[107] and loan[108] contracts. They have also

---

the relationships which result from this contract[,]'" forbid class arbitration. *Id.* at 448, 451–54 (citation omitted) (alteration in original).

104. *See* Gilles, *supra* note 19, at 408–13

105. Quizno's Franchise Agreement, § 21.4, *available at* http://www.secinfo.com/dsV1x.71a.a.htm (emphasis added). Recent news stories have reported franchisee class action suits against the sandwich chain. *See* Julie Creswell, *Some Quiznos Franchisees Take Chain to Court*, N.Y. TIMES, Feb. 24, 2007, at C1. It appears the corporation has not yet raised the collective action waiver as a defense to the class action suit. For other examples of collective actions waivers in franchise agreements, see Marron v. Snap-On Tools, Co., No. Civ. 03-4563, 2006 WL 51193, at *1 (D.N.J. Jan. 9, 2006); Blimpie Int'l, Inc. v. Blimpie of the Keys, 371 F. Supp. 2d 469, 471 (S.D.N.Y. 2005); and Indep. Assoc. of Mailbox Ctr. Owners, Inc. v. Superior Court, 34 Cal. Rptr. 3d 659, 662 (Ct. App. 2005).

106. Konig v. U-Haul Co. of Cal., 52 Cal. Rptr. 3d 244, 247 (Ct. App. 2006) (alteration in original) (emphasis added), *rev. granted* 55 Cal. Rptr. 3d 864 (2007). For other examples of collective action waivers in employment agreements, see Ingle v. Circuit City Stores, Inc., 328 F.3d 1165, 1176 (9th Cir. 2003); and Skirchak v. Dynamics Research Corp., 432 F. Supp. 2d 175, 178–80 (D. Mass. 2006).

107. *See, e.g.*, Dale v. Comcast Corp., 453 F. Supp. 2d 1367, 1374 (N.D. Ga. 2006) (cable television service subscriber agreement); Stern v. Cingular Wireless Corp., 453 F. Supp. 2d 1138, 1141–42 (C.D. Cal. 2006) (cellular telephone service contract); Cunningham v. Citigroup, No. Civ. 05-3476, 2005 WL 3454312, at *1 (D.N.J. Dec. 16, 2005) (credit card agreement); Stenzel v. Dell, Inc., 870 A.2d 133, 137–38 (Me. 2005) (optional contract for

been used in insurance contracts,[109] and one has even arisen in a standardized enrollment contract between an educational institution and its students.[110] These collective action waivers have generally been enforced, with courts holding they are "not per se invalid."[111] With mixed and sometimes irreconcilable results, courts have mostly left the enforceability of collective action waivers to a case-by-case unconscionability inquiry.[112]

With this inquiry left to extant contract doctrine, companies have had some success in contracting around class actions and class arbitrations. Yet, this fiat by contract has not been the focus of the public or scholarly debate. Rather, the focus of the discussion of the "decline and fall" of class actions has been legislative reform.[113] Indeed, Professor Myriam Gilles has recently suggested that the focus on legislative reform may be somewhat misplaced.[114] The significant threat to collective procedural mechanisms is not legislative but, rather, private standardized contracts.[115] In what Professor Jean Sternlight dubs a "'do-it-yourself' approach to law reform," businesses are contracting out of the existing class action, class arbitration, and joinder procedures by inserting collective action waivers into their arbitration clauses.[116]

---

computer service).

108. *See, e.g.*, Sprague v. Household Int'l, 473 F. Supp. 2d 966, 969–70 (W.D. Mo. 2005) (consumer loan agreement); Walther v. Sovereign Bank, 872 A.2d 735, 739 (Md. 2005) (mortgage agreement); Vasquez-Lopez v. Beneficial Or., Inc., 152 P.3d 940, 949 (Or. Ct. App. 2007) (mortgage agreement).

109. *See, e.g.*, Lomax v. Woodmen of the World Life Ins. Soc'y, 228 F. Supp. 2d 1360, 1365 (N.D. Ga. 2002) (life insurance agreement); Peach v. CIM Ins. Corp., 816 N.E.2d 668, 670 (Ill. App. Ct. 2004) (automobile "extended protection plan").

110. Davis v. ECPI Coll. of Tech., L.C., No. 05-2122, 2007 WL 840506, at *1 (4th Cir. Mar. 20, 2007).

111. *See supra* note 71 and accompanying text.

112. *See, e.g.*, Skirchak v. Dynamics Research Corp., 508 F.3d 49, 57 (1st Cir. 2007); Discover Bank v. Superior Court, 113 P.3d 1100, 1108 (Cal. 2005).

113. *See* Edward F. Sherman, *Decline & Fall*, 93 A.B.A. J. 51, 51 (2007). Since the class action's introduction into the Federal Rules of Civil Procedure, *see* FED. R. CIV. P. 23, it has met with staunch supporters and detractors. *See* Sherman, *supra*, at 51. Professor Sherman describes the heated legal climate in which "business organizations pursued an intensive campaign to sway public opinion, and to lobby Congress and state legislatures for change in substantive and procedural law that would put the clamps on consumer class actions." *Id.* As Profesor Sherman observes, "[d]epending upon one's point of view, the class action is a powerful vehicle for protecting the rights of individuals confronting powerful corporations—or a legal version of Frankenstein's monster." *Id.* Most recently, the class action has been restricted through legislation like the Class Action Fairness Act of 2005, 28 U.S.C. §§ 1332(d), 1453, 1711–15 (2000 & Supp. V 2007), and the Private Securities Litigation Reform Act of 1995, 15 U.S.C. § 78u-4(b) (2001). Much of the scholarship concerning restricting class actions has focused on these legislative reforms.

114. *See* Gilles, *supra* note 19, at 375.

115. *Id.*

116. Jean R. Sternlight, *As Mandatory Binding Arbitration Meets the Class Action, Will the Class Action Survive?*, 42 WM. & MARY L. REV. 1, 11 (2000); *see also* Gilles, *supra* note 19, at

This "opting out" of liability[117] is particularly problematic because it defeats the prevailing benefit of collective action: to aggregate claims that would, on an individual basis, otherwise have a negative or minimal value.[118] Likewise, these waivers may also defeat the primary rationale for the class mechanism—judicial (or arbitral) economy—by permitting numerous claims that share common issues of law and fact to be adjudicated separately, thereby resulting in the duplication of efforts and the waste of resources that class mechanisms aim to minimize.[119]

Certainly, the concerns with collective action waivers devolve into a battle of competing efficiency arguments, with the underlying justifications for collective action on the one hand and the corporations' arguments justifying the use of collective action waivers on the other. The aggregation of claims, especially when all contracting parties have signed the same standardized contract, streamlines the process of enforcing substantive laws and avoids the duplication of efforts and resources of litigants, courts, and arbitrators. However, corporations that use collective action waivers point to economic efficiency; they argue that the collective action waiver allows them to provide their services at a cheaper rate because they will not have to defend against costly class actions or arbitrations.[120] There is simply no empirical evidence, however, to suggest that corporations actually do pass these hypothetical savings on to consumers.[121]

---

430 (discussing how companies are using their arbitration clauses to expressly "opt out of all potential classwide liability").

117. *See* Gilles, *supra* note 19, at 430.

118. Phillips Petroleum Co. v. Shutts, 472 U.S. 797, 809 (1985). Justice Rehnquist observed that "[m]odern plaintiff class actions follow the same goals, permitting litigation of a suit involving common questions when there are too many plaintiffs for proper joinder. Class actions also may permit the plaintiffs to pool claims which would be uneconomical to litigate individually." *Id.* That case, for instance, "involve[d] claims averaging about $100 per plaintiff [and therefore] most of the plaintiffs would have no realistic day in court if a class action were not available." *Id.*

119. Devlin v. Scardelletti, 536 U.S. 1, 10 (2002) (noting that one of the major goals of class action litigation is "to simplify litigation involving a large number of class members with similar claims").

120. Jean R. Sternlight & Elizabeth J. Jensen, *Using Arbitration to Eliminate Consumer Class Actions: Efficient Business Practice or Unconscionable Abuse?*, 67 LAW & CONTEMP. PROBS. 75, 77–92 (2004) (examining this consumer savings argument for allowing collective action waivers).

121. Sternlight & Jensen, *supra* note 120, at 95 ("In sum, economic theory alone raises significant doubts that companies pass on to consumers the entire cost-savings from using arbitration clauses to eliminate class actions. It is not surprising that, to date, no published studies show that the imposition of mandatory arbitration leads to lower prices."); *see also* Carrington, *supra* note 21, at 76 ("[A]n arbitration clause may be merely a disguised provision requiring [that] the weaker party . . . bear additional costs, such as those associated with contesting a matter in a distant forum or paying the salary of the neutral . . . [or] risk paying the stronger party's legal fees if the claim fails . . . .").

As an example, consider the facts of the widely-cited California case of *Discover Bank v. Superior Court.*[122] Christopher Boehr, a credit card holder, was charged a $29 late fee by Discover Bank despite making his payment on the due date.[123] Boehr alleged that the late fee resulted from the bank's undisclosed policy of crediting payments made after 1 p.m. to the next business day.[124] Boehr sought to pursue class arbitration, alleging that class members suffered losses as a result of Discover Bank's deception in representing the time at which fees would be assessed. [125] However, Discover Bank's cardholder agreement contained a clause forbidding classwide arbitration.[126]

Discover Bank insisted that, under the cardholder agreement, Boehr could only proceed on an individual basis.[127] Because Boehr had only a negligible $29 claim, his individual lawsuit was not worth enough money to prosecute and the harm would go without redress. Nonetheless, Boehr challenged the clause as unconscionable.[128] In a decision that appeared to come just a little shy of declaring collective action waivers in consumer contracts against public policy, the California Supreme Court held that the collective action waiver was unconscionable and, thus, unenforceable.[129]

California requires a showing of both procedural and substantive unconscionability.[130] In *Discover Bank*, the court held that the collective action waiver was procedurally unconscionable because Boehr's cardholder agreement did not initially contain an arbitration clause.[131] Rather, the arbitration clause was purportedly made a part of the cardholder agreement by a "notice of amendment" in a "bill stuffer."[132] Further, the court held that the collective action waiver was substantively unconscionable because it, in effect, operated to insulate Discover Bank from liability.[133]

Given the relatively small sum of Boehr's claim, the collective action waiver served as a disincentive for consumers to hold Discover Bank

---

122. 113 P.3d 1100 (Cal. 2005).

123. *Id.* at 1104.

124. *Id.* at 1103.

125. *Id.*

126. The clause provided: "NEITHER YOU NOR WE SHALL BE ENTITLED TO JOIN OR CONSOLIDATE CLAIMS IN ARBITRATION BY OR AGAINST OTHER CARDMEMBERS WITH RESPECT TO OTHER ACCOUNTS, OR ARBITRATE ANY CLAIM AS A REPRESENTATIVE OR MEMBER OF A CLASS OR IN A PRIVATE ATTORNEY GENERAL CAPACITY." *Id.*

127. *Id.*

128. *Id.* at 1104.

129. *Id.* at 1110. "[W]hen the waiver is found in a consumer contract of adhesion . . . involv[ing] small amounts of damages, and when . . . the party with the superior bargaining power has carried out a scheme to deliberately cheat large numbers of consumers out of individually small sums of money . . . such waivers are unconscionable . . . ." *Id.*

130. *Id.* at 1108.

131. *Id.*

132. *Id.*

133. *Id.*

accountable.[134] Correspondingly, Discover Bank insulated itself from the threat of any significant liability for the overcharge amounts that, on an individual basis, were comparatively meager but, in the aggregate, could have reaped the company a handsome return.[135] Indeed, in holding that this particular collective action waiver was substantively unconscionable, the California Supreme Court analogized it to an exculpatory clause, remarking that Discover Bank had sought to use the clause as a "'get out of jail free' card."[136] The court noted that, while collective action is a procedural mechanism, it is "often inextricably linked to the vindication of substantive rights."[137] A case involving a similar waiver in a cellular telephone contract led another court to quote the French novelist Anatole France: "The law in its majesty prohibits rich and poor alike from sleeping under bridges."[138]

But not all courts have likened collective action waivers to exculpatory clauses.[139] In *Jenkins v. First American Cash Advance of Georgia, LLC*, for example, the Eleventh Circuit enforced a collective action waiver.[140] Plaintiff Charlene Jenkins attempted to file a class action lawsuit against a payday lender and its affiliate, alleging that the agreements under which she received loans violated Georgia usury laws.[141] Jenkins and the defendants entered into at least eight payday lending transactions, with each of the loans each valued at less than $500 with maturity dates between seven and fourteen days.[142] The annual

---

134. Notably, the clause also expressly aimed to quell arbitration in any "private attorney general capacity." *Id.* at 1103.

135. The number of customers charged with this fee is unclear. Assuming Discover Bank charged 10,000 cardholders the $29 fee, Discover Bank would have billed $290,000.00 in fees, before interest.

136. *Discover Bank*, 113 P.3d at 1108 (citations omitted).

137. *Id.* at 1109.

138. Thibodeau v. Comcast Afroilan, No. 4526, 2006 WL 416863, at *1 (Pa. Ct. C.P. Jan. 27, 2006), *aff'd* 912 A.2d 874 (Pa. Super. Ct. 2006).

139. Many courts have upheld "collective action waivers." *See* Jenkins v. First Am. Cash Advance of Ga., LLC, 400 F.3d 868, 878 (11th Cir. 2005); Iberia Credit Bureau, Inc. v. Cingular Wireless LLC, 379 F.3d 159, 174–75 (5th Cir. 2004); Blaz v. Belfer, 368 F.3d 501, 505 (5th Cir. 2004); Snowden v. Checkpoint Check Cashing, 290 F.3d 631, 638–39 (4th Cir. 2002); Champ v. Siegel Trading Co., 55 F.3d 269, 274–77 (7th Cir. 1995); Dale v. Comcast Corp., 453 F. Supp. 2d 1367, 1376–77 (N.D. Ga. 2006); Tillman v. Commercial Credit Loans, Inc., 629 S.E.2d 865, 872–75 (N.C. Ct. App. 2006); Strand v. U.S. Bank Nat'l Ass'n ND, 693 N.W.2d 918, 921–27 (N.D. 2005); AutoNation USA Corp. v. Leroy, 105 S.W.3d 190, 199–200 (Tex. Ct. App. 2003).

140. *Jenkins*, 400 F.3d 868 at 882–83.

141. *Id.* at 870.

142. *Id.* at 871. The court described the typical payday loan transaction:

> In such transactions, a borrower receives a modest cash advance that becomes due for repayment within a short period of time, usually about 14 days. As security for the loan, the borrower gives a check to the payday lender in the amount of the cash advance, plus the interest charged by the lender. The interest rates in payday lending transactions typically range from 20% to 30% for a two-week advance, which computes to an annual

percentage rates charged by the defendants for the loans ranged from 438% to 938.57%.[143] To consummate each loan transaction, Jenkins executed a promissory note and an arbitration agreement.[144] Among other things, the arbitration agreements provided that Jenkins was waiving her right to participate in a class action or class arbitration against the lender.[145]

Pointing to these clauses, the lender moved to stay the proceeding and compel arbitration on an individual basis.[146] The district court held that the arbitration clause was procedurally unconscionable, noting that a payday loan in a pre-printed adhesion contract "unquestionably places the consumer at a severe bargaining disadvantage."[147] Moreover, given the high interest rates, the loans "would only appeal to extremely desperate consumers."[148] The district court likewise held that the arbitration clause was substantively unconscionable.[149] First, the court recognized a lack of mutuality; to the extant that the arbitration clause limited Jenkins' access to small claims court, it was one-sided in favor of the lender.[150] Second, the court noted that the collective action waiver was unfair because "[e]ach arbitration clause was attached to a small loan of under $500," and thus, "[a] class action is the only way that borrowers with claims as small as the individual loan transactions can obtain relief."[151]

The district court recognized that denying Jenkins access to the class action mechanism could effectively insulate the lender from liability.[152] To illustrate,

---

percentage rate of about 520% to 780%. If the borrower has not repaid the lender by the due date, the lender can negotiate the check. Alternatively, the borrower may be able to extend the loan's due date by paying a fee. This type of extension is referred to as a renewal or a rollover.

Id. (footnote omitted).

143. *Id.*

144. *Id.*

145. *Id.* at 872. The agreements' class action waiver provided: "YOU ARE WAIVING YOUR RIGHT TO SERVE AS A REPRESENTATIVE, AS A PRIVATE ATTORNEY GENERAL, OR IN ANY OTHER REPRESENTATIVE CAPACITY, AND/OR TO PARTICIPATE AS A MEMBER OF A CLASS OF CLAIMANTS, IN ANY LAWSUIT FILED AGAINST US AND/OR RELATED THIRD PARTIES." *Id.* The agreements also contained a class arbitration waiver: "THE ARBITRATOR SHALL NOT CONDUCT CLASS ARBITRATION; THAT IS, THE ARBITRATOR SHALL NOT ALLOW YOU TO SERVE AS A REPRESENTATIVE, AS A PRIVATE ATTORNEY GENERAL, OR IN ANY OTHER REPRESENTATIVE CAPACITY FOR OTHERS IN THE ARBITRATION." *Id.* at 872 n.2.

146. *See id.* at 873.

147. Jenkins v. First Am. Cash Advance of Ga., LLC, 313 F. Supp. 2d 1370, 1374 (S.D. Ga. 2003), *rev'd* 400 F.3d 868 (11th Cir. 2005).

148. *Id.*

149. *See id.* at 1374–76.

150. *Id.* at 1374–75.

151. *Id.* at 1375.

152. *Id.*

assume that Georgia usury law governs[153] and the interest collected on the payday loans exceeds the legal limit. Further assume that Jenkins received eight $400 loans and owed 20% interest over a two-week period. In other words, fourteen days after receiving her first loan, she would owe the lender $480. Assuming she paid on time and did not carry over the balance on each loan, the interest paid on the eight loans would total $640. Thus, even if Jenkins was able to recover the interest paid plus treble damages, there would not be sufficient incentive for the lawsuit to be prosecuted on an individual basis. However, given that many borrowers signed identical loan documents agreeing to the same, allegedly usurious interest rates, even if only twenty of the borrowers were able to proceed collectively against the lender, the lawsuit would become worthwhile. Yet, by having Jenkins and other borrowers waive their right to proceed collectively, the lender thereby rendered the usury laws virtually unenforceable and insulated itself from any liability for charging excessive interest rates. In effect, by prohibiting aggregation of claims, the lender was able to disincentivize prosecution of the usury laws and, correspondingly, exempt itself from the reach of usury regulations.

Although the district court declined to enforce the arbitration agreements on the basis of unconscionability, the Eleventh Circuit reversed and remanded with instructions to grant defendants' motion to compel arbitration.[154] The Eleventh Circuit refused to entertain the plaintiffs' procedural unconscionability arguments because it characterized them as directed at the contract generally, not just at the arbitration clause, and thus held that it was an issue for the arbitrator to decide.[155] The appellate court did, however, address the substantive unconscionability of the collective action waivers.[156] In reversing the district court's substantive unconscionability finding and upholding the collective action waivers, the Eleventh Circuit noted that the arbitration agreements and usury statutes allowed for the recovery of expenses and attorneys' fees.[157] Thus, the court held that the statutory availability of

---

153. Georgia usury laws might not have applied to this dispute because the loan documents included a choice of law provision stating that South Dakota law governed. Jenkins v. First Am. Cash Advance of Ga., LLC, 400 F.3d 868, 871 (11th Cir. 2005). Of course, the choice of South Dakota was no coincidence as South Dakota generally has no usury limits. *See* S.D. CODIFIED LAWS § 54-3-1.1 (2004) The relevant statute provides the following minimal conditions for bypassing usury limits:

> 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.

Id.

154. *Jenkins*, 400 F.3d at 882–83.

155. *Id.* at 876–77.

156. *Id.* at 877–78.

157. *Id.* at 878 (citing GA. CODE ANN. § 16-14-6(c)) ("The Arbitration Agreements expressly permit Jenkins and other consumers to recover attorneys' fees and expenses '[i]f allowed by statute or applicable law.' Under the Georgia RICO statute, a prevailing plaintiff

costs and attorneys' fees ensured that there was incentive for representation of borrowers on an individual basis and, thus, no barrier to the vindication of those borrowers' substantive rights.[158]

The Eleventh Circuit's analysis failed to recognize, however, that without the possibility of aggregation, fewer (and perhaps no) attorneys would take the case given the risk involved. By aggregating claims, a class action may produce a sizeable fund to compensate both the class members and their attorneys. In a similar case addressing whether a collective action waiver in a payday loan was unconscionable, the New Jersey Supreme Court stated that the "'substantial fund' not only covers the attorney's actual fees, but also provides incentive in the form of possible contingency fees for attorneys to risk the prospect of receiving no recovery for their efforts."[159]

In addition, the arbitration of claims on an individual basis might not result in meaningful enforcement of the law. As Professor Sternlight has argued, a few successful suits for individual relief might not induce a company to make company-wide policy changes to comply with the substantive law.[160] In this regard, liability to a few individual plaintiffs does not equate to general accountability, especially when the company's gain from the wrongdoing is an aggregate one. Thus, Jenkins might be able to recover the excessive interest the payday lender charged her, but she might not be able to effect a change in the lender's company-wide policy concerning the interest rates it charges borrowers.

Moreover, given that the *Discover Bank*[161] line of reasoning heavily considers the incentives to bring the lawsuit, many courts have focused on the size of the individual claims.[162] This focus has not always fared well for employees seeking to bring a class action or class arbitration.[163] If a plaintiff's individual claim would provide a large enough monetary award, a court is less likely to view the collective action waiver as an effective bar to substantive relief. For example, in *Konig v. U-Haul Co.*,[164] the plaintiff employee sought to commence class action against his employer, U-Haul, for unpaid wages and

---

may be awarded attorney's fees.") (alterations in original).

158. *Id.*

159. Muhammad v. County Bank of Rehoboth Beach, 912 A.2d 88, 97 (N.J. 2006) (quoting *In re* Cadillac, 461 A.2d 736, 741 (N.J. 1983)).

160. Sternlight & Jensen, *supra* note 120, at 90–91 ("A company may find it worthwhile to pay off a few individual claims but keep its overall policy.").

161. Discover Bank v. Superior Court, 113 P.3d 1100, 1105 (Cal. 2005).

162. *See, e.g.*, Konig v. U-Haul Co. of Cal., 52 Cal. Rptr. 3d 244, 251–52 (Ct. App. 2006), *rev'd*, 153 P.3d 955 (Cal. 2007) ("[P]laintiff failed to establish 'predictably . . . small amounts' of damages payable to class members are at issue . . . ." (quoting *Discover Bank*, 113 P.3d at 1110)); Gentry v. Superior Court, 37 Cal. Rptr. 3d 790, 794–95 (Ct. App. 2006), *rev. granted* 135 P.3d 1 (Cal. 2006) (refusing to hold unconscionable a class action waiver where plaintiff "alleged statutory violations that could result in substantial damages and penalties should he prevail on his individual claims").

163. *See, e.g.*, cases cited *supra* note 162.

164. 52 Cal. Rptr. 3d 244 (Ct. App. 2006).

unfair business practices.[165] However, because each plaintiff's individual claim could exceed $1,000, the California appellate court enforced the collective action waiver in the plaintiff's employment agreement.[166] Relying on *Discover Bank*, the court reasoned that each of the individual plaintiffs' damages was not "predictably . . . small" enough that the inability to proceed collectively would insulate the employer from liability.[167]

The California Supreme Court, however, may have recently taken the focus of the collective action waiver analysis in a new (or slightly revised) direction. In *Gentry v. Superior Court*, the court addressed the enforceability of a class action waiver in a standardized employment contract.[168] The court noted that a finding of procedural unconscionability was not required to invalidate a collective action waiver that implicated unwaivable statutory rights: the plaintiffs' claimed violations of California's overtime laws.[169] Apparently moving beyond the concerns enunciated in *Discover Bank*,[170] a majority of the California Supreme Court articulated the following standard to determine a class action waiver's enforceability: whether "class arbitration would be a significantly more effective way of vindicating the rights of affected employees than individual arbitration."[171] The court enunciated certain factors to aid in this determination, including "the modest size of the potential individual recovery [and] the potential for retaliation against members of the class."[172] The court's decision, however, seems to be narrowly limited to overtime pay claims, or at least to collective action waivers where the underlying substantive claim was based on an unwaivable statutory right. The reach and significance of *Gentry* will only become apparent as it is applied in California courts and interpreted by other jurisdictions.[173]

The norm has been to use the unconscionability analysis to police these problematic clauses; the results of this case-by-case analysis are dizzying. After reviewing many of the numerous cases concerning collective action waiver to determine whether such a waiver in a cellular telephone contract was unconscionable, the Supreme Court of Illinois recently concluded that "it is not useful to do a simple head count of the number of state courts to have ruled a

---

165. *Id.* at 246.

166. *Id.* at 251–53.

167. *Id.* at 252 (quoting *Discover Bank*, 113 P.3d at 1110).

168. 165 P.3d 556, 559 (Cal. 2007).

169. *Id.* However, a finding of procedural unconscionability "is a prerequisite to determining that the *arbitration agreement* as a whole is unconscionable." *Id.* (emphasis added).

170. 113 P.3d 1100.

171. *Gentry*, 165 P.3d at 559.

172. *Id.* at 568.

173. *See* Samuel Estreicher & Steven C. Bennett, *California Court Creates Class-Arbitration Waiver Test*, 238 N.Y.L.J. at *3 (2007), *available at* http://www.jonesday.com (search for "class-arbitration waiver test"; then follow "EstreicherBennet110807" hyperlink).

certain way on class action waivers."[174] For example, collective action waivers in AT&T's substantially similar (if not identical) standardized customer services agreements have, under substantially similar facts, been enforced by some courts and deemed unconscionable by others.[175] The same panoply of disparate outcomes happened with the collective action waiver in the standardized contract of Cingular Wireless.[176] Indeed, the case law largely defies any reliable pattern.[177]

In short, to the extent that the courts enforce collective action waivers, which are effected through adhesion contracts, the availability of collective action may be written out of the procedural law by private contract. Consequently, as Professor Gilles predicts, the class action could meet its

---

174. Kinkel v. Cingular Wireless LLC, 857 N.E.2d 250, 271 (Ill. 2006).

175. *See, e.g.*, Ting v. AT&T, 319 F.3d 1126 (9th Cir. 2003). Applying California law, the Ninth Circuit found the collective action waiver in AT&T's consumer services agreement procedurally unconscionable as consumers had to either accept the waiver through a contractual amendment or cancel their service. *Id.* at 1149. The court also found the waiver substantively unconscionable because of a lack of mutuality—the proscription on class relief affected only the customers as there was no chance AT&T would file a class action against its customers. *Id.* at 1150 & n.14.

However, in *Ragan v. AT&T Corp.*, 824 N.E.2d 1183 (Ill. App. Ct. 2005), an Illinois Appellate Court, applying New York law under a choice of law provision, addressed essentially the same boilerplate collective action waiver as addressed in *Ting*, 319 F.3d 1126. *See Ragan*, 824 N.E.2d at 1189. Nevertheless, the Illinois court held that the provision was not procedurally unconscionable because, although presented as an amendment after the contract's formation, the consumer had an opportunity to reject the amendment by cancelling the cellular telephone service. *See id.* The Illinois court also held that the provision was not substantively unconscionable. *See id.* at 1193–97.

176. *See* Iberia Credit Bureau, Inc. v. Cingular Wireless LLC, 379 F.3d 159, 174 (5th Cir. 2004) (enforcing the provision); Merritt v. Cingular Wireless LLC, No. B178747, 2006 WL 2744357, at *1 (Cal. Ct. App. Sept. 27, 2006) (refusing to enforce the provision); Paton v. Cingular Wireless, No. A108816, 2006 WL 1413537, at *1 (Cal. Ct. App. May 23, 2006) (refusing to enforce the provision); Kinkel v. Cellular Wireless, LLC, 828 N.E.2d 812, 824 (Ill. App. Ct. 2005) (refusing to enforce the provision).

177. *See generally* Martin C. Bryce, Jr., *Red State Versus Blue State: Surprisingly Most (But Not All) Courts in Both "Red" and "Blue" States Enforce Express Class Action Waivers in Consumer Arbitration Agreements*, 59 CONSUMER FIN. L.Q. REP. 222, 223 (2005) (explaining that the majority, but not all federal courts, enforce express class action waivers in consumer arbitration agreements); Alan S. Kaplinksy & Mark J. Levin, *Consensus or Conflict? Most (But Not All) Courts Enforce Express Class Action Waivers in Consumer Arbitration Agreements*, 60 BUS. LAW. 775, 776 (2005) ("[T]he absence of definitive guidance from the nation's highest court [concerning whether express collective action waivers are enforceable under the Federal Arbitration Act] leaves room for disagreement by the lower federal courts and the state courts as to the enforceability of express class action waivers."); William M. Howard, Annotation, *Validity of Arbitration Clause Precluding Class Actions*, 13 A.L.R. 6th 145 (2006) (surveying "the state and federal cases that consider whether an arbitration clause in a contract . . . renders the agreement to arbitrate unconscionable or otherwise unenforceable").

"near-total demise."[178] This ability to prevent collective action is particularly troublesome because it allows corporations to circumvent fairly-debated and deliberatively-enacted legislation[179] and could, in effect, enable corporate "self-deregulation."[180] Many of the substantive policies against usurious interest rates,[181] deceptive bank fees, and payment of fair wages are the product of carefully debated legislation. Likewise, the class action system exists through the federal rule-making process, which is quasi-legislative.[182]

B. Limits on Discovery

Part of the "simplicity, informality, and expedi[ency]"[183] of arbitration is the relative lack of procedure—that is, the aspects of litigation that are implicitly relinquished with the agreement to arbitrate. Without the right to a jury trial, or the right to appeal an arbitral award, the adjudicatory process is presumed to be faster and less expensive.[184] Another common example of the added expediency of arbitration is that discovery is generally more limited in this context than in the courts.[185]

The FAA does not address the availability of discovery in arbitration, and absent a statutory or contractual right to discovery, a party has no legal entitlement to discovery prior to the arbitration hearing.[186] For the most part, if

---

178. *See generally* Gilles, *supra* note 19, at 373, 375 (arguing that "with a handful of exceptions, class actions will soon be virtually extinct").

179. Taylor & Cliffe, *supra* note 16, at 1099, 1100–01 (describing standardized arbitration and pre-litigation agreements as end runs around fairly-enacted legislation).

180. The term "self-deregulation" is borrowed from David A. Schwartz's comprehensive critique of the pre-dispute arbitration regime. Schwartz, *supra* note 23, at 37 ("The enforcement of adhesive arbitration clauses allows firms to lessen the regulatory impact of statutory claims—in short, to deregulate themselves.").

181. *See, e.g.*, TENN. CODE ANN. § 47-14-103 (2007) (setting the maximum interest rate at ten percent).

182. *See* Rules Enabling Act, 28 U.S.C. §§ 2071–77 (2000) (giving judicial branch power to promulgate federal rules of civil procedure); Paul R. Rice, *Back to the Future with Privileges: Abandon Codification, Not the Common Law*, 38 LOY. L.A. L. REV. 739, 743 (2004) (describing the federal rule-making process through Advisory Committee as "quasi-legislative"). *See generally* Peter G. McCabe, *Renewal of the Federal Rulemaking Process*, 44 AM. U. L. REV. 1655 (1995) (describing federal rule-making process, its historical background, and future initiatives).

183. *See* Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 31 (1991) ("[B]y agreeing to arbitrate, a party 'trades the procedures and opportunity for review of the courtroom for the simplicity, informality, and expedition of arbitration.'" (quoting Mitsubishi Motors Corp. v. Soler Chrysler-Plymouth, Inc., 473 U.S. 614, 628 (1985))).

184. *See generally* Dawes, *supra* note 29, at 603 ("The risks [of arbitration include] lack of appeal rights [and] waiver of other procedural and substantive rights . . . ."); Schwartz, *supra* note 23, at 60–61 (listing "[w]hy corporate defendants like arbitration").

185. *Gilmer*, 500 U.S. at 31.

186. 21 SAMUEL WILLISTON & RICHARD A. LORD, A TREATISE ON THE LAW OF CONTRACTS §

the parties' agreement is silent concerning the availability of discovery or incorporates institutional rules such as those of the American Arbitration Association (AAA), the question will be left to the discretion of the arbitrator.[187] In determining whether a party is entitled to document requests or depositions, the arbitrator is presumably weighing the expediency and cost-effectiveness of the arbitral forum against the need of a party to obtain the information necessary to prove her claim.[188]

In *Gilmer v. Interstate/Johnson Lane Corp.*, the United States Supreme Court held that limitations on discovery do not necessarily render an arbitration clause invalid.[189] Robert Gilmer claimed he was terminated as Manager of Financial Services at Interstate/Johnson Lane Corporation because of his age, in violation of the Age Discrimination in Employment Act of 1967 (ADEA).[190] As part of his employment, Gilmer was registered with the New York Stock Exchange (NYSE).[191] As required in part of the standardized registration application, Gilmer agreed to arbitrate any disputes arising out of his employment pursuant to NYSE rules.[192] The NYSE rules governing arbitration allowed for limited discovery, including "document production, information requests, depositions, and subpoenas."[193] The Court held that that the NYSE's discovery provision was sufficient to allow Gilmer a fair opportunity to present his claims against his employer under the ADEA.[194]

Unlike the NYSE rules, however, not all pre-dispute arbitration clauses allow for the possibility of basic discovery devices. The following examples illustrate how some corporations have used pre-dispute arbitration clauses to substantially limit the availability of discovery or to eliminate it entirely. Plaintiffs have responded by challenging such limitations on discovery as unconscionable.

57:90, at 511 (4th ed. 2001) (citing Burton v. Bush, 614 F.2d 389 (4th Cir. 1980)).

187. REVISED UNIF. ARBITRATION ACT § 17 (2000); Transwestern Pipeline Co. v. Blackburn, 831 S.W.2d 72, 78 (Tex. App. 1992); 21 WILLISTON & LORD, *supra* note 186, § 57:90, at 511–12 (citing Stanton v. Paine Webber Jackson & Curtis, Inc., 685 F. Supp. 1241 (S.D. Fla. 1988)). For example, in employment disputes under agreements referencing the AAA Rules, the arbitrator has discretion to allow the employee's use of many basic discovery tools. *See* AAA, EMPLOYMENT ARBITRATION RULES & MEDIATION PROCEDURES R.9, *available at* http://www.adr.org/sp.asp?id=28481 ("The arbitrator shall have the authority to order such discovery, by way of deposition, interrogatory, document production, or otherwise, as the arbitrator considers necessary to a full and fair exploration of the issues in dispute, consistent with the expedited nature of arbitration.").

188. REVISED UNIF. ARBITRATION ACT § 17(c) (2000).

189. *See Gilmer*, 500 U.S. at 31.

190. *Id.* at 23–24.

191. *Id.*

192. *Id.*

193. *Id.* at 31.

194. *Id.*

In *Ostroff v. Alterra Healthcare Corp.*,[195] plaintiff Lillian Restine signed a thirty-one-page "Residency Agreement" on the day she moved into the defendant's assisted living facility.[196] The agreement provided that any claims relating to Restine's residency at the facility, except eviction, would be submitted to binding arbitration.[197] Among other things, the arbitration clause provided that "[d]iscovery in the arbitration proceeding is governed by the Pennsylvania Rules of Civil Procedure, but on a shortened timeline, and with only depositions of experts allowed."[198]

Ms. Restine sued the corporation, alleging that she was injured by the negligent action of an employee of the residence facility, "suffered a broken hip and numerous other physical and mental injuries, including memory loss," and was unable to walk as a result of the accident.[199] Pursuant to the Residency Agreement, the defendant corporation moved to compel arbitration.[200] Restine then challenged the arbitration clause on unconscionability grounds, pointing to the limitation that she could only depose expert witnesses.[201]

The district court denied the corporation's motion to compel arbitration, noting that, while limitations on discovery are not per se invalid, the clause in this case was severely restrictive and, thus, substantively unconscionable.[202] The restriction preventing Restine from deposing lay witnesses would limit her access to information and "put her at a distinct disadvantage in arbitration."[203] Unlike the plaintiff in *Gilmer*, the restriction to depose only experts "may well deny Restine a 'fair opportunity to present [her] claims.'"[204] Indeed, due to the nature of Restine's accident,[205] and with only expert depositions, it would likely be impossible to gather the information necessary to prove the claim.

The court reached the appropriate result in *Ostroff* and recognized the corporation's overreaching in limiting discovery. Other jurisdictions, in various contexts, have found arbitration clauses substantively unconscionable based in part on similar, pre-dispute limitations on discovery.[206] Generally,

---

195. 433 F. Supp. 2d 538 (E.D. Pa. 2006).

196. *Id.* at 540.

197. *Id.* at 541.

198. *Id.* (citation omitted).

199. *Id.* at 540–41.

200. *Id.* at 541.

201. *Id.* at 541, 545.

202. *Id.* at 545.

203. *Id.*

204. *Id.* (quoting Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 31 (1991) (alteration in original)).

205. Restine's injuries were allegedly caused by an employee who swung open Restine's door without knocking, striking Restine and causing her to fall. *Id.* at 540.

206. *See* Walker v. Ryan's Family Steak Houses, Inc., 400 F.3d 370, 387 (6th Cir. 2005); Domingo v. Ameriquest Mortgage Co., 70 F. App'x 919, 920 (9th Cir. 2003); Booker v. Robert Half Int'l, 315 F. Supp. 2d 94, 103–04 (D.C. 2004); Geiger v. Ryan's Family Steak Houses, Inc., 134 F. Supp. 2d 985, 996 (S.D. Ind. 2001); Hooters of Am. v. Phillips, 39 F. Supp. 2d 582, 614 (D.S.C. 1998); Hoffman v. Cargill, Inc., 968 F. Supp. 465, 475 (N.D. Iowa 1997).

courts have treated discovery limitations on a case-by-case basis under the unconscionability doctrine, looking to the *Gilmer* standard that requires a plaintiff have a "fair opportunity" to present her claims.[207]

Yet, not all unduly restrictive discovery limitations have been invalidated. For example, in *Bar-Ayal v. Time Warner Cable, Inc.*, plaintiff Shlomo Bar-Ayal brought a putative class action against Time Warner, a cable provider, for alleged practices of improperly levying additional charges against customers.[208] The service agreement's arbitration clause not only waived collective action and shortened the statute of limitations but also eliminated pre-hearing discovery.[209] The court cited *Gilmer*[210] in holding that the discovery limitation was allowable,[211] but did not discuss the *Gilmer* standard that such limitations should not prevent plaintiffs from fairly presenting their claims.[212]

Moreover, in *Pony Express Courier Corp. v. Morris*, a court again upheld an arbitration clause limiting discovery.[213] In that case, employee Diane Morris brought an action against her employer alleging sexual harassment.[214] The employer moved to compel arbitration pursuant to the parties' arbitration agreement.[215] Although the agreement eliminated discovery and the court conceded that "binding arbitration seem[ed] harsh," it held that the limitations were not unconscionable and enforced the arbitration clause.[216]

A corporation might not use an arbitration clause to attempt to eliminate discovery entirely, but it could place an onerous burden on a plaintiff seeking discovery. For example, in *Martinez v. Master Protection Corp.*, employee Tony Martinez sued his employer Master Protection Corporation, alleging various claims based on national origin discrimination.[217] Among other things, the parties' arbitration agreement limited discovery to one single deposition unless there was a showing of "substantial need."[218] Martinez successfully challenged the arbitration clause on unconscionability grounds—not on the basis of the discovery limitation, however.[219] The court was persuaded that the clause was substantive unconscionability because its one-sidedness was evidence by the "lack of mutuality," a shortened statute of limitations, and unduly burdensome costs.[220] Thus, the court found it unnecessary to decide

207. *See* Ostroff v. Alterra Healthcare Corp., 433 F. Supp. 2d 538, 546 (E.D. Pa. 2006).
208. No. 03 CV 9905(KMW), 2006 WL 2990032, at *1 (S.D.N.Y. Oct. 16, 2006).
209. *Id.* at *5.
210. Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 31 (1991).
211. *Bar-Ayal*, 2006 WL 2990032, at *16.
212. *Gilmer*, 500 U.S. at 31.
213. 921 S.W.2d 817, 822 (Tex. App. 1996).
214. *Id.* at 819.
215. *Id.*
216. *Id.* at 822.
217. 12 Cal. Rptr. 3d 663, 667 (Ct. App. 2004).
218. *Id.* at 672.
219. *Id.* at 672–73.
220. *Id.* at 673.

whether the discovery restrictions, themselves, "prevent[ed] Martinez from vindicating his rights."[221] The court noted, however, that, "considered against the backdrop of the other indisputably unconscionable provisions, the limitations on discovery do . . . compound the one-sidedness of the arbitration agreement."[222]

The analysis of the *Martinez* court elucidates the inadequacy of this ad hoc treatment of limitations on discovery. On one hand, the court acknowledged that adequate discovery was "indispensable" for Martinez to vindicate his statutory rights.[223] On the other hand, it also acknowledged "discovery *limitations* are an integral part of the arbitration process."[224] Addressing this balance, the court reasoned that, "given the relatively straightforward allegations of misconduct," the discovery limitation did not, as a matter of law, prevent Martinez from vindicating his statutory rights.[225]

The reasoning used to justify the *Martinez* holding is problematic because the discovery restrictions are set by a form arbitration clause well before the dispute arises, and the employer is presumably in a much better position than the employee to predict the types of disputes that are likely to arise. At the time Martinez signed his employment contract containing the arbitration clause, he likely did not anticipate that a dispute would arise and, if so, whether it would be the type requiring extensive discovery to vindicate his rights. Whether or not the dispute turns out to be "straightforward"[226] cannot be fairly assessed until the dispute arises.

Given, as the court acknowledged, that these limitations on discovery could put a plaintiff employee at a "serious disadvantage"[227] in preparing his case, it

---

221. *Id.*

222. *Id.*

223. *Id.* at 672 (quoting Armendariz v. Found. Health Psychcare Servs., Inc., 6 P.3d 669, 683 (Cal. 2000)).

224. *Martinez*, 12 Cal. Rptr. 3d at 672 (emphasis added) (citation omitted).

225. *Id.* at 673. The court stated:

> We recognize that, in many employment disputes, restricting a plaintiff to a single deposition and document request could place him at a serious disadvantage if testimony from numerous witnesses is necessary to prepare his case. We are also aware the same restriction could operate to the employer's advantage, because it has ready access to most of the relevant documents and many of the witnesses remain in its employ. Consequently, the employer typically has far less need for discovery in order to prepare for arbitration then [sic] the employee.
>
> However, given the relatively straightforward allegations of misconduct involved in this action, and the possibility that proof of Martinez's Labor Code claims will rest largely on documentation rather than testimony, we are unable to state as a matter of law that [the employer's] "arbitration agreement does not afford adequate discovery rights to employees seeking to vindicate statutory rights . . . ."

Id. at 672–73 (citations omitted).

226. *Id.* at 673.

227. *Id.* at 672.

seems patently unjust that the relative "straightforward[ness]"[228] of the claims saves the limitation from being deemed unconscionable, especially when unconscionability is ordinarily assessed as of the time the agreement is made.[229] Moreover, such limitations essentially place a difficult burden of proof on the plaintiff in two regards—first, the plaintiff must show that the discovery limitations are unconscionable and, if that is unsuccessful, second, the plaintiff must show that she has a substantial need for additional deposition testimony. As a result, the burden of litigation (and proof) falls on the party that is least able to bear it financially.

Certainly arbitration is more efficient and expedient if discovery is limited; however, these benefits should not come at the expense of procedural fairness, the enforcement of substantive laws, or both. Implicit in the agreement to arbitrate is some informality and limitation on discovery; however, at some point, too many limitations on discovery will defeat the purpose of arbitration as a forum to hear the plaintiff's claims. By expressly limiting discovery, and thereby removing a determination of the availability of discovery from the arbitrator's discretion, a corporation may effectively weaken the enforcement of substantive laws.

C. Shortening the Statute of Limitations

Another example of a pre-dispute limitation is the shortening of the applicable statute of limitations,[230] which plaintiffs have challenged as unconscionable.[231] As with limitations on discovery, the courts have reached different conclusions concerning the enforceability of shortened statutes of limitations. Some courts have allowed companies to use their arbitration clauses to effectively shorten a statute of limitations,[232] while others have declined to enforce these provisions based on the doctrine of unconscionability.[233]

---

228. *Id.* at 673

229. 8 WILLISTON & LORD, *supra* note 83, § 18:12, at 77 ("The determination of whether a given clause or contract is in fact unconscionable is to be made at the time of its making rather than at some subsequent point in time.") (footnote omitted); *see also In re* FirstMerit Bank, N.A., 52 S.W.3d 749, 757 (Tex. 2001) (unconscionability assessed based on "circumstances existing when the parties made the contract").

230. *See, e.g.*, Soltani v. W. & S. Life Ins. Co., 258 F.3d 1038, 1042–45 (9th Cir. 2001) (requiring that suits be brought within six months).

231. *See, e.g., id.* at 1042 ("[Plaintiffs] . . . argued under a . . . general unconscionability analsysis that they were presented with contracts of adhesion, could not negotiate terms, and thus should not be held to the shortened limitations period.").

232. *See, e.g., id.* at 1044–45; Aull v. McKeon-Grano Assocs., No. 06-2752, 2007 WL 655484, at *8 (D.N.J. Feb. 26, 2007); Bar-Ayal v. Time Warner Cable, Inc., No. 03 CV 9905, 2006 WL 2990032, at *5, 16 (S.D.N.Y. Oct. 16, 2006); *In re* Standard Meat Co., No. 05-06-01470-CV, 2007 WL 730660, at *3–5 (Tex. App. Mar. 9, 2007).

233. *See, e.g.*, Davis v. O'Melveny & Myers, 485 F.3d 1066, 1070 (9th Cir. 2007); Parilla v. IAP Worldwide Servs. VI, Inc., 368 F.3d 269, 278 (3d Cir. 2004); Ingle v. Circuit

For example, in *In re Standard Meat Co.*, employee Adriana Chagoya, sued her employer for negligence based on injuries she allegedly sustained while working on a food assembly line.[234] Her employer moved to compel arbitration, referring to the arbitration agreements signed by Chagoya when she applied for the job and during the orientation process.[235] Chagoya argued the agreements were unconscionable on numerous grounds, one of which was that the agreement gave her one year to file a notice of her intent to arbitrate and, thereby, effectively shortened the statute of limitations on her claims.[236] The Texas appellate court enforced the clause and noted that, in *EZ Pawn Corp. v. Mancias*,[237] the Texas Supreme Court allowed the modification of a statute of limitations on a wrongful discharge claim through an arbitration agreement.[238] Indeed, in *EZ Pawn*, that court enforced a pre-dispute arbitration clause requiring an employee to initiate arbitration within 180 days of the date that the claim accrued.[239]

In contrast, the Ninth Circuit recently declined, on unconscionability grounds, to enforce an arbitration clause that, among other things, contained an effective shortening of the statute of limitations.[240] In *Davis v. O'Melveny & Myers*, Jacqueline Davis, a paralegal, sued her employer for violations of the Fair Labor Standards Act (FLSA).[241] However, an arbitration clause in the firm's "Dispute Resolution Program" governing employees provided that the employee had one year to give notice of any claim that is "known to the employee or with reasonable effort . . . should have been known to him or her."[242]

The court expressed concern that this "notice provision" had the effect of shortening the limitations period by effectively barring the employee from pressing a "continuing violations" theory, which allows consideration of related acts that began prior to the limitations period if they constitute part of a "systematic policy of discrimination."[243] The court found it particularly

---

City Stores, Inc., 328 F.3d 1165, 1173 (9th Cir. 2003); *Martinez*, 12 Cal. Rptr. 3d at 669; Covenant Health Rehab of Picayune, L.P. v. Brown, 949 So.2d 732, 739 (Miss. 2007).

234. *In re Standard Meat Co.*, 2007 WL 730660, at *1.

235. *Id.*

236. *Id.* at *4.

237. 934 S.W.2d 87, 89 (Tex. 1996).

238. *In re Standard Meat Co.*, 2007 WL 730660 at *4. The court noted that the arbitrator should determine the appropriate limitations period. *Id.*

239. 934 S.W.2d at 89. In *EZ Pawn*, while preparing for depositions, the employer's counsel realized an arbitration agreement existed and, thus, only first moved to compel arbitration after litigating the case in court for ten months. *Id.*

240. Davis v. O'Melveny & Myers, 485 F.3d 1006, 1070 (9th Cir. 2007).

241. *Id.*

242. *Id.* at 1071. This provision effectively limited the statute of limitations because, under the FLSA, the limitations period is either two or three years, depending on the type of violation. Fair Labor Standards Act, 29 U.S.C. § 255(a) (2000) (providing that the statute of limitations is two years, unless the violation is willful, in which case it is three years).

243. *Davis*, 485 F.3d at 1077.

troublesome that the notice period in the invalidated arbitration clause ran from the date that the employee knew or should have known of the alleged violation.[244] On this basis, the court distinguished *Davis* from other California cases upholding arbitration clauses that shortened the statute of limitations to a six-month period.[245] The *Davis* court reasoned that the six-month limitations periods in those cases were reasonable because they ran from the date the employee left employment and, thus, did not bar a "continuing violations" theory.[246]

However, the potential bar of a "continuing violations" theory, while troubling, should not drive the unconscionability analysis. Rather, it certainly is arguable that the cases allowing a six-month limitations period were wrongly decided and the provisions should not have been enforced, even though they measured the limitations period from the last day of employment. In other words, even narrow statutes of limitations that do not bar a continuing violations theory should be struck down. Any *pre-dispute* shortening of a statute of limitations not permitted by the substantive statute should be considered suspect. Otherwise, as with class action waivers and discovery limitations, corporations may use such provisions in standard form agreements to circumvent fairly-debated and deliberatively-enacted legislation.[247] First, businesses may use standard form agreements as an end run around legislatively enacted limitations periods. Second, these shortened limitations periods can thwart the enforcement of substantive policies against, for example, unfair labor practices. While perhaps not as obvious or reliable as limitations on collective action, a shortened statute of limitations does surreptitiously, in effect, serve to weaken the remedial and deterrent functions of underlying substantive laws.

III. THE PROBLEMS PRESENTED BY EXPRESS, PRE-DISPUTE LIMITATIONS OF PROCESS AND THE USE OF THE UNCONSCIONABILITY DOCTRINE TO POLICE THEM

The Supreme Court's policy favoring arbitration should not be repeated as an empty mantra to support the enforcement of all limitations contained within

---

244. *Id.*

245. *Id.* (citing Soltani v. W. & S. Life Ins. Co., 258 F.3d 1038, 1044 (9th Cir. 2001)).

246. *Davis*, 485 F.3d at 1077. The court discerned:

> The time to file [in *Soltani*] did not depend upon when the employee knew of the claim, or otherwise when it arose. A three-year old claim could still be filed, as long as it was also filed within six-months from when the employee stopped working (and as long as it was not otherwise barred by the relevant statute of limitations). This type of provision does not raise the concerns about nullifying the "continuing violations" theory, as the employee would during that six-month period still be able to take full advantage of the ability to reach back to the start of the violation.

Id.

247. *See* Schwartz, *supra* note 23, at 37; Taylor & Cliffe, *supra* note 16, at 1100.

arbitration clauses. Many of the critiques that have been aimed generally at pre-dispute arbitration, especially in the consumer and employment contexts, can likewise be aimed at the express, procedural limitations contained within those clauses. At the same time, it is conceptually possible to accept the critiques of additional, pre-dispute limitations and still allow the continuation of the current arbitration regime. This is because these limits on process can work as an effective barrier to holding corporations accountable. Thus, the Court's stated policy can be upheld by enforcing arbitration clauses in general, while striking down certain other express limitations contained therein.

The aims of arbitration are not necessarily defeated when certain express limitations within arbitration clauses are struck down. Curtailing pre-dispute limitations on discovery, for example, would not impede the policy generally favoring arbitration. Likewise, refusing to enforce collective action waivers is not necessarily inconsistent with the efficiency goals of arbitration—actually, such refusal may advance efficiency by allowing many similar individual disputes to proceed in a class arbitration. At some point, certainly, procedural fairness and corporate accountability outweigh the efficiencies associated with arbitration. Thus, these perceived benefits of arbitration should not be a distraction from its underlying purpose: providing an alternative forum *to present substantive claims*.

The problem with express pre-dispute limitations is compounded by some of the objections to pre-dispute arbitration more generally. Namely, corporations that draft arbitration clauses into their standardized agreements are usually in a position of superior bargaining power, with a wider knowledge of the intricacies of the deal and the potential disputes that might arise. As "repeat-players" in the marketplace, these businesses also have more incentive to keep disputes out of court, as well as more resources to invest toward this goal.[248] The potential for corporate abuse of express, pre-dispute limitations is compounded by the fact that the vast majority of arbitration clauses are contained in contracts of adhesion, which "bear little resemblance to the voluntary agreements envisioned when one thinks of 'consent.'"[249]

The idea of consent in this context is mythical and, thus, so too is the notion of party autonomy. To assess whether the parties formed a binding arbitration clause, the courts have looked to contractual standards of assent: an objective manifestation of a willingness to enter into a bargain, whether or not the party has read or understood the arbitration clause.[250] This standard has allowed the enforcement of arbitration clauses such as those contained in the now infamous *Gateway* "terms and conditions," which were presented to a

248. *See* Alderman, *supra* note 27, at 1253–58 (discussing the benefits of the "repeat-player" in consumer arbitration); *see also* Knapp, *supra* note 13, at 790 (discussing the relationship between possible measures for reform and the "'repeat player' advantage").

249. Alderman, *supra* note 27, at 1247.

250. *See* Stephen J. Ware, *Arbitration Clauses, Jury-Waiver Clauses, and Other Contractual Waivers of Constitutional Rights*, 67 LAW & CONTEMP. PROBS. 167, 170–72 (2004).

consumer only after purchasing the computer and defined "assent" as simply keeping the computer longer than thirty days.[251] Generally, pre-dispute agreements to arbitrate have been treated no differently than any other contract; a party may assent to its terms by signing human resources paperwork, filling out a credit card application, or keeping a product for a certain amount of time.

Critics have made the case for more exacting standards of pre-dispute consent to arbitration because, by agreeing to arbitrate, a party might be waiving the Seventh Amendment right to trial by jury.[252] Some have argued that consent should be "knowing"—that the party agreeing to arbitration knew the clause was contained within the agreement.[253] Others have argued even further that a pre-dispute agreement to arbitration should require "knowing and intelligent" consent such that the party agreeing to arbitration was not only aware of the arbitration clause but understood it as well.[254]

Although these arguments are compelling,[255] courts have consistently upheld pre-dispute, contractual waivers of the right to a jury trial,[256] and the contractual standard of a "manifestation of assent" remains the norm in assessing the validity of arbitration clauses.[257] Moreover, it is not clear that a pre-dispute agreement to arbitrate or to limit certain procedures can be "knowing and intelligent." For example, the very nature of a pre-dispute agreement to eliminate discovery cannot *really* be knowing and intelligent until the underlying dispute becomes known. In other words, one cannot know what discovery might uncover unless discovery is actually conducted.

Thus, because express, pre-dispute limitations in arbitration clauses are often contained in contracts of adhesion and are formed by the contractual standard of objective assent, the idea of party autonomy is strained. Further, the potential for corporate self-deregulation is amplified. Limitations on class relief and discovery, as well as shortened statutes of limitations, may act as a barrier to a party's substantive relief. Moreover, these barriers are being effected in standardized forms that stakeholders generally do not read (or do not understand).[258] To the extent that corporations issue these form contracts en

---

251. *See* Hill v. Gateway 2000, Inc., 105 F.3d 1147, 1147–48, 1151 (7th Cir. 1997).

252. *See generally, e.g.*, Jean R. Sternlight, *Mandatory Binding Arbitration and the Demise of the Seventh Amendment Right to a Jury Trial*, 16 OHIO ST. J. ON DISP. RESOL. 669, 676 (2001) (arguing that "the Seventh Amendment jury trial waiver standard is applicable in many arbitration cases").

253. *See* Ware, *supra* note 250, at 172–76 (discussing Sternlight, *supra* note 252, at 680–710).

254. *See id.* at 175.

255. *But see id.* at 197–204 (arguing that a heightened standard of consent is not required to effectively waive the right to a jury trial).

256. *See* Noyes, *supra* note 33, at 604–07 (discussing the standard of assent to a waiver of the right to a jury trial).

257. *See generally* Ware, *supra* note 250, at 205 (predicting that arbitration clauses will continue to be assessed by a contractual standard of consent).

258. Indeed, the Restatement of Contracts assumes that signatories to standardized form agreements do not read them. *See* RESTATEMENT (SECOND) OF CONTRACTS § 211 cmt. b (1981).

masse, they become the standard for transactions and, in effect, the legislation governing an industry. As illustrated by the examples discussed in this Article, the opportunity to draft standardized arbitration clauses invites companies to privately enact contractual limitations on process, which may effectively weaken the deterrent and remedial effects of existing substantive laws.

Moreover, it is certainly true that the doctrine of unconscionability has served to temper the formalism of arbitration.[259] Some scholars have argued that unconscionability sufficiently polices arbitration abuses.[260] However, the unconscionability doctrine appears ill-equipped to address the express limitations contained within pre-dispute arbitration clauses.

Unconscionability is not readily definable.[261] Indeed, one of the charms of this doctrine is its flexibility to serve as a counterbalance when a contract does not quite involve fraud or duress but would leave the conscience uneasy if it were enforced. The weakness of such a vague standard, however, is its ex ante unpredictability. While the ex ante unpredictability of unconscionability is not unique to arbitration, this frailty has presented itself as a particularly problematic theme in the context of pre-dispute procedural limitations. This is because courts are applying a case-by-case analysis to clauses that most often arise in one-size-fits-all, standardized form agreements. As illustrated in the foregoing examples,[262] this analytical approach leads to inconsistent results among substantially similar (if not identical) facts and, further, defeats the efficiency goals of arbitration. Thus, in any given dispute, it is not certain whether an express contractual limit on procedure contained in a standardized contract will be enforced.

Yet, one of the general goals of contract law is to provide the comfort of certainty in the marketplace.[263] In addition, the enforcement of pre-dispute arbitration clauses seeks to avoid the delay and costs of litigating in court. In this context, neither of these objectives is served by an unconscionability analysis. Rather, the corporations that draft these clauses cannot be certain that their pre-dispute procedural limitations will be enforced. Furthermore, if the limitations are determined to be unconscionable, corporations cannot fairly predict whether the entire arbitration clause will be defeated. Thus, at much time and expense to the corporation and to the contracting stakeholders, both parties are likely to end up in court, litigating the enforceability of the

---

259. *See generally* Stempel, *supra* note 25, at 763 ("[M]any scholars have suggested that unconscionability is simply too plastic a concept that permits too much post-hoc judicial meddling with contracts.").

260. *See, e.g.*, Robert A. Hillman & Jeffrey J. Rachlinksi, *Standard-Form Contracting in the Electronic Age*, 77 N.Y.U. L. REV. 429, 456–58 (2002).

261. Indeed, the Uniform Commercial Code does not even attempt to define "unconscionability." *See* U.C.C. § 2-302 (2005); *see also* Leff, *supra* note 25, at 487 ("If reading this section makes anything clear it is that reading this section alone makes nothing clear about the meaning of 'unconscionable' except perhaps that it is pejorative.").

262. *See supra* Part II.

263. *See supra* note 34 and accompanying text.

procedural limitation. This has the effect of seriously undermining the supposed expediency and cost effectiveness of arbitration.

Even where the unconscionability analysis appears to lead to a fair result when the express limitation is not enforced, unfairness nonetheless results. This is because the plaintiff bore the burden of proving unconscionability and, thus, had to decide to undertake the significant time and costs associated with challenging the provision.[264] This burden, in turn, serves as another barrier to corporate accountability.

Additionally, the unpredictability concerning the enforceability of these clauses, and the attendant worry that they will not be enforced ex ante, has not deterred corporations from drafting such overreaching clauses. Because a plaintiff must prove unconscionability, and because many courts will simply strike an offending clause from the agreement and otherwise compel the parties to proceed in arbitration, corporations have little incentive to refrain from overreaching.

Moreover, the separability doctrine, which dissects the arbitration provision from the contract in which it is contained, presents confusion in the application of the unconscionability analysis. As a threshold matter, courts should be determining the unconscionability of the arbitration clause, not the entire contract. However, it is not always easy to discern where the arbitration clause ends and the rest of the contract begins. For example, where a collective action waiver was contained within a larger contract of adhesion, at least one circuit court has refused to entertain *procedural* unconscionability arguments directed at the contract as a whole, determining that the matter was for an arbitrator to decide, not the court.[265] However, the same court addressed the *substantive* unconscionability of the collective action waiver by parsing it from the agreement.[266] Thus, in states where a showing of both procedural and substantive unconscionability is required, adherence to the separability doctrine can prevent the determination that an arbitration clause is unconscionable.

Further, the focus of the unconscionability doctrine is inapt. The procedural unconscionability inquiry, which asks whether there was an "absence of meaningful choice," is usually simply window dressing because the limitations are contained in adhesion contracts, which do not allow negotiation. Moreover, the substantive unconscionability analysis, with its focus on whether the agreement is one-sided, does not squarely address the serious problem with collective action waivers, discovery limitations, and shortened statute of limitations: they can serve as an effective barrier to the vindication of substantive rights. Thus, for example, in assessing the enforceability of a discovery limitation, it should not matter that, after the fact, the case turns out to be "straightforward." Rather, the inquiry should focus on whether a pre-dispute limitation could, if a dispute arises, fail to provide an alternative forum

264. Indeed, Professor Todd Rakoff has argued that terms in adhesion contracts should be presumptively unenforceable. *See* Rakoff, *supra* note 90, at 1173–74.

265. Jenkins v. First Am. Cash Advance of Ga., LLC, 400 F.3d 868, 877 (11th Cir. 2005).

266. *Id.*

and instead weaken the deterrent and remedial purposes of the underlying substantive law.[267] Because this potential exists wherever a standardized form contract includes a pre-dispute limitation on procedural rights, such clauses should be per se unenforceable.

IV. REGULATING PRE-DISPUTE LIMITATIONS ON PROCESS: BALANCING EFFICIENCY, AUTONOMY AND ACCOUNTABILITY

Assessing whether pre-dispute procedural limitations should be enforced requires a delicate balancing of efficiency and private autonomy against procedural fairness and corporate accountability. The unconscionability doctrine has not been an appropriate or dependable tool to achieve this balance.[268] In light of unpredictable results, the parameters for enforceable pre-dispute arbitration terms need to be more reliably articulated. Drawing on the analogy to exculpatory clauses, one solution calls for prohibition of express, pre-dispute limitations of procedural rights in standardized form agreements. If the parties agree to these limitations once a dispute has arisen, they should be free to do so. However, federal legislation should be enacted to clarify that pre-dispute collective action waivers, limitations on discovery, and shortened statute of limitations in standardized form agreements will not be valid. Congress could, for example, amend the FAA to specify that such limitations

---

267. These concerns are heightened by companies' aggressive use of choice of law and choice of forum clauses. Companies concerned with the enforceability of their pre-dispute procedural limitations can require, in their standard form agreements, that the law of a state with weaker unconscionability jurisprudence be applied. *See generally* William J. Woodward, Jr., *Finding the Contract in Contracts for Law, Forum and Arbitration*, 2 HASTINGS BUS. L.J. 1, 3, 12–13 (2006). Professor Woodward highlights the irony in this approach:

> Very few would think that a vendor could avoid an unconscionability challenge by simply adding a "waiver of unconscionability" to that very form. Yet a modern drafter might very well accomplish the same thing by 'choosing' the law of a place with weaker consumer protection and arguing that, as a matter of contract, the customer is bound by that "choice of law."

Id.

268. This Article is certainly not the first to suggest that the law should look "beyond unconscionability" when assessing pre-dispute limitations of procedural rights. *See generally* J. Maria Glover, Note, *Beyond Unconscionability: Class Action Waivers and Mandatory Arbitration Agreements*, 59 VAND. L. REV. 1735, 1760 (2006) (arguing that courts should consider "whether there is a sufficiently close nexus between the class action waiver and non-waivable substantive rights such that these waivers should not be left to private bargaining"); Robert S. Safi, Note, *Beyond Unconscionability: Preserving the Class Mechanism Under State Law in the Era of Consumer Arbitration*, 83 TEX. L. REV. 1715, 1717–18 (2005) (asserting that "unconscionability serves mainly as a point of departure for the discussion of other, more effective tools available to states that want to protect their consumers from [class action waivers]").

are per se unenforceable,[269] whether they are expressly stated in the agreement or incorporated by reference to institutional rules such as those of the AAA.

As this Article has shown, pre-dispute limitations have the potential to weaken the deterrent and remedial aims of the underlying substantive law and, thus, have been analogized to exculpatory clauses.[270] Therefore, to the extent that the underlying substantive right is a statutory one, the clauses limiting these procedural rights should be unenforceable as a matter of public policy.[271] The law should not permit corporations (or any party) to effectively contract around statutory liability.[272]

Moreover, to the extent the underlying substantive law derives from common law sources, section 195 of the Restatement (Second) of Contracts provides guidance. Under Restatement Second section 195(1), "[a] term exempting a party from tort liability for harm caused intentionally or recklessly is unenforceable on grounds of public policy."[273] Thus, to the extent the underlying common law cause of action involves intentional or reckless acts, the pre-dispute limitations should be unenforceable.

Further, the Restatement Second recognizes that, for harm caused by negligence, an exculpatory clause will not be enforced if the plaintiff is "a member of a protected class."[274] The term "protected class" includes, for example, the employment relationship, recognizing that an employer may not seek to exculpate itself from liability for negligent harm caused to an employee.[275] While corporate law theory does not place on managers any generalized duty to stakeholders as contracting parties, other constituencies that have been the subject of this Article—consumers, franchisees, and insured parties—are also arguably within the contemplation of that class of protected relationships under Restatement Second section 195(2).[276]

---

269. This Article also is not the first to advocate for legislative reform prohibiting such clauses. *See* Sternlight, *supra* note 116, at 121.

270. *See supra* notes 22, 136 and accompanying text.

271. *See* 57A AM. JUR. 2D *Negligence* § 55 (2004) ("Statutory liability for negligence cannot be contracted away . . . ."). As stated in American Jurisprudence, "if an injury results from a violation of a statute that establishes a certain standard of conduct for the protection and benefit of the members of a class, an immunity contract or clause exculpating a defendant from liability for negligence is unenforceable as contrary to public policy." *Id.* (footnotes omitted).

272. However, in places where the underlying statutory law allows a corporation to limits its own liability, the analogy to exculpatory clauses no longer holds. For example, the Uniform Commercial Code allows a seller to disclaim warranties. *See* U.C.C. § 2-316 (2007). Thus, in the narrow circumstance where the underlying substantive right involves a disclaimed warranty, the procedures to enforce that warranty might, likewise, be disclaimable. In other words, if an underlying statutory scheme allows contractual exculpation from substantive liability, it follows that it may also permit narrow procedural limitations to enforce that statute.

273. RESTATEMENT (SECOND) OF CONTRACTS § 195(1) (1981).

274. *Id.* at § 195 cmt. a.

275. *Id.*

276. *See* Yang v. Voyagaire Houseboats, Inc., 701 N.W.2d 783, 789 n.3, 790–91 (Minn. 2005) (en banc) (refusing to enforce an exculpatory clause in a houseboat rental agreement due

Thus, whether the underlying substantive law is a creature of statute or common law, a strong argument exists by analogy to prohibit pre-dispute limitations on procedural rights that could effectively exculpate corporations from liability. Such a rule is consistent with the Supreme Court jurisprudence describing arbitration as an alternative forum but not an avenue for weakening the deterrent and remedial functions of the underlying substantive law.[277] Again, rather than leaving the courts in a position of discerning this policy, which is likely to lead to a patchwork of irreconcilable results, Congress should act to amend the FAA to render these limitations unenforceable per se.[278]

Admittedly, it must be recognized that this solution likely suffers from political infeasibility. Corporations generally favor arbitration over civil litigation in handling disputes with stakeholder constituencies and have the lobbying power and influence to frustrate legislation that threatens their interests in arbitration. Likewise, arbitration is itself a big business and not without its own political influence.[279] Yet, hopeful progress has been made in some states; a handful of state legislatures have prohibited or required heightened scrutiny of collective action waivers.[280]

---

to nature of relationship between lessor and lessee and disparity in bargaining power); *cf.* Valley Nat'l Bank v. Nat'l Ass'n for Stock Car Auto Racing, Inc., 736 P.2d 1186, 1189 (Ariz. Ct. App. 1987) (enforcing spectator's release of liability of automobile race track because of absence of protecting relationship between the parties); Kellar v. Lloyd, 509 N.W2d 87, 93 (Wis. Ct. App. 1993) (enforcing racetrack volunteer's release of liability because volunteer was not an employee, and thus, not within protected class).

277. *See* Gilmer v. Interstate/Johnson Lane Corp., 500 U.S. 20, 27–28 (1991). This approach seems to garner the sentiment of Justice Steven's dissent in *Gilmer*, which expressed concern that "an essential purpose of the ADEA is frustrated by compulsory arbitration of employment discrimination claims." *Id.* at 42 (Stevens, J., dissenting).

278. *See supra* note 269 and accompanying text. Given this Article's definition of "stakeholder," this per se rule would apply to all standardized agreements, even those between two business entities. There are two potential responses to whether the prohibitions here proposed should be applied to business-to-business arbitration agreements. First, to the extent there are arguments that the rules applicable to business-to-business arbitration should be different than those applicable to consumers, employees, and franchisees, these arguments could be addressed to the default rules that apply to the parties' contract—i.e., as a baseline, how much discovery is available in arbitration—and not the problematic pre-dispute terms identified in this Article. For a discussion of the baselines for collective action, discovery, and limitations periods, see *infra* notes 292–99 and accompanying text. Second, anecdotally, these problematic clauses do not appear to arise with any frequency in business-to-business contracts and, thus, there is likely a compelling argument that the restrictions on these clauses need not apply in the business-to-business context. Moreover, it seems that empirical evidence now supports this anecdotal claim. *See* Eisenberg et al., *supra* note 17, at 15.

279. *See* Alderman, *supra* note 27, at 1256 ("The provision of arbitration services . . . is a competitive business involving large profits.").

280. *See* CONN. GEN. STAT. ANN. § 36a-746c(7) (West 2004) (prohibiting the inclusion of class action waivers in high cost home loans); GA. CODE ANN. § 16-17-2(c)(2)(C) (2007) (directing courts to consider waiver of class action rights in determining an arbitration agreements' unconscionability); N.M. STAT. §§ 44-7A-1(b)(4)(f), 44-7A-5 (prohibiting

At the same time, however, what has occurred in Utah is disquieting. In 2006, the Utah legislature enacted the first statute to validate collective action waivers in all types of consumer credit card and other loan agreements.[281] Interestingly, a press release from the law firm of Ballard Spahr Andrews & Ingersoll, LLP, boasts having "shepherded" the enactment of the Utah law, "which will help banks and finance companies defeat class actions filed against them throughout the country."[282] One of the firm's partners commented: "Given that Utah has dozens of large banks that extend consumer credit throughout the country, this is very significant legislation. This statute will serve as significant protection against unnecessary and unwarranted class action suits."[283] The press release notes that "[n]ot only will the statute apply to class actions brought in Utah, it should also apply to class actions filed elsewhere whenever a valid contractual choice of Utah law provision has been included in the agreement."[284]

Thus, the proposed legislative reform must recognize this race to the bottom. The simplest option is to enact reform on a federal level by amending the FAA.[285] The FAA preempts state laws concerning arbitration; thus, corporations would be unable to contract around the reach of the reforms.[286] If reforms occur on a state level, as has happened in Utah, aggressive use of choice of law clauses might allow one state's corporation-friendly laws to effectively deny access to justice for all contracting parties.[287] In essence, absent federal legislative reforms, corporations could use express procedural limitations in arbitration clauses to contract around substantive accountability and, in tandem, use a choice of law clause to write themselves into the law of a state that promises to enforce such limitations.

---

collective action waivers in adhesive consumer arbitration agreements); OKLA. STAT. ANN. tit. 12, § 1880 (West 2007) (providing that collective action waivers appearing in arbitration agreements within adhesion contracts are subject to heightened scrutiny in determining unconscionability).

281. UTAH CODE ANN. § 70C-3-104 (Supp. 2007) (enforcing conspicuous collective action waivers); Press Release, Ballard Spahr Andrews & Ingersoll, LLP, Ballard Attorneys Pilot Unprecedented Utah Law (Mar. 30, 2006), http://www.ballardspahr.com/press/press_detail.asp?ID=964 [hereinafter Press Release].

282. Press Release, *supra* note 281.

283. *Id.*

284. *Id.*

285. To the extent that a general choice of law clause could invoke state arbitration law and thereby opt out of the FAA, state reforms should be encouraged as well. *See supra* note 87.

286. *See* Federal Arbitration Act, 9 U.S.C. § 2 (2000). Another option, which would allow the reforms to occur on an incremental, state-by-state level, is for states to use "bomb shelter" provisions in their statutes to prohibit the enforcement of these problematic clauses against their citizens. *See* Jack M. Graves, *Party Autonomy in Choice of Commercial Law: The Failure of Revised U.C.C. § 1-301 and a Proposal for Broader Reform*, 36 SETON HALL L. REV. 59, 69 n.73 (2005) (discussing use of "bomb shelter" provisions in context of UCITA).

287. Notably, this also raises federalism concerns as it allows Utah to effectively trump the contrary policies of other states, such as Connecticut and Georgia.

The compelling breadth of scholarship recommending various reforms to the arbitration regime has largely gone unheeded.[288] If Congress is not prepared to ban pre-dispute arbitration clauses outright, it should at least act to prohibit these further procedural limitations contained within such clauses. This solution would not undermine the policy favoring arbitration. Instead, it would actually be consistent with the efficiency goals of arbitration by eliminating litigation over the enforceability of common arbitration terms.

Moreover, regulation of these problematic clauses is warranted. The contract of adhesion apologists argue that it is appropriate to simply leave equalizing effects to market forces. At least in the context of consumer contracts, scholars have argued that a corporation's reputational concerns will prevent it from acting opportunistically.[289] However, arbitration terms set up a sturdy "wall of silence" to protect corporate reputation. Indeed, one of the very purposes of diverting disputes to the arbitral forum is confidentiality.[290] It is true that, at much expense and with the burden of proof, a contracting stakeholder may take the dispute out of the private sphere and into court to challenge an overreaching term as unconscionable. However, with all the barriers of pre-dispute arbitration erected, it is doubtful that corporations' reputational concerns will reliably prevent them from using overreaching terms, especially when the doctrine of unconscionability does not reliably inform parties which terms are overreaching in this context.[291]

Finally, a per se rule against these procedural limitations begs the question of what the default rules are and what they should be. If a pre-dispute limit on process will not be enforced, what, then, are the parties left with? Regarding class action waivers, the common law default rule provides that courts may not order consolidation of arbitration unless the parties' express agreement allows for it.[292] However, the Revised Uniform Arbitration Act would reverse this default rule and allow courts to consider consolidation unless the parties'

---

288. *But see* Arbitration Fairness Act of 2007, S. 1782, 110th Cong. § 3 (1st Sess. 2007) (proposing to amend the Federal Arbitration Act to invalidate pre-dispute agreements to arbitrate franchise, consumer and employment disputes); Arbitration Fairness Act of 2007, H.R. 3010, 110th Cong. § 3 (1st Sess. 2007) (proposing identical amendments as S. 1782, *supra*).

289. Lucian A. Bebchuk & Richard A. Posner, *One-Sided Contracts in Competitive Consumer Markets*, 104 MICH. L. REV. 827, 830 (2006) ("[A] rule of unconscionability that condemned one-sided terms would systematically favor opportunistic buyers without protecting fair buyers, because the latter are protected by the sellers' investment in reputation.").

290. David P. Pierce, *The Federal Arbitration Act: Conflicting Interpretations of Its Scope*, 61 U. CIN. L. REV. 623, 625 (1992).

291. *See* Todd D. Rakoff, *The Law and Sociology of Boilerplate*, 104 MICH. L. REV. 1235, 1236 (2006) ("Without belaboring the issue, Bebchuk and Posner seem to me to do nothing to show that this combination of judicial enforcement and the reputational concerns of firms will produce systematically desirable results.").

292. *See* Alan Scott Rau, *Federal Common Law and Arbitral Power*, 8 NEV. L.J. 169, 174–75 & n.16 (2007) (citing ALAN SCOTT RAU ET AL., PROCESSES OF DISPUTE RESOLUTION: THE ROLE OF LAWYERS 897–901 (4th ed. 2006)).

agreement explicitly forecloses the possibility.[293] For the reasons asserted in this Article, unless the parties agree *after the dispute has arisen* to foreclose aggregation of claims, the mechanism of collective action should be unequivocally permitted. Thus, to the extent that the existing common law default rules would not accomplish this end, the federal legislative reforms must, likewise, establish a rule that unequivocally allows collective action.

The same issues arise with regard to discovery. If the parties cannot limit discovery, the question becomes what the baseline of discovery should be. The permissive relevance standard of the Federal Rules of Civil Procedure[294] arguably defeats the efficiency of arbitration, which is ordinarily marked by less discovery. At the same time, an arbitration agreement providing for a heightened standard of "substantial need"[295] perhaps sets too high a burden on the party seeking discovery and presents the problems outlined in this Article. For this reason, consistent with many institutional rules concerning discovery,[296] the default should furnish the arbitrator with discretion to order discovery.[297] The parties should not be permitted to contractually limit the arbitrator's discretion by pre-dispute agreement. Using this discretion, the arbitrator should balance the parties' need to obtain the information required to establish a claim or defense against the efficiency goals of arbitration.[298]

The minimum limitations period is an easier question because it is set by the laws applicable to the substantive claim. It should be noted, however, that if a particular substantive statute expressly allows the parties to contractually reduce the limitations period, that statute should trump this Article's proposed legislative reform.[299] In this regard, the treatment of a pre-dispute reduction of

---

293. *Id.*; REVISED UNIF. ARBITRATION ACT § 10(c) (2000) ("The court may not order consolidation of the claims of a party to an agreement to arbitrate if the agreement prohibits consolidation.").

294. FED. R. CIV. P. 26(b)(1). The Federal Rules of Civil Procedure require that the evidence sought is relevant to a claim or defense in the action:

> Parties may obtain discovery regarding any matter, not privileged, that is relevant to the claim or defense of any party, including the existence, description, nature, custody, condition, and location of any books, documents, or other tangible things and the identity and location of persons having knowledge of any discoverable matter.

Id.

295. *See* Martinez v. Master Prot. Corp., 12 Cal. Rptr. 3d 663, 672 (Ct. App. 2004) (discussed *supra* Part II.B).

296. *See supra* note 187 and accompanying text.

297. Certainly, limited discovery is part of arbitration. Perhaps the per se rule against discovery limitations or the minimum quantum of available discovery should be limited or amended to cater to certain categories of arbitration—for example, business-to-business arbitration.

298. *See* REVISED UNIF. ARBITRATION ACT § 17(c) (2000) ("An arbitrator may permit such discovery as the arbitrator decides is appropriate in the circumstances, taking into account the needs of the parties to the arbitration proceeding and other affected persons and the desirability of making the proceeding fair, expeditious, and cost effective.").

299. For example, U.C.C. § 2-725(1) (2007) allows parties, in their "original agreement,"

the applicable statute of limitations would be consistent in both litigation and arbitration.

CONCLUSION

Tellingly, the law first met reluctantly with corporations and with arbitration. In the words of Justice Brandeis, early restrictions on corporate activity and financing were born of "a sense of some insidious menace inherent in large aggregations of capital, particularly when held by corporations."[300] These restrictions were eventually liberalized to the current system of corporation statutes, which serve largely as enabling laws.[301] Likewise, with arbitration, the common law was initially hostile to the notion that private contractual arrangements could "oust the courts of the jurisdiction conferred by law."[302] The FAA has reversed this hostility and the Supreme Court precedent has embraced pre-dispute arbitration with open arms.[303] The Supreme Court's stated policy favoring arbitration, as well as the increased popularity of arbitration, have in turn fostered the increased use of express terms limiting certain procedural rights in arbitration. If Congress is not prepared to prohibit pre-dispute arbitration clauses generally, it should at least address certain additional, specific express procedural limitations contained within such clauses in standardized forms.

The examples discussed in this Article—collective action waivers, limits on discovery, and shortened statute of limitations—may serve to weaken the enforcement of underlying substantive laws and, in effect, allow corporations to evade accountability to those stakeholders with whom they have contractual relationships. The doctrine of unconscionability, which has been used to police these limitations, does not provide clear guidelines. Rather than defer to the law of extant contract defenses, Congress should articulate clearer standards for the enforcement of these terms. This articulation will only serve to further the aims of arbitration: expediency and simplicity. Moreover, it will ensure that corporations are prevented from using standardized form agreements with stakeholder constituencies to contract out of process, and with that, contract away accountability for statutory and common law violations.

---

to "reduce the period of limitation to not less than one year," except in consumer contracts. This UCC provision would not be trumped by the reforms proposed in this Article.

300. Louis K. Liggett Co. v. Lee, 288 U.S. 517, 549 (1933) (Brandeis, J., dissenting in part) (addressing the development of laws concerning corporations).

301. *See* JAMES D. COX & THOMAS LEE HAZEN, COX & HAZEN ON CORPORATIONS § 2.06, at 92 (2d ed. 2003) (discussing how modern corporation statutes became "'enabling,' 'permissive' and 'liberal'" as a result of reduced restrictions on corporations).

302. Taylor & Cliffe, *supra* note 16, at 1092–94 (quoting Home Ins. Co. v. Morse, 87 U.S. (1 Wall.) 445, 451 (1874)).

303. *See* cases cited *supra* note 12.

ALL SPRAWLED OUT: HOW THE FEDERAL REGULATORY SYSTEM HAS DRIVEN UNSUSTAINABLE GROWTH

CHAD D. EMERSON*

---

* Associate Professor of Law, Thomas Goode Jones School of Law, Faulkner University. Professor Emerson is extremely grateful to Professor George Kuney for his wise counsel and editorial suggestions during the development of this article. He would also like to thank the SEALS Young Scholars Program where this topic was originally presented and vetted by various colleagues. Finally, Professor Emerson extends his thanks to Davy Hay for his excellent research assistance on this critical topic.

I. INTRODUCTION

When it comes to land planning and development, the United States faces a serious threat that grows more troublesome every year—one whose negative effects run the gamut from environmental concerns to social and fiscal harms.[1] This threat, often called sprawl, is evidenced by the proliferation of unsustainable land development patterns throughout the country.[2]

---

1. While the harmful effects of sprawl on society are noteworthy, this article focuses more on how specific federal regulatory policies have induced sprawl. This focus is, in part, because other articles have already effectively discussed the harmful effects of sprawl. *See, e.g.*, Robert H. Freilich & Bruce G. Peshoff, *The Social Costs of Sprawl*, 29 URB. LAW. 183, 189–93 (1997) (discussing how sprawl has generated negative community, housing, employment, and political effects); Chad Lamer, *Why Government Policies Encourage Urban Sprawl and the Alternatives Offered by New Urbanism*, 13 KAN. J.L. & PUB. POL'Y 391, 399–401 (2004) (considering the effects of sprawl on economic, environmental, health, and general quality of life issues); *see also* Michael E. Lewyn, Policy Review, *The Urban Crisis: Made in Washington*, 4 J.L. & POL'Y 513, 518–26 (1996) (discussing the impact of sprawl on city residents and suburbanites).

2. There are many definitions for sprawl, though most have in common the idea that sprawl is defined by unsustainable land development practices on a community's fringe. Testifying before Congress on this issue, Richard Moe, President of the National Trust for Historic Preservation, offered one particularly concise definition: "the poorly planned, low-density, auto-oriented development that spreads out from the edges of communities." *Hearing on Community Growth and Environmental Quality Before the S. Comm. on Env't & Pub. Works*, 107th Cong. (1999) (statement of Richard Moe, President, Nat'l Trust for Historic Pres.) [hereinafter *Hearing on Community Growth*], *available at* http://epw.senate.gov/107th/moe_3-17.htm.

Significantly, sprawl is not simply a problem of bad design or planning. These are merely symptoms of a more profound cause.[3] The true driving force behind sprawl is a series of federal laws and regulations that, over the last century, have facilitated development patterns in the United States that are neither fiscally sound nor physically sustainable.[4]

This Article examines three specific areas of federal regulation that have exacerbated sprawl: tax policy, transportation policy, and housing policy. The Article surveys specific examples of federal laws within each of these three areas that have promoted the near-unfettered growth of American sprawl. The laws and regulations are analyzed within a historical context to determine why and how they came to be. Thereby, areas of federal regulation are identified that, if modified or repealed, would facilitate a move away from sprawl growth and toward a more sustainable land development strategy. Ultimately, this Article exposes the federal laws that have driven sprawl in this country and, thus, have intensified the numerous negative effects of sprawl on our society.[5]

II. WHAT IS SPRAWL?

Before analyzing the laws that created sprawl, one point warrants clarification: *Not all suburban growth constitutes sprawl.* Indeed, "[b]eing anti-sprawl is not being anti-growth. The question is not whether our communities should grow, but rather how they will grow."[6] The reality is that

---

3. *See* Angela Glover Blackwell, *It Takes a Region*, 31 FORDHAM URB. L.J. 1303, 1305 (2004) ("Sprawl and regional inequity are not natural results of a free market economy. Rather, they are direct results of public policies that have provided incentives for suburban growth at the expense of central cities and older suburbs and their low-income residents.").

4. *Id.* Federal laws and regulations are not the only causes of sprawl, but they are the leading ones. *See generally* Richard K. Green, *Nine Causes of Sprawl*, ILL. REAL ESTATE LETTER 1 (1999) (discussing other non-regulatory causes of sprawl). Additionally, state and local regulatory laws and policies can facilitate sprawl. *See, e.g.*, GERRIT KNAAP ET AL., GOVERNMENT POLICY AND URBAN SPRAWL 3–6 (2000), *available at* http://dnr.state.il.us/orep/c2000/balancedgrowth/pdfs/government.pdf (discussing how state transportation policy, infrastructure financing, and local government financing encourage sprawl).

5. While this Article primarily focuses on the worst offenders among federal regulations—federal tax, transportation, and housing policies—these areas are not the only ways that the federal government encourages sprawl. Other examples include the Department of Housing and Urban Development's grant program for communities to expand sewer and water infrastructure into the undeveloped periphery of the city. "Once the water, sewage, and utility lines are established outside of the city, housing and other developments follow. The HUD grants act as a subsidy for the construction of new housing developments that in turn contribute to sprawl." *See* Lamer, *supra* note 1, at 398. Other federal regulations that have promoted sprawl include the Superfund legislation, the Clean Air and Water Acts, and other national agricultural policies. *See* KNAAP ET AL., *supra* note 4, at 4.

6. *Hearing on Community Growth*, *supra* note 2.

suburban land development in the United States dates back to the country's origins.

A. The Origins of Suburban Growth

Land development outside the city center is hardly a new phenomenon. Indeed, dating back to early Babylon, the wealthy built retreats in the countryside.[7] This trend continued, with well-known examples including the early Italian city-states, London by the 1500s, and Paris in the 1600s.[8] As early as the eighteenth century, prominent individuals in the United States sought housing in places that were outside of, yet still accessible to, major cities such as Boston and Philadelphia.[9] Specific examples of this early suburban growth in the United States included Chestnut Hill near Philadelphia, Tuxedo Park near Manhattan, and Lake Forest near Chicago—each an exclusive enclave where only the wealthy could afford the time and expense of commuting back to the urban center.[10]

Yet life outside the city was often neither ideal nor idyllic. Indeed, many of those who lived on the city's periphery did so because they could not afford to live near the city center.[11] This condition dates back to ancient communities, where the poor often dwelled outside the protection of the city walls.[12] However, even in those instances, living on the periphery did not mean an isolated existence, as those outside the walls lived within reasonable proximity to the city.[13]

Dolores Hayden, a leading researcher on historical development patterns, coined the terms "borderlands" and "picturesque enclaves" for two early models of American suburban living.[14] With borderland developments beginning in the 1820s and picturesque enclaves in the 1850s, both early types of development demonstrated extra-urban growth well before the advent of the personal automobile.[15] While both of these early models developed outside of

---

7. HOWARD FRUMKIN, LAWRENCE FRANK & RICHARD JACKSON, URBAN SPRAWL AND PUBLIC HEALTH: DESIGNING, PLANNING, AND BUILDING FOR HEALTHY COMMUNITIES 27 (2004).

8. *Id.*

9. *See id.*

10. WITOLD RYBCZYNSKI, LAST HARVEST 87 (2007).

11. *See* Joel Schwartz, *Evolution of the Suburbs*, *in* SUBURBIA: THE AMERICAN DREAM AND DILEMMA 1, 2 (Philip C. Dolce ed., 1976).

12. ROBERT BRUEGMANN, SPRAWL: A COMPACT HISTORY 21 (2005) (discussing the area outside the city that "housed marginal social or political groups and families too poor to afford dwellings inside the walls").

13. *See id.* (describing the suburban "transitional zone" between the city walls and the rural farmland only "[a] few miles outside the walls").

14. DOLORES HAYDEN, BUILDING SUBURBIA: GREEN FIELDS AND URBAN GROWTH, 1820-2000, at 4 (2003).

15. *Id.* at 4; *see id.* at 21 (describing contemporary modes of commuting to the city). In

the traditional city limits, the primary difference between the two was that borderlands were generally more isolated, individual residences, while "[t]he designers of enclaves added a sense of community to the borderland goals of house and land, becoming the first to express [this] triple dream."[16]

These early suburban models resulted from a desire of the wealthy class to retreat from the crowded, and often dirty, streets of the city.[17] Yet during those times, what constituted a suburb was much different than the suburbs of today. Unlike modern times, most cities of that vintage were small and compact, with the distance from the center of the city to the edge generally no more than a mile or two.[18] This short distance was primarily due to the fact that transportation modes were limited to walking, sailboats, riding on horseback, and horse-drawn carriages, none of which made distant exurban growth a viable option for those commuting on a daily basis.[19]

The advent of the steamboat altered this equation and allowed regular commutes to the city from greater distances.[20] For example, the steamboat bolstered the growth of Brooklyn Heights, an early suburb that developed across New York Harbor from lower Manhattan.[21] Historian Kenneth Jackson opines that Brooklyn Heights was the nation's first commuter suburb.[22] It differed from other early suburbs because of "the number of commuters, the easy access to a large city, and the bucolic atmosphere" that defined the early Brooklyn Heights development.[23]

While Brooklyn Heights is evidence that suburban developments have existed since the early days of this country, suburbia's rise to prominence as the nation's dominant growth pattern did not occur until the 1970 census, which, for the first time ever, "declared America a suburban nation."[24] By this time,

---

addition to these two types of suburban living, Hayden identifies five additional suburban eras: the 1870s streetcar buildouts, the 1900s mail-order and self-built suburbs, the 1940s "sitcom" suburbs, the edge nodes of the 1960s, and the 1980s rural fringes. *Id.* at 4–5.

16. *Id.* at 45.

17. *Id.* at 21–22.

18. *Id.* at 21.

19. *Id.*

20. *See* KENNETH T. JACKSON, CRABGRASS FRONTIER: THE SUBURBANIZATION OF THE UNITED STATES 27 (1985) (noting the "phenomenal growth" of Brooklyn after regular steam ferry service to New York City began).

21. *Id.* at 25–27.

22. *Id.* at 25. Interestingly, though, while their works reach many similar conclusions, Dolores Hayden appears to disagree with Jackson on this point: "Although some historians have called Brooklyn Heights the first suburb, neither Brooklyn Heights nor the Boston projects provided models for the picturesque enclave. They are better understood as extensions of urban housing models from the affluent neighborhoods of Manhattan and Beacon Hill." HAYDEN, *supra* note 14, at 46.

23. JACKSON, *supra* note 20, at 25.

24. Schwartz, *supra* note 11, at 1.

suburban growth had become the most prolific type of commercial and residential development.[25]

Regrettably, this trend continues today as more residents locate in the suburbs than in the cities.[26] This is unfortunate because today's suburban growth differs from historical extra-urban growth in two important respects. First, early suburban development in the United States was the exception rather than the norm—often limited to the few who could afford it.[27] In considering America's early suburbs, one researcher has noted that "middle-income city dwellers could not afford this [suburban] living pattern because of the extra time and travel costs it demanded."[28] In the post-World War II era, though, relatively inexpensive suburban development has become the dominant growth pattern, enabling people of all socioeconomic profiles to reside outside the city.[29]

Second, early suburban growth, though located on the periphery, was developed in a much more sustainable way. Generally, the earliest suburban growth was fairly limited in scope and was not treated as a place from which workers would make a daily commute to the city center.[30] Rather, these early suburbs were often filled with second homes for the wealthy.[31] Then, as the nineteenth century closed, suburban growth began to develop primarily along transit corridors for trolleys and trains.[32]

A classic example of this growth pattern[33] was the Frederick Law Olmstead-planned village of Riverside, an early 1860s suburb of Chicago located at the final stop of the Burlington-Northern Railroad commuter train.[34] Identified as one of the earliest picturesque enclaves,[35] Riverside actually represented a historical version of today's increasingly popular "transit-oriented development," which includes a small commercial district centered on the rail

---

25. *See* HAYDEN, *supra* note 14, at 10; RYBCZYNSKI, *supra* note 10, at 86.

26. *See* HAYDEN, *supra* note 14, at 10 ("By 2000, more Americans lived in suburbs than in central cities and rural areas combined.").

27. Peter O. Muller, *Transportation and Urban Form: Stages in the Spatial Evolution of the American Metropolis*, *in* THE GEOGRAPHY OF URBAN TRANSPORTATION 59, 60 (Susan Hanson & Genevieve Giuliano eds., 3d ed. 2004).

28. *Id.*

29. *See id.*; David L. Callies & Glenn H. Sonoda, *Providing Infrastructure for Smart Growth: Land Development Conditions*, 43 IDAHO L. REV. 351, 352 (2007).

30. *See* RYBCZYNSKI, *supra* note 10, at 87.

31. *Id.*

32. Muller, *supra* note 27, at 60.

33. *See* RYBCZYNSKI, *supra* note 10, at 111 ("Nineteenth-century garden suburbs such as Riverside . . . were far from the city but were firmly tied to downtown by railroads and streetcars.").

34. Riverside Community Web Site, History, http://www.riverside-illinois.com/History.htm (last visited Apr. 4, 2008).

35. HAYDEN, *supra* note 14, at 45.

station.[36] While not as ideal as urban redevelopment, Riverside still provided a reasonably sustainable growth pattern by centering itself along a mass transit option and fostering development around the transit stop.[37]

Another well-known example of a streetcar suburb was Chevy Chase, Maryland.[38] Established in the 1890s by Nevada's two United States senators, Chevy Chase was accessible via a new electric streetcar line.[39] The developers wisely built the streetcar line themselves,[40] an adroit recognition that proximity to mass transit would create value for their project. Indeed, this type of development became so popular that by 1915, Los Angeles contained roughly 1200 miles of streetcar tracks.[41]

This organized approach to land development would change as federal and state laws began to cement massive, automobile-centric suburban development as the driving force in American land growth patterns.[42] While these laws were not always intended to promote unsustainable suburban growth, that was often their effect.[43]

Today's sprawling, low-density, auto-centric growth pattern continues to become more and more endemic. In fact, one commentator has identified a new phenomenon called "the new suburban poverty."[44] This phenomenon, in which middle-class residents can no longer afford their suburban, automobile-dominated lifestyle, "result[s] i[n] a historic milestone that has gone strangely ignored: For the first time ever, more poor Americans live in the suburbs than in all our cities combined."[45] If an individual looks past the unsupported rhetoric that single-use, low-density suburbs are simply a market-driven choice for many, one confronts a stark reality:

---

36. *See generally* Center for Transit Oriented Development, http://www.newurbanism.org/centerfortod.html (last visited Apr. 4, 2008) (introducing the Center for Transit-Oriented Development, its motivating factors, and its goals); Transit Oriented Development, http://www.transitorienteddevelopment.org/tod.html (last visited Apr. 4, 2008) (describing the components of transit oriented design and the trend toward such developments).

37. *See* RYBCZYNSKI, *supra* note 10, at 111.

38. HAYDEN, *supra* note 14, at 73.

39. *Id.*

40. *Id.*

41. *Id.* at 98. This is not to say that problems did not exist with streetcar suburbs. In particular, residents disliked the cumbersome and unattractive network of poles and wires required to power the streetcars. *Id.* at 76.

42. *See* Muller, *supra* note 27, at 60–61.

43. *See id.* Commentators have also analyzed how racial and socioeconomic factors contributed to sprawl-promoting federal policies and regulations. *See, e.g.*, Blackwell, *supra* note 3, at 1305–06 (proposing that the race-restricted federal housing regulations of the mid-twentieth century promoted sprawl by whites while concentrating minorities in the increasingly decayed city centers).

44. Eyal Press, *The New Suburban Poverty*, THE NATION, Apr. 23, 2007, *available at* http://www.thenation.com/doc/20070423/press.

45. *Id.*

> Stories of downward mobility in America's suburbs have not exactly cluttered the headlines over the past decade. Low-wage jobs, houses under foreclosure, families unable to afford food and medical care are not [the images typically evoked]. But venture beyond the city limits of any major metropolitan area today, and you will encounter these things, in forms less concentrated—and therefore less visible—than in the more blighted pockets of our cities perhaps, but with growing frequency all the same.[46]

Worse still, there is no indication that this teetering house of cards called suburban sprawl is likely to become more stable. Rather, this crisis—a term that is probably too weak, rather than too strong—may very well evolve into what one commentator has termed "the long emergency," which, even if only partly accurate, will still result in a dramatic re-ordering of the country's built environment.[47] For this reason, an important step in addressing the situation is to carefully consider the laws and policies fostering the problem.

B. Two Types of Sprawl

While sprawl is a complex problem, it generally can be reduced into two broad, historical categories: unsustainable residential growth and unsustainable commercial growth. These two divisions are the natural result of the single, separated-use zoning scheme that resulted from the decision in *Village of Euclid v. Ambler Realty Co.*[48]

In this famous 1926 decision, the United State Supreme Court essentially held that zoning systems requiring the separation of land uses—even if those land uses would otherwise be entirely compatible—do not constitute a regulatory taking or violate substantive due process, and therefore can be implemented by municipalities with near impunity.[49] This short-sighted approach was best evidenced by the Court's own language: "[I]t may thereby happen that not only offensive or dangerous industries will be excluded, but those which are neither offensive nor dangerous will share the same fate."[50]

The impact of this decision was swift and decisive, especially when coupled with the United States Department of Commerce's creation of model

---

46. *Id.*

47. *See generally* JAMES HOWARD KUNSTLER, THE LONG EMERGENCY: SURVIVING THE END OF OIL, CLIMATE CHANGE, AND OTHER CONVERGING CATASTROPHES OF THE TWENTY-FIRST CENTURY (2006) (examining the causes and results of unsustainable suburban sprawl).

48. 272 U.S. 365 (1926).

49. *Id.* at 386–88.

50. *Id.* at 388. Of course, the *Euclid* court's vision of what constituted compatible land uses was amazingly myopic, as evidenced by its conclusion that, rather than complementing detached residential units, "the apartment house is a mere parasite." *Id.* at 394. If multi-family units were parasitic, the idea of mixing residential uses with office or commercial ones must have been near apocalyptic for the *Euclid* majority.

zoning-enabling legislation in the 1920s, which also favored the separation of uses in land development.[51] Within years of these two seminal events, most states had adopted zoning-enabling legislation and many cities had, in turn, passed zoning ordinances that facilitated, if not absolutely required, the separation of virtually all land uses.[52] These types of policy decisions encouraged what one commentator has described as "the nation's mid-to-late twentieth century urban diaspora."[53]

1. Commercial Sprawl

Commercial sprawl began with the development of the retail sector. In the early twentieth century, most commercial development was located within a walkable area.[54] Commercial growth primarily occurred in the form of "business nodes"—compact commercial areas located near transit stops that "offer[ed] 'convenience' shops including drug stores, small groceries, and doctors' offices."[55] With the arrival of *Euclid* and zoning-enabling legislation, though, communities began to see the growth of commercial uses separated from both residential areas and fixed transit routes.

The business nodes of the past expanded into the so-called "miracle miles," large groupings of retailers located along major paved roads.[56] The miracle miles grew in response to the proliferation of motor vehicles in the *Euclid* era.[57] They were necessarily long, narrow strips of commercial development that catered almost exclusively to motor vehicles and provided very little, if any, access for pedestrians or mass transit. Because the miracle miles were auto-centric, they presumably could sell much more in volume, since the purchasers could transport their wares home inside their spacious vehicles rather than having to carry them on the streetcar.

---

51. *See* Lee R. Epstein, *Where Yards Are Wide: Have Land Use Planning and Law Gone Astray?*, 21 WM. & MARY ENVTL. L. & POL'Y REV. 345, 357–58, 357 n.50 (1997). The Standard State Zoning Enabling Act granted broad power to municipalities' legislative bodies so they could

> regulate and restrict the height, number of stories, and size of buildings and other structures, the percentage of lot that may be occupied, the size of yards, courts, and other open spaces, the density of population, and the location and use of buildings, structures, and land for trade, industry, residence, or other purposes.

Id. at 357 n.50 (quoting ADVISORY COMM. ON ZONING, U.S. DEP'T OF COMMERCE, A STANDARD STATE ZONING ENABLING ACT § 1 (2d ed. 1926)).

52. *See id.* at 357–58.

53. *Id.* at 355.

54. *See* Thomas W. Hanchett, *U.S. Tax Policy and the Shopping-Center Boom of the 1950s and 1960s*, 101 AM. HIST. REV. 1082, 1088–89 (1996).

55. *Id.*

56. *See id.* at 1089.

57. *See id.*

Even though they encompassed larger tracts of suburban land than ever before, the miracle miles still did not beat downtown retail cores in terms of scope and variety.[58] That would soon change, though, with the advent of "regional center" malls—massive commercial developments that could compete directly with downtowns in terms of both the number and the variety of retail options.[59] As one commentator has noted,

> [B]y combining the small local shopping center with at least one large branch department store, the regional center competed favorably with the downtown area, offering a full range of merchandise and services (everything for "one-stop shopping"), but with the added convenience of traffic-free access and ample free parking.[60]

Commentators generally point to J.C. Nichols's Country Club Plaza, built in 1922, as the first shopping center in the United States.[61] However, unlike most regional center malls, the Plaza incorporated both residential and retail units into its overall master plan.[62] This would prove to be the exception rather than the norm.[63] For example, when the Minneapolis-based Southdale Mall opened in 1956, it became the country's first indoor mall and did not include any type of residential units.[64] Nor did other 1950s-era regional center malls, such as Northgate Mall in Seattle[65] or the Detroit-area Northland Shopping Center.[66]

Originally, regional center malls in the United States were slow to develop—even with the increase in motor vehicle ownership—for three primary reasons. First, constructing the infrastructure and buildings for these retail centers cost significantly more than developing the same land for

---

58. *See* Meredith L. Clausen, *Northgate Regional Shopping Center—Paradigm From the Provinces*, 43 J. SOC'Y ARCHITECTURAL HISTORIANS 144, 146–47 (1984) (discussing the fact that regional shopping centers, as compared to smaller shopping centers, competed more favorably with downtown retail because of the "full range of merchandise and services" offered).

59. *See id.* at 147.

60. *Id.*

61. *See* Hanchett, *supra* note 54, at 1089.

62. *See id.*

63. *See* Malcolm Gladwell, *The Terrazzo Jungle*, NEW YORKER, Mar. 15, 2004, at 120, 120–22, *available at* http://www.newyorker.com/archive/2004/03/15/040315fa_fact1 (discussing the advent of the "introvert" shopping mall—the blueprint for virtually every mall in America today).

64. *Id.* at 120–22, 125.

65. *See generally* Clausen, *supra* note 58 (providing a thorough analysis of the origins of Northgate Regional Shopping Center).

66. *See generally* Greta Guest, *Golden Northland: Pioneering Shopping Mall Marks a Faded 50th*, DETROIT FREE PRESS, Mar. 22, 2004, at 1E (discussing the origins of Northland Shopping Center).

residential use.[67] Second, the increased initial cost resulted in a slower return on investments than what a developer would normally realize in a residential development.[68] Third, because of the slow return on investment, retail centers generally required the developer to treat the centers as a longer-term investment than they might have considered a typical residential project to be.[69]

As a result of these factors, large retail centers did not immediately thrive.[70] Not until Congress made a key change to U.S. tax law did regional center malls become the sprawl form of choice for many developers.[71]

2. Residential Sprawl

In addition to inducing commercial sprawl, federal tax, transportation, and housing regulations have also stimulated residential sprawl.

While not marking its absolute beginning, the 1929 stock market crash and the subsequent economic depression fostered an environment ripe for breeding residential sprawl.[72] In particular, the federal government's passage of laws and regulations designed to combat the Great Depression by putting Americans to work and protecting their homes actually facilitated a sprawl-conducive environment.[73] The regulations created a situation in which "public policy and public spending . . . played an important part in creating incentives for suburbanization and sprawl."[74]

These Depression-era laws represented a significant change, as they led the federal government into areas of land use and housing regulation that had traditionally been left to local and state governments. Specific examples of this expansion of federal power include the Department of Commerce's drafting of

---

67. *See* Hanchett, *supra* note 54, at 1089; *see also* Gladwell, *supra* note 63, at 125 (noting the prohibitive costs of large shopping centers in the early 1950s).

68. *See* Hanchett, *supra* note 54, at 1091; Gladwell, *supra* note 63, at 125.

69. *See* Hanchett, *supra* note 54, at 1089, 1091.

70. *See id.* at 1091 ("At the midpoint of the 1950s, fewer than two dozen regional shopping centers existed in all of the United States.").

71. *See infra* Part III.A–B. Of course, this is not to say that a roofed shopping center automatically evidences commercial sprawl. In fact, covered or enclosed commercial gatherings have existed for hundreds of years in places such as the Middle Eastern bazaars in Isfahan, Iran, *see generally* Mohammad Gharipour, Bazaar of Isfahan (Dec. 12, 2003), http://www.iranchamber.com/architecture/bazaar_of_isfahan1.php (providing a detailed history of the Bazaar of Isfahan), and Istanbul, Turkey, *see generally* ArchNet, Covered Bazaar, http://archnet.org/library/sites/one-site.tcl?site_id=7441 (last visited Jan. 29, 2008) (discussing the history of the covered Istanbul bazaar known as Kapaliçarsi). The distinction is that these examples were generally integrated into areas that included other land uses such as residential units—eliminating the requirement that almost all vendors and purchasers travel long distances to reach these establishments. *See* Gharipour, *supra*.

72. *See* Epstein, *supra* note 51, at 354–55.

73. *See id.*

74. *Id.*

a national model building code in 1922 and national model zoning codes in 1924 and 1928.[75] Through these efforts, "the federal government began to engage directly with housing as an important area of national policy."[76]

Significantly, the Federal Housing Administration (FHA) attempted to address the housing crisis by adopting policies designed to spur residential development.[77] As discussed in more detail in Part V, the FHA promulgated a series of regulations—and lobbied for a series of laws—that, when combined with the national model zoning act and its separation of land uses, facilitated an unprecedented wave of new residential construction that was separated from other, non-residential land uses.[78]

Unfortunately, little effort was made to determine whether the separated uses were compatible. This resulted in perfectly compatible uses (such as an architect living above her office or an accountant living behind his office) being prohibited in the same way as incompatible uses (such as workers living adjacent to noxious, heavy industry).[79] Because of the collision of these sprawl-friendly federal policies, promoting the renovation of existing residential areas or even the mixture of residential uses with compatible non-residential ones soon became financially impractical.[80]

The result, if not exactly anticipated, quickly became obvious: a systematic separation of land uses unlike previous development patterns in the United States or, for that matter, much of the rest of the world.[81] Because of this sprawling form of growth, "government[s] at every level struggled to provide adequate infrastructure—roads, water, sewer, schools, and parks—to accommodate the seemingly voracious demands of such rapid Greenfield development."[82] Quite clearly, the perfect environment for sprawl had been set.

III. THE FEDERAL TAX LAWS THAT FACILITATED SPRAWL

In terms of inducing sprawl, two federal tax policies have played a prominent role in facilitating this type of unsustainable growth: the federal accelerated depreciation deduction[83] and the federal mortgage interest

75. HAYDEN, *supra* note 14, at 121.

76. *Id.*

77. *See* Freilich & Peshoff, *supra* note 1, at 186.

78. *Cf.* JACKSON, *supra* note 20, at 6–7 (noting that typical Americans live in suburban areas that are far from their jobs).

79. *Cf., e.g.*, Village of Euclid v. Ambler Realty Co., 272 U.S. 365, 388 (1926) (predicting this outcome).

80. *See infra* Part V.

81. *See* JACKSON, *supra* note 20, at 6–10.

82. Callies & Sonoda, *supra* note 29, at 353.

83. *See infra* Part III.A–B.

deduction.[84] The first served to exacerbate commercial sprawl while the second increased residential sprawl.[85]

Although neither policy expressly advocated suburban sprawl, their effects directed such a pattern.[86] As one commentator has noted,

> The federal tax code, in all its complexity, is heavily tilted toward new development and the consumption of open space. The tax code has historically subsidized upper middle class homeownership in the suburbs. It needs to put at least as much emphasis on promoting opportunities for revitalization and stabilization of older communities. Federal tax policy needs to provide incentives—which are currently lacking—for middle-class and moderate-income households to become urban homeowners.[87]

This Part examines how Congress facilitated sprawl through these two key tax deductions.

A. The History of the Accelerated Depreciation Deduction

Congress's initial decision to permit businesses to deduct depreciation values from machinery and buildings, along with its later decision allowing businesses to accelerate the rate at which they utilized this deduction, served as one of the major drivers of commercial sprawl. Over the past fifty years, Congress has regularly expanded and contracted the scope of the accelerated depreciation deduction.[88] Yet through it all, commercial sprawl has thrived in a tax environment that has created incentives to construct new buildings designed for short shelf lives.[89]

1. The 1913 and 1916 Tax Laws

When Congress passed the federal income tax in 1913, the original legislation allowed businesses to deduct "a reasonable allowance for depreciation by use, wear and tear of property."[90] Congress's rationale was to encourage businesses to set aside money saved by the deduction for use in

---

84. *See infra* Part III.C–D. These two areas of tax policy strongly encouraged sprawl but were not the only federal tax regulations to do so. *See* Epstein, *supra* note 51, at 355; Hanchett, *supra* note 54, at 1093.

85. *See infra* Part III.B, III.D.

86. *See infra* Part III.B, III.D.

87. *Hearing on Community Growth*, *supra* note 2.

88. *See generally* Hanchett, *supra* note 54 (describing the history of the accelerated tax deduction).

89. *See id.* at 1083.

90. Revenue Act of 1913, Pub. L. No. 63-16, 38 Stat. 114, 172 (1913).

maintaining or replacing their factories' worn machinery.[91] Essentially, the deduction sought to advance a policy of long-term infrastructure planning by companies.[92]

In 1916, Congress revised its definition of the deduction to "[a] reasonable allowance for the exhaustion, wear and tear of property,"[93] but again failed to define even loose parameters of what constituted "reasonable" depreciation. Rather, Congress allowed businesses to use their own interpretation of the term.[94] Predictably, companies adopted very broad definitions of what constituted a "reasonable allowance" so they could take larger deductions.[95]

The result of this broad definition was that "[b]y 1931, the deductions taken for depreciation in America exceeded the total taxable net income of all corporations."[96] Worse still, the depreciation deduction did not actually require companies to save the deducted monies for future machinery and building maintenance.[97] As a result, the money that companies saved due to the depreciation deduction could be used for virtually any reason.[98] The depreciation deduction had expanded well beyond its anticipated scope.

2. The 1934 Tax Regulations

In response to the expansive definitions used by corporations as to what constituted a "reasonable allowance," the United States Treasury Department promulgated regulations that set a specific, uniform method for calculating depreciation expenses.[99] Known as the straight-line method, one of the key changes was that the regulation set the useful life of most buildings at forty years.[100] This regulation meant that each year a building owner could deduct one-fortieth of the original cost of the building for depreciation.[101] The immediate effect of this change was to significantly reduce the expansive amounts that corporations had been deducting from their taxes for their buildings' depreciation.[102]

---

91. Hanchett, *supra* note 54, at 1092.
92. *See id.*
93. Revenue Act of 1916, ch. 463, § 5(a), 39 Stat. 756, 759 (1916).
94. Hanchett, *supra* note 54, at 1092.
95. *Id.*
96. *Id.*
97. *Id.* at 1093.
98. *Id.*
99. *Id.* at 1092.
100. *Id.*
101. *Id.* at 1092–93.
102. *See id.*

3. The 1954 Tax Law

In the early 1950s, the country began to experience a mild recession following the post-World War II boom.[103] This downturn increased pressure on Congress to enact new laws that would encourage business growth.[104] One such law passed by Congress was the amendment to the Internal Revenue Code of 1954.[105]

The 1954 law reinvigorated the depreciation tax deduction by adopting two new methods for calculating depreciation: the double-declining balance method and the sum-of-the-years method.[106] Both methods allowed building owners to increase their depreciation deductions by "simply shift[ing] tax deductions toward the first years of a project's life, which enabled investors swiftly to reap the benefits."[107] By increasing the percentage of the original cost that was deducted, businesses could accelerate the amount that they deducted each year.[108] Thus, the concept of accelerated depreciation was born.[109]

4. The Tax Act of 1969

Accelerated depreciation quickly led businesses to increase their use of the depreciation deduction.[110] Once again, Congress and the Treasury Department found themselves in search of a way to balance the pro-business growth features of the deduction with its effect of reducing the overall federal income tax receipts.[111] In 1969, Congress made a measured move to reign in accelerated depreciation by reducing the 200% declining balance figure to 150% for new construction and requiring all purchasers of used buildings to revert to the straight-line method.[112] This policy decision provided new construction with a tax advantage greater than that afforded to renovated construction.[113]

---

103. *Id.* at 1093.

104. *Id.*

105. Pub. L. No. 83-591, 68A Stat. 3 (1954).

106. David W. Brazell, Lowell Dworin & Michael Walsh, *A History of Federal Tax Depreciation Policy* 12 (U.S. Treas. Dep't, Office of Tax Analysis, OTA Paper 64, May 1989), *available at* http://www.treasury.gov/offices/tax-policy/library/ota64.pdf.

107. Hanchett, *supra* note 54, at 1094.

108. *See id.*

109. *See id.* at 1095.

110. *Id.*

111. *Id.* at 1102–03.

112. *Id.* at 1105.

113. *Id.*

5. The 1981 Tax Law

While the 1969 change did provide some control over accelerated depreciation, the country soon faced another recession—and with it, renewed efforts to pass legislation that would spur business growth.[114] Congress decided in 1981 to replace all existing methods for calculating depreciation with a new formula, the Accelerated Cost Recovery System (ACRS).[115] A major component of this new system reduced a building's useful life from forty years to fifteen years.[116] By compacting the useful-life period, Congress further decreased the depreciation time frame.[117] Thus, "a developer could [actually] deduct an astounding 31 percent of a building's cost as depreciation during its first three years."[118] Ultimately, the ACRS became one of the most significant drivers of commercial sprawl.

B. How the Accelerated Depreciation Deduction Has Induced Sprawl

Following the adoption of the accelerated depreciation deduction in 1954, real estate developers quickly recognized its business benefits.[119] Within five years, 97.9% of all real estate partnerships had switched to this method of calculating depreciation.[120]

As an immediate result, the accelerated depreciation deduction provoked a large wave of new suburban shopping centers.[121] For example, the total square footage of seventeen new regional center malls opened in the fourth quarter of 1956 exceeded the total square footage of all regional center malls opened prior to 1956.[122] Moreover, "construction shot up . . . from an average of 6 million square feet yearly in the early 1950s, to an average of 30 million each year beginning in 1956."[123]

Interestingly, some evidence indicates that Congress intended accelerated depreciation to apply only to machinery and factories, not to built structures.[124] Nevertheless, as drafted, the depreciation legislation ended up applying to most

114. *See* Brazell, *supra* note 106, at 20.

115. Hanchett, *supra* note 54, at 1105–06; Beth B. Kern, *The Role of Depreciation and the Investment Tax Credit in Tax Policy and Their Influence on Financial Reporting During the 20th Century*, 27 ACCT. HISTORIANS J. 145, 157 (2000).

116. Kern, *supra* note 115, at 157.

117. *See id.*

118. Hanchett, *supra* note 54, at 1106.

119. *Id.* at 1095.

120. *Id.*

121. *Id.* at 1097.

122. *Id.*

123. *Id.* at 1098.

124. *See id.* at 1094–95.

new commercial buildings.[125] This limitation created an incentive to construct new buildings instead of renovating existing ones: "Investors seeking the best return on their dollars now looked away from established downtowns, where vacant land was scarce and new construction difficult. Instead, they rushed to put their money into projects at the suburban fringe—especially into shopping centers."[126]

Not surprisingly, the rapidly increasing use of the accelerated depreciation deduction continued through the 1960s.[127] By 1967, overall accelerated depreciation deductions for buildings reached $750 million.[128] Even more amazingly, by 1970, "this single tax expenditure . . . equal[ed] fully one-fourth of the federal annual budget deficit."[129]

This trend led to a dramatic shift in the geographic location of new shopping centers.[130] As one commentator noted, "[t]hrough the mid-1950s, developers had sought locations *within* growing suburban areas. Now [post-1950s] shopping centers began appearing in the cornfields *beyond* the edge of existing development."[131] Federal tax regulations had fundamentally altered the location and scope of commercial development in the United States.[132] They directed development away from the mixed-use buildings in existing city centers and toward sprawling, single-use structures on the undeveloped fringe.[133]

C. The History of the Home Mortgage Interest Deduction

When Congress passed the original 1913 federal income tax, it provided a tax deduction for interest paid by consumers.[134] This deduction included interest paid on home mortgages.[135] Significantly, though, Congress does not appear to have specifically intended the deduction to induce home ownership.[136] Congress "certainly wasn't thinking of the interest deduction as a

125. *Id.* at 1095, 1097.
126. *Id.* at 1097.
127. *See id.* at 1102–03.
128. *Id.* at 1103.
129. *Id.*
130. *See id.* at 1098.
131. *Id.*
132. *See id.* at 1106–07.
133. *See id.*
134. Roberta F. Mann, *The (Not So) Little House on the Prairie: The Hidden Costs of the Home Mortgage Interest Deduction*, 32 ARIZ. ST. L.J. 1347, 1351–52 (2000). Congress passed the first income tax package in 1894, but the Supreme Court held that the income tax was unconstitutional. *See* Roger Lowenstein, *Who Needs the Mortgage-Interest Deduction?*, N.Y. TIMES, March 5, 2006 (Magazine), at 78.
135. Mann, *supra* note 134, at 1351–52; *accord* 26 U.S.C.A. § 163(h) (Supp. 2007).
136. Mann, *supra* note 134, at 1352; Lowenstein, *supra* note 134, at 78.

stepping-stone to middle-class homeownership, because the tax excluded the first $3,000 . . . of income," and only one percent of the population of the time earned more than that amount.[137] Even more persuasively, most homeowners of that era (except farmers) did not have home mortgages.[138] Indeed, rather than meaning to subsidize homeownership, Congress probably designed the interest deduction to aid business interests.[139] In particular, Congress sought to facilitate small business growth by allowing the deduction of all interest payments from proprietors' taxable income.[140] Yet even if Congress's intent was not to encourage people to purchase homes, the effect was a homeownership subsidy.[141]

Over time, Congress became concerned that the interest payment deduction "'provided an incentive to invest in consumer durables rather than assets which produce taxable income and, therefore, [was] an incentive to consume rather than [to] save.'"[142] Therefore, Congress eliminated nearly all personal interest deductions, with the major exception of the mortgage deduction, in passing the Tax Reform Act of 1986.[143]

By allowing homeowners to continue to deduct most mortgage-related interest payments but not allowing renters to deduct a proportional amount of rent payments, Congress obviously continued to implicitly favor homeownership over renting.[144] Politically, this was hardly unexpected.[145] In debating the Tax Reform Act of 1986, Senator Phil Gramm succinctly summarized the reality of the mortgage interest deduction:

> There is no basic principle in tax law that is more supported by the American people than the principle that you ought to be able to deduct interest on your home from your taxes. We have taken a position that home ownership is

137. Lowenstein, *supra* note 134.

138. *Id.*

139. *Id.*

140. *Id.*

141. Mann, *supra* note 134, at 1352–53 (explaining that even if not specifically so intended, the "likely effect of the interest deduction . . . was to encourage home ownership, at least for those taxpayers who benefit from the deduction").

142. Dean M. Maki, *Household Debt and the Tax Reform Act of 1986*, 91 AM. ECON. REV. 305, 305 (2001) (quoting STAFF OF JOINT COMM. ON TAXATION, 99th CONG., GENERAL EXPLANATION OF THE TAX REFORM ACT OF 1986, at 263 (Joint Comm. Print 1987)).

143. John Y. Taggart, *Denial of the Personal Interest Deduction*, 41 TAX LAW. 195, 222 (1988); *accord* Tax Reform Act of 1986, Pub. L. No. 99-514, 100 Stat. 2085 (1986).

144. Joseph A. Snoe, *My Home, My Debt: Remodeling the Home Mortgage Interest Deduction*, 80 KY. L.J. 431, 432 (1992).

145. *See id.* at 433 ("The preferential treatment afforded home mortgage interest partially reflects political reality: [T]he average taxpayer has become accustomed to deducting mortgage interest and likely would be outraged if Congress eliminated the deduction.").

> something that we want to promote, that that is an objective of our tax policy that is strongly supported, and it is reflected in this bill.[146]

In light of this attitude, Congress's retention of the home mortgage deduction is unsurprising.[147] Congress ultimately made only minor changes to the mortgage interest deduction, such as limiting how many homes a taxpayer could include in the deduction.[148] In the end, Congress continued to adhere to the policy of promoting new homeownership through the mortgage interest deduction. After all, is "an elected official really going to risk fooling with the mortgage deduction?"[149]

D. How the Mortgage Interest Deduction Has Induced Sprawl

The financial impact of the mortgage interest deduction can be measured in the billions.[150] By allowing deductions for home mortgage interest and property taxes, the federal tax code significantly lowers homeowners' tax liability.[151] Specifically, federal tax subsidies in favor of homeowners exceed $40 billion annually, amounting to roughly $2,800 per year for each homeowner with a mortgage.[152] These tax subsidies prioritize homeownership over renting[153] because the same deductions are not available to renters.

For a taxpayer to fully realize the benefits of the mortgage interest deduction, the taxpayer is, at least implicitly, given incentive to purchase new construction. "To take full advantage of the deduction, higher incomes require higher home mortgages (and higher housing costs). This deduction encourages sprawl by providing the means to protect *more income* by buying *more home*."[154]

Most new home construction occurs on the suburban and exurban fringes of a city, where land is cheapest.[155] Infrastructure, in turn, must be extended to service the new development.[156] As a result, the individual homeowner realizes

---

146. *Id.* (quoting 132 CONG. REC. S7387 (daily ed. June 12, 1986) (statement of Sen. Gramm)).

147. *See id.* at 432–33.

148. Taggart, *supra* note 143, at 214–15.

149. Lowenstein, *supra* note 134.

150. Mann, *supra* note 134, at 1353 ("Over the next five years, the home mortgage interest deduction alone is estimated to cost the United States government $262.6 billion in lost tax revenue, steadily increasing from $48.5 billion in 1999 to $56.8 billion in 2003.").

151. Freilich & Peshoff, *supra* note 1, at 187.

152. James A. Kushner, *Brownfield Redevelopment Strategies in the United States*, 22 GA. ST. U. L. REV. 857, 861 (2006).

153. *See* Mann, *supra* note 134, at 1352.

154. Freilich & Peshoff, *supra* note 1, at 187.

155. *See* KNAAP ET AL., *supra* note 4, at 3, 5–6.

156. *See id.* at 6.

an increased tax benefit because the more expensive home generates more interest to deduct—but the municipality actually incurs greater costs because of the new schools, fire stations, roads, and the like that must be built in order to service this growth.[157]

While the deduction has nevertheless been justified on grounds that it promotes homeownership, in reality, the deduction is not widely available to most potential homeowners.[158] In fact, it is only available to taxpayers who itemize their deductions.[159] Research indicates that the majority of homeowners who benefit from the mortgage interest deduction have incomes of at least $100,000—a tax bracket including only eight percent of homeowners.[160] Significantly more wealthy than unwealthy taxpayers are able to take advantage of the deduction.[161] This is important because the wealthy are the very homeowners who are most likely to be able to afford large, new homes on the suburban fringes of the city even without the deduction.

An important step toward halting the proliferation of sprawl would be for Congress to rework or even eliminate the mortgage interest deduction. Historical data indicate that eliminating the deduction would not negatively affect homeownership.[162] Repeal of the deduction would actually "facilitate the entry of lower income buyers into the housing market and improve the distribution of home ownership."[163] Ultimately, "eliminating the home mortgage interest deduction would reduce urban sprawl by reducing incentives to develop rural land" without harming the national housing market.[164]

IV. THE FEDERAL TRANSPORTATION LAWS THAT FACILITATED SPRAWL

Similarly to the tax laws that induced sprawl, federal transportation policy has also instigated unsustainable growth patterns.[165] Indeed, federal transportation laws have been one of the largest regulatory drivers of sprawl.[166]

---

157. *See id.* at 5–6.

158. *Id.* at 1359.

159. *Id.*

160. *Id.* at 1360.

161. *See id.* at 1365. Not only does the mortgage interest deduction favor the wealthy, it also "systematically disfavors the financial interests of minorities." *Id.*

162. Mann, *supra* note 134, at 1391.

163. *Id.*

164. *Id.* at 1391–92.

165. *See* Oliver A. Pollard, III, *Smart Growth and Sustainable Transportation: Can We Get There from Here?*, 29 FORDHAM URB. L.J. 1529, 1532 (2002) ("Federal, state, and local transportation policies have fueled auto-dependence and sprawl.").

166. *See id.* at 1532–35 (providing an overview of some of the prominent legislation and public policies that promote motor vehicle use and, thus, sprawl). Pollard also outlines several of the leading harms caused by unsustainable transportation policies, specifically environmental and economic harms resulting from the sprawl-conducive regulations. *See id.* at 1538.

The federal government has cemented a system of sprawl through its road-building legislation and federal gas tax regulations.

A. How Federal Road Building Laws Have Induced Sprawl

Unlike fixed rail transportation, non-fixed rail transportation—primarily motor vehicles—provides a traveler with much more flexibility.[167] When traveling by train, travelers are subject to the train's schedule and fixed destinations, whereas travelers by motor vehicle can basically elect to travel whenever and wherever they choose.[168] As motor vehicles became more widely available in the early twentieth century, the federal government unsurprisingly supported this enhanced flexibility by beginning to enact laws and regulations that promoted, and in some cases even subsidized, a massive expansion of non-fixed rail infrastructure in the United States.

One of the biggest elements of the non-fixed rail infrastructure is the federal road system, which is composed of the federal interstate system and other federal highways.[169] Each year, the federal government spends tens of billions of dollars on road construction in this system.[170] Yet road building has not always been a federal function.

In fact, the debate about whether the federal government should expend monies for road-building involved some of the nation's most influential leaders, including Thomas Jefferson and James Madison.[171] For example, as his final presidential act, Madison vetoed a bill that would have essentially empowered federal road-building:

> I am not unaware of the great importance of roads . . . and that a power in the National Legislature to provide for them might be exercised with signal advantage to the general prosperity. But seeing that such a power is not expressly given by the Constitution, and believing that it can not [sic] be deduced from any part of it without an inadmissible latitude of construction

---

167. *See* JAMES A. KUSHNER, THE POST-AUTOMOBILE CITY: LEGAL MECHANISMS TO ESTABLISH THE PEDESTRIAN-FRIENDLY CITY 31–35 (2004) (discussing the advantages of automobile travel).

168. *See id.* The United States government estimates that there are 3.9 million miles of public roads in the country but only 120,000 miles of "major railroads." *See* Nationalatlas.gov, Transportation of the United States, http://www.nationalatlas.gov/transportation.html (last visited Apr. 4, 2008); *see also* The Public Purpose, Transport Fact Book: Road & Rail Mileage in US Urbanized Areas, http://www.publicpurpose.com/tfb-usuaroadrail.htm (last visited Apr. 4, 2008) (comparing road mileage and rail mileage in select urban areas of the United States).

169. *See* Federal Highway Admininstration, Who We Are, http://www.fhwa.dot.gov/whoweare/whoweare.htm (last visited Apr. 4, 2008).

170. *See id.*

171. *See* John Lambert, *Control, Supervision & Management Issues*, 12541 NAT'L BUS. INST. CONTINUING LEGAL EDUC. 97, 100–02 (2004), *available at* 12541 NBI-CLE 97 (Westlaw).

> and a reliance on insufficient precedents; . . . I have no option but to withhold my signature from it . . . .[172]

During the 1800s and early 1900s, local governments remained primarily responsible for the construction of vehicular thoroughfares.[173] When considering adoption of the nation's first highway bill in 1916, Congress noted that "[p]rimarily[,] roads are local concerns and jurisdiction over them belongs to the States and local authorities. This jurisdiction should never be disturbed by the General Government."[174]

Despite the early opposition to federal funding, the prevailing view soon became that Congress did indeed have the power to fund road-building.[175] As manufacturers began to mass-produce motor vehicles in the early 1900s, there was a gradual shift from fixed rail as the primary form of transportation toward a greater emphasis on motor vehicles.[176] With this shift, a need developed for a more extensive thoroughfare network that could handle the increased traffic.[177] The federal government thus created "subsidies mak[ing] it cheaper for people to live further from where they work[ed], shop[ped], and engage[d] in other activities, which spur[red] development on the fringes of existing communities and necessitate[d] increased driving distances and frequency."[178] One commentator has noted the staggering extent of these subsidies: "$257 billion annually in tax subsidies for automobile use and fuel, an average of $2,000 for each taxpayer."[179] The following sections demonstrate how we reached this point by analyzing several concrete examples of federal regulations that have contributed to the proliferation of American sprawl.

1. The Federal-Aid Road Act of 1916

As arguably the first federal transportation law to adopt an anti-city slant, the Federal-Aid Road Act of 1916[180] played an important role in the facilitation

---

172. *Id.* at 102.

173. *Cf. id.* at 99–104 (discussing the early nineteenth century debate over whether the states had granted, or could grant, any road-building power to the federal government).

174. H.R. REP. NO. 64-26, at 4 (1916).

175. Lambert, *supra* note 171, at 102–05 (describing the first effort to build a national road and stating that "[w]hile there are those who will still argue that 'strict construction' does not allow the federal government to build a general national road system, the serious debate was lost long ago both politically and in the courts").

176. Laurence Gerckens, *Ten Events That Shaped the 20th Century American City*, PLAN. COMMISSIONERS J., Spring 1998, at 12, 15.

177. *See* Pollard, *supra* note 165, at 1532–33.

178. *Id.* at 1533.

179. Kushner, *supra* note 152, at 861.

180. Pub. L. No. 64-156, 39 Stat. 355 (1916).

of sprawl growth patterns in this country.[181] Resulting from a series of policy resolutions that originated with the first American Road Congress, the 1916 Act incorporated several provisions that laid the groundwork for unsustainable growth patterns.[182]

The 1916 Act used a decidedly anti-urban funding formula, including limiting federal aid to $10,000 per mile.[183] While this provision may not appear anti-urban on its face, its effect was to enable more rural road expenditures, since land was cheaper in rural areas than urban ones.[184] Similarly, the Act only permitted federal road aid grants in towns with populations greater than 2,500 residents if the average distance between houses was more than 200 feet.[185] This limitation essentially prohibited road monies from being spent in densely populated urban areas.[186]

Furthermore, the Act used a funding formula that determined how much money should be allocated to an area by analyzing its population, overall area, and existing road mileage.[187] The theory behind the formula would seem to be that areas with more roads would need more money for maintenance. Inexplicably, however, urban streets were entirely excluded when making this calculation.[188] Urban areas thus received significantly less funding despite having more roads used by more citizens.[189]

Each of these provisions of the 1916 Act set the tone for decades of federal spending on rural highways, which facilitated massive suburban and exurban growth while simultaneously providing only minimal funding to urban street networks and alternative forms of transit.

2. The Federal-Aid Highway Act of 1921

Five years later, Congress revisited the issue of federal highway funding when it passed the Federal Highway Act of 1921.[190] The 1921 Act classified "federal-aid highways" as either "primary" interstate routes or "secondary"

---

181. *See* OWEN D. GUTFREUND, TWENTIETH-CENTURY SPRAWL: HIGHWAYS AND THE RESHAPING OF THE AMERICAN LANDSCAPE 22 (2004).

182. *See id.* at 17, 22.

183. *Id.* at 22.

184. *Id.* (stating that settled urban areas had "higher land values" and "more demanding traffic flows").

185. *Id.*

186. *See id.*

187. *Id.*

188. *Id.*

189. *See id.* ("[T]he formula effectively stipulated a strong funding bias in favor of sparsely settled states, at the expense of those states that were more urbanized.").

190. Pub. L. No. 67-87, 42 Stat. 212 (1923).

intercounty routes.[191] The Act apportioned sixty percent of federal funding to primary routes and forty percent to secondary routes.[192]

Significantly, the Act failed to allocate any of the funding to intra-urban, or city, streets.[193] The urban inequity of this approach has been characterized as "a system of transfer payments, from urbanized regions to rural regions, and from all taxpayers to those who drove automobiles."[194] In effect, the benefits of the extensive federal road funds were realized by only a small segment of the overall population.[195] This disparity was evidenced by the fact that, as of 1921, "users of the [nine] million motor vehicles in the nation paid only twelve percent of *all* highway costs."[196]

Thus, the 1921 Act, coupled with the 1916 Act, further entrenched a federal road policy that prioritized rural roads over urban ones, often to the point of excluding urban streets from all federal funding. "The original 1916 and 1921 highway legislation [was] explicit in rendering all roads in urban areas ineligible for federal aid."[197] As the federal government continued to devalue urban areas, sprawl flourished into the suburban and exurban areas.

3. The Federal-Aid Highway Acts of 1934 and 1944

The anti-urban approach toward federal road funding continued into the 1930s and 1940s, albeit differently than in past decades. While the 1916 and 1921 Acts almost completely barred urban road funding, the Federal-Aid Highway Act of 1934[198] and the Federal-Aid Highway Act of 1944[199] did not include such an express prohibition.[200] Instead, the federal government assigned the primary responsibility for allocating federal road funds to state highway departments.[201]

The 1934 Act represented an incremental improvement, but still did not require states to allocate federal road funds to urban projects.[202] The 1944 Act did establish a minimum threshold for urban road funding, but the percentage remained much lower than that for non-urban areas.[203] Of the available $1.5

---

191. GUTFREUND, *supra* note 181, at 25.
192. *Id.*
193. *Id.* at 26.
194. *Id.* at 27.
195. *See id.*
196. *Id.*
197. Gary T. Schwartz, *Urban Freeways and the Interstate System*, 8 TRANSP. L.J. 167, 175 (1976).
198. Pub. L. No. 73-393, 48 Stat. 993 (1934).
199. Pub. L. No. 78-521, 58 Stat. 838 (1944).
200. Schwartz, *supra* note 197, at 176.
201. *See id.*
202. *See id.*
203. *See* GUTFREUND, *supra* note 181, at 46–47; Schwartz, *supra* note 197, at 176.

billion in total federal road funds, the 1944 Act allocated only twenty-five percent to urban areas.[204]

Yet the amount of funding available for urban areas was deceiving; that money was allocated to highways that happened to cut through urban areas, instead of to enhancements of the city street network itself.[205] Worse still, the 1944 Act's definition of "urban" included any city with over 5,000 residents.[206] The Act's overly broad definition served to further dilute the total urban allocation by forcing major metropolitan areas to share funds with small towns of only several thousand residents.[207]

Thus, while the 1934 and 1944 highway acts permitted states to use federal road funds in urban areas, the acts nevertheless continued to foster sprawl by broadly defining define urban areas and by apportioning a significantly smaller amount of the overall federal funds to those areas.

4. The Federal-Aid Highway Act of 1956

The Federal-Aid Highway Act of 1956[208] is argued to have facilitated more prolific sprawl than any other federal transportation law.[209] By initiating plans for a 41,000-mile national interstate highway network, the Act created a system whereby single-vehicle usage took priority over mass transit in terms of both scope and funding.[210] "'More than any other single measure, the 1956 Act created the decentralized, automobile-dependent metropolis we know today.'"[211]

The 1956 Act had its effect primarily through the Highway Trust Fund, a mechanism created by sister legislation to the Act.[212] The fund, formally

---

204. *See* GUTFREUND, *supra* note 181, at 46–47; Schwartz, *supra* note 197, at 176.

205. *See* GUTFREUND, *supra* note 181, at 46–47 (noting that the funds were for urban extensions, while "ordinary urban streets" remained ineligible for funding); Schwartz, *supra* note 197, at 176 ("[A]n urban area route would qualify . . . only if its function was to 'extend' into a city an otherwise intercity highway . . . .").

206. GUTFREUND, *supra* note 181, at 48. This amounted to a "misappropriation of the 'urban' moniker." *Id.*

207. *Id.*

208. Pub. L. No. 84-627, tit. I, 70 Stat. 374, 374 (1956).

209. *See, e.g.*, Pollard, *supra* note 165, at 1532–33. For a detailed history of the 1956 Act and the events leading up to its passage, see generally Richard F. Weingroff, *Federal-Aid Highway Act of 1956: Creating the Interstate System*, PUB. ROADS, Summer 1996, at 10, *available at* http://www.tfhrc.gov/pubrds/summer96/p96su10.htm.

210. Pollard, *supra* note 165, at 1532–33 ("During this period of unprecedented road-building, public transit received comparatively meager federal funding. This further skewed transportation decision making in favor of highway construction.").

211. *Id.* (quoting U.S. EPA, OUR BUILT AND NATURAL ENVIRONMENTS: A TECHNICAL REVIEW OF THE INTERACTIONS BETWEEN LAND USE, TRANSPORTATION, AND ENVIRONMENTAL QUALITY 10 (2001)).

212. Lambert, *supra* note 171, at 112.

established as part of Hale Boggs's Highway Revenue Act of 1956,[213] served as a depository for revenues resulting from a series of taxes designed to fund the massive interstate highway system.[214] The receipts from the taxes were to be deposited into the fund and redistributed as federal road funding.[215] Even though the fund was supported by a variety of federal user fees, the Hale Boggs Act prohibited those monies from being used on non-highway construction projects, like mass transit.[216]

By passing these two acts, the federal government consummated its "explicit objective . . . to make highway transportation as close to free as possible."[217] In fact, the highways are far from free; automobile use as a whole costs "every man, woman and child in America over four thousand dollars per year."[218]

5. How Subsequent Highway Acts Have Failed to Effectively Confront Sprawl

Fortunately, Congress eventually recognized the unsustainable effects of federal transportation regulations and began to take some preliminary steps toward addressing the problem.[219] In 1991, Congress passed the Intermodal Surface Transportation Efficiency Act (ISTEA), which provided, "It is the policy of the United States to develop a National Intermodal Transportation System that is economically efficient and environmentally sound, provides the foundation for the Nation to compete in the global economy, and will move people and goods in an energy efficient manner."[220] The ISTEA amended prior transportation legislation by allowing some federal transportation funds that had previously been restricted to highway construction to now also be used for alternative transportation options.[221]

The ISTEA was reauthorized in 1998 by the Transportation Equity Act for the 21st Century (TEA-21),[222] which actually required that certain levels of

---

213. Highway Revenue Act of 1956, Pub. L. No. 84-627, tit. II, 70 Stat. 374, 387 (1956).

214. *See* Lambert, *supra* note 171, at 112.

215. *See id.*

216. GUTFREUND, *supra* note 181, at 55.

217. Rachel Weinberger, *The High Cost of Free Highways*, 43 IDAHO L. REV. 475, 496 (2007).

218. Liam A. McCann, *Tea-21: Paving Over Efforts to Stem Urban Sprawl and Reduce America's Dependence on the Automobile*, 23 WM. & MARY ENVTL. L. & POL'Y REV. 857, 880 (1999).

219. *See* Pollard, *supra* note 165, at 1540–42.

220. Intermodal Surface Transportation Efficiency Act of 1991, Pub. L. No. 102-240, 105 Stat. 1914, 1914 (1991).

221. Pollard, *supra* note 165, at 1541.

222. Pub. L. No. 105-178, 112 Stat. 107 (1998).

federal transportation monies be allocated to non-automobile transit projects.[223] The ISTEA and TEA-21 certainly did not level the playing field, but they did represent a shift in transportation policy away from more than seventy-five years of sprawl-conducive regulations by allowing federal monies to be spent on the construction of mass transit projects such as light rail and buses.[224]

Still, when considered as a whole, federal transportation funding remains a key driver in the spread of unsustainable land development patterns in this country.[225] In fact, the idea that the ISTEA and TEA-21 legislation effectively confront sprawl is quite inaccurate. For instance, Kansas City, Missouri is a metropolitan area that maintains the most highway miles per capita in the country.[226] Yet under TEA-21, "[t]he disparity in Kansas City between highway spending and transit spending equal[ed] almost $600 million."[227] In overall dollars, the legislation favored highway funding over transit funding by a five to one margin.[228] In truth, TEA-21 does little more than pay lip service to the ideal of promoting transit alternatives. When coupled with federal gas tax policies, one can see that the overall effect of federal transportation laws remains extremely sprawl-oriented.

B. How Federal Gas Tax Laws Have Induced Sprawl

Though a significant source of sprawl, road building regulations are certainly not the only federal transportation laws that have driven unsustainable growth. Federal gas tax laws have also played a significant role in facilitating sprawl.

1. A Brief History of Gas Taxes

During the early years of the twentieth century, motor vehicle ownership in the United States grew at an astounding rate. Approximately 78,000 motor

223. Pollard, *supra* note 165, at 1541–42.

224. *See id.* Significantly though, at least one commentator has noted that even this change toward funding commuter rail projects could facilitate unsustainable growth patterns "by allowing the marketing and development of sprawling subdivisions beyond the edge cities." Kushner, *supra* note 152, at 862.

225. Congress clearly has moved toward a more sustainable national transportation policy in recent federal highway acts. *See* 49 U.S.C.A. §§ 5501-5506 (Supp. 2007). Because the primary goal of this Article is to examine the historical laws and regulations that originally provoked sprawl, it does not extensively analyze recent trends other than to note that, despite improvements, the legislature continues to prioritize an auto-centric strategy in their policy decisions.

226. McCann, *supra* note 218, at 871–72.

227. *Id.* at 875.

228. *Id.* at 881–82.

vehicles were registered in 1905.[229] By 1921, this number had soared to 1.6 million, and by the end of the decade had increased to more than 23 million.[230] This huge increase in car ownership necessitated a corresponding growth in roads on which to drive them.[231] Yet while consumers paid for the cars, the expense of public roads was generally left to the government.[232]

In 1891, New Jersey became the first state to fund a program that contributed government money to county road construction projects.[233] Within the next several decades, Congress followed suit and began to spend federal money on road building under a system that matched state expenditures dollar for dollar—an approach that resulted in Congress authorizing roughly $75 million between 1916 and 1921.[234] By 1917, every state in the country had created a governmental body charged with road building in order to capitalize on the newly available federal road funds.[235]

Prior to 1920, most states paid for road building from existing revenue sources such as property taxes.[236] As growth in vehicle ownership exploded, though, existing property taxes could no longer pay for the large number of new roads required.[237] Rather than dramatically raise property taxes, states created a new type of user tax—the gas tax—to meet the increased fiscal demand.[238] Oregon, New Mexico, and Colorado passed laws creating a penny per gallon gas tax as early as 1919.[239] By the end of the 1920s, every state in the country had adopted a gas tax of some amount.[240] Then, in 1932, the federal government introduced a national gas tax.[241]

With the new gas tax revenues and skyrocketing auto sales, costs related to road building and maintenance became the second highest government

229. MARK H. ROSE, INTERSTATE: EXPRESS HIGHWAY POLITICS 1939–1989, at 2 (rev. ed. 1990).

230. *Id.*

231. *See id.* at 2–5.

232. *Id.* at 4–5. The obvious exception to government-funded roads is toll roads, where drivers pay at least a portion of the road building and maintenance cost through the tolls. *See generally* Daniel Klein & John Majewski, Turnpikes and Toll Roads in Nineteenth-Century America (Aug. 12, 2004), http://eh.net/encyclopedia/article/Klein.Majewski.Turnpikes (discussing the history of toll roads in America).

233. ROSE, *supra* note 229, at 8. New Jersey initially agreed to pay one-third of the costs for county road building. *Id.*

234. *Id.*

235. *Id.*

236. *Id.* at 4.

237. *See id.*

238. *See id.*

239. *Id.*

240. *Id.*

241. *Id.*; Robert Puentes & Ryan Prince, *Fueling Transportation Finance: A Primer on the Gas Tax*, *in* TAKING THE HIGH ROAD: A METROPOLITAN AGENDA FOR TRANSPORTATION REFORM 45, 48 (Bruce Katz & Robert Puentes eds., 2003).

expenditure of any kind during the 1920s and 1930s.[242] Between 1921 and 1940, this expenditure exceeded $34 billion across all levels of government.[243] With this amount of money at play, it was hardly surprising that, in 1939, the United States Department of Agriculture's Bureau of Public Roads introduced one of the first initiatives to create a national road network that would cover roughly 30,000 miles.[244]

The federal government originally intended that the national gas tax would be a temporary measure used to meet immediate transportation needs and balance the budget.[245] However, rather than allowing it to lapse, Congress has kept the tax in place.[246] Originally, the receipts from the tax were deposited into the general fund and Congress had discretion to appropriate the receipts for non-road-building purposes.[247] Yet interest groups began to strongly oppose any efforts to allocate gas tax revenues to non-highway-construction projects.[248] This opposition prompted legislation such as the Hayden-Cartwright Act of 1934, which expressly reduced federal road funding to states that used gas tax revenue for non-highway-construction projects.[249] This was taken a step further in 1956, when the Federal-Aid Highway Act and the Highway Revenue Act required that gas tax receipts be almost entirely earmarked for roadway building and maintenance expenses.[250]

Notably, this earmark requirement did not provide that gas tax receipts could be used for alternative transportation methods such as mass transit, which falls outside the building and maintenance category.[251] This omission became a major contributing factor to the growth of sprawl.

2. How the Federal Gas Tax Facilitates Sprawl

Unlike several of the other federal laws that have driven sprawl, the unsustainable effects of the federal gas tax are not due to the existence of the tax itself. Indeed, a gas tax could be a useful tool in combating unsustainable growth when the tax is high enough to make long commutes from sprawling suburbs and exurbs—the lifeblood of unsustainable growth—financially infeasible for the average citizen.[252] The federal gas tax induces sprawl not

242. *See* ROSE, *supra* note 229, at 4.
243. *See id.*
244. *See id.*
245. *Id.*; Puentes & Prince, *supra* note 241, at 48.
246. Puentes & Prince, *supra* note 241, at 48.
247. *Id.*
248. GUTFREUND, *supra* note 181, at 32.
249. *Id.*
250. *See* Puentes & Prince, *supra* note 241, at 48; *supra* Part IV.A.4.
251. *See* Puentes & Prince, *supra* note 241, at 48–49; *supra* Part IV.A.4.
252. *But cf.* ANDRES DUANY, ELIZABETH PLATER-ZYBERK & JEFF SPECK, SUBURBAN NATION: THE RISE OF SPRAWL AND THE DECLINE OF THE AMERICAN DREAM 96 (2000) ("[I]t is probably

merely because it is levied, or even because it is too low to reflect the true cost of suburban sprawl. Rather, the true driver of sprawl is the manner in which the federal government appropriates the tax receipts.

The federal earmarking of gas tax receipts solely for road construction and maintenance essentially constitutes a massive federal subsidy of a model of highway building that otherwise could not come close to paying for itself. Ironically, this is exactly the reasoning by which many criticize mass rail transit—that it requires a federal subsidy to stay in business.[253]

For instance, one commentator has argued:

> Since 1972 Amtrak has received more than $13 billion of federal subsidies. Twenty-five years later, Amtrak appears no closer to financial independence than the day taxpayer assistance began. Worse, Amtrak has no apparent plan to become self-sufficient. In fact, it is now pressing for a half-cent of the federal gasoline tax in order to have a permanent umbilical cord to the federal treasury.[254]

Exactly the same thing can be said about federal highways. The federal highway system—the roadway system most often connecting a city center to its suburban and exurban sprawl—received estimated annual funding of $30–60 billion per year during the 1990s.[255] This amounts to a massive subsidy, as not all car owners use the federal highways to the same degree, if at all. Those who live in the suburban outreaches drive the federal highways significantly more than those who live in the cities. Yet all Americans pay the same federal gas tax.

In addition, this "subsidy" means that the tax receipts are inequitably distributed among geographic regions. In particular, statistics show that citizens of urban areas contribute much more in gas taxes than the amount those areas received from its allocation.[256] According to one report,

> Taxpayers in an estimated 158 metropolitan areas received 90 cents or less for each dollar they paid in gas taxes. Some 104 metro areas, including Dallas, Orlando, Tucson and New Orleans, received 75 cents or less; sixty-nine metro areas got back less than two thirds of what their drivers paid in gas taxes.
>
> The result of this funding shortfall is increased traffic congestion, fewer transit options, and more sprawl in outlying areas that is paid for by the

---

unrealistic to hope that legislators will soon take steps, such as enacting a substantial gasoline tax, to allocate fairly the costs of driving.").

253. *See* Stephen Moore, Amtrak Subsidies: This Is No Way to Run a Railroad (May 22, 1997), http://www.cato.org/pub_display.php?pub_id=6146.

254. *Id.*

255. *See* KUSHNER, *supra* note 167, at 12 & n.42.

256. Environmental Working Group, Gas Tax Losers: Metropolitan Areas Get Short End of Federal Gas Tax Funds (Mar. 30, 2004), http://www.ewg.org/reports/gastaxlosers.

> suburban drivers who are increasingly stuck in traffic in and around our nation's cities.[257]

Furthermore, the highway-only earmark has resulted in a complete underfunding of mass transit, making it almost financially impossible for mass transit agencies to invest in the infrastructure necessary to serve as a viable alternative to car travel. The federal government has invested upwards of $182 billion to create the national interstate system, but no comparable investment was ever made in passenger transit infrastructure.[258] For this reason, a recent report proposed that the federal government should address unsustainable growth patterns by "allow[ing] application of gas tax revenues to [a] balanced variety of transportation modes and projects."[259] The report reasoned that "[r]estricting the available resources to roads only inhibits a balanced network by greatly limiting the ability of transit agencies and others to pursue sufficient funding."[260]

Finally, while Congress's 1982 passage of the Surface Transportation Act[261] did require that some federal gas tax receipts be placed into a Mass Transit Account, the present amount of the deposit—only 2.86 cents per gallon of the overall 18.4 cent per gallon tax—remains significantly less than the 15.44 cents apportioned to the highway fund.[262] As a result, the federal gas tax continues to facilitate sprawl by earmarking over 80% of the tax receipts to the federal highways that facilitate unsustainable suburban and exurban growth patterns.

V. THE FEDERAL HOUSING LAWS THAT FACILITATED SPRAWL

Thus far, we have examined how the United States government has facilitated sprawl through federal tax and transportation policies. A survey of the laws that helped make sprawl would not be complete without considering how federal housing laws and regulations have historically contributed to unsustainable growth patterns.

In 1949, the average new home was just over 980 square feet.[263] By 1999, that number had grown to about 2000 square feet, despite the fact that the average household size during that same time period decreased from 3.37

---

257. *Id.*

258. KUSHNER, *supra* note 167, at 11.

259. Puentes & Prince, *supra* note 241, at 69.

260. *Id.*

261. Surface Transportation Assistance Act of 1982, Pub. L. No. 97-424, 96 Stat. 2097 (1983).

262. Puentes & Prince, *supra* note 241, at 49–50.

263. Robert E. Lang & Rebecca R. Sohmer, Editors' Introduction, *Legacy of the Housing Act of 1949: The Past, Present, and Future of Federal Housing and Urban Policy*, 11 HOUSING POL'Y DEBATE 291, 292 (2000).

people to 2.62 people.[264] In short, our homes have grown while our families have shrunk.

One of the main factors in this incongruous result has been a series of federal housing laws and programs that prioritized newly constructed residential developments over the renovation of existing housing.[265] The leading force behind this effort has been the Federal Housing Administration (FHA), whose loan programs have promoted new construction while providing very little for efforts to renovate and repair existing homes.[266] In addition, FHA regulations made it much less expensive to borrow money for new, detached housing than for multi-family, attached units or homes in mixed-use developments.[267] As a result, "families opted to leave their older homes within the central city and move to new homes in the suburbs."[268] The following section examines several of the federal policies that contributed to this residential exodus to the suburban fringes.

A. Early Federal Programs

In September 1931, President Herbert Hoover announced the President's Conference on Home Building and Home Ownership.[269] Charged with "developing the facts and a better understanding of the questions involved and inspiring better organization and the removal of influences which seriously limit the spread of homeownership, both town and country," the conference convened a group of over 1000 participants to consider a national strategy for housing.[270] This initial federal foray into housing policy resulted in a conference report that made claims such as, "[m]ore industries should move to the country, where workers may have better home surroundings," and "[r]ural homes can be made as beautiful and convenient as city homes."[271] Such assertions unfortunately foreshadowed governmental favoritism toward new homes on the suburban, if not rural, fringes.[272]

---

264. *Id.*

265. Lamer, *supra* note 1, at 397.

266. *Id.*

267. *Id.*

268. *Id.*

269. Herbert Hoover, President of the U.S., Statement Announcing the White House Conference on Home Building and Home Ownership (Sept. 15, 1931), *available at* http://www.presidency.ucsb.edu/ws/print.php?pid=22804.

270. *Id.*

271. *Id.* To be fair, though, while the committee's conclusions leaned decidedly toward new, suburban, single-use construction, it did provide some recognition of the value of the existing built environment when it urged that "[o]ld homes should be modernized for the sake of health and convenience." *Id.*

272. *See* Herbert Hoover, *Foreword* to PLANNING FOR RESIDENTIAL DISTRICTS, at xi, xi (John M. Gries & James Ford eds., 1932), *available at* http://www.presidency.ucsb.edu

On the heels of the 1931 conference, Congress passed the 1934 National Housing Act.[273] The 1934 Act was, at least in part, a response to the large number of home mortgage defaults that occurred during the Great Depression.[274] To guard against a reoccurrence of this phenomenon, the Act mandated certain minimum size and building quality standards for homes.[275] The strategy behind creating these minimum standards was to ensure a mandated level of quality, so that even if another economic downturn struck, the homes in default could more easily be resold.[276] Moreover, the Act also promoted a system for longer-term mortgage financing to stimulate home ownership.[277]

Unfortunately, the effect of these legislative efforts, though likely unintended, was the facilitation of sprawl.

> The policies encouraged home ownership by introducing a low-interest, long-term, fully amortized loan with uniform payments over the life of the debt. These policies did not apply evenly to all housing types but favored the development of new single family detached housing at a distance away from the urban core. On the other hand, more urban housing types such as multi-family homes or improvements on existing homes were left unfunded, and there was a disinvestment in inner city neighborhoods as potential home owners moved to the suburbs to take advantage of the available assistance.[278]

In addition to the 1934 Act, the FHA also enabled unsustainable growth patterns through its own administrative regulations. The 1935 FHA Building Codes facilitated sprawl by introducing regulations that would ultimately "make it more profitable for builders to invest in new construction, rather than improve existing structures."[279]

Similarly, the 1938 FHA Underwriting Manual enabled sprawl by giving assurances to lending institutions that the FHA would back loans if the banks

---

/ws/index.php?pid=23502 ("The next great lift in elevating the living conditions of the American family must come from a concerted and nationwide movement to provide new and better homes.").

273. Pub. L. No. 73-479, 48 Stat. 1246 (1934).

274. Gerckens, *supra* note 176, at 17.

275. *Id.*

276. *Id.*

277. National Building Museum, Where Do We Go from Here? Smart Growth and Choices for Change (Apr. 20, 1999) (unpublished exhibition script), http://www.nbm.org/Exhibits/past/2000_1996/Where_Do_We_Go_Script.html.

278. DIV. OF TRANSP. PLANNING, CAL. DEP'T OF TRANSP., *Key Issues and Policy Options Chapter 8: Transportation Planning and Energy*, *in* FUELING THE FUTURE: TRANSPORTATION ENERGY IN CALIFORNIA 2 (2003), *available at* http://www.jfaucett.com/caltransenergy/KIPOPCh8.pdf.

279. National Building Museum, *supra* note 277.

required builders to comply with FHA's new construction standards.[280] This essentially made loans for renovated construction an impractical business decision for banks because those loans could not obtain the same federal backing as new construction.[281]

These 1930s-era efforts served as the foundation for a sprawl-friendly housing policy that Congress would formally embrace in 1949.

B. The Housing Act of 1949

In 1999, the Fannie Mae Foundation commissioned a survey in which scholars identified the "top 10 influences on the American metropolis of the past 50 years."[282] Only one federal policy made the list twice.[283] The Housing Act of 1949, number four on the list, served as the cornerstone for housing sprawl throughout the post-World War II United States.[284] Similarly, the FHA's mortgage financing and subdivision regulations—number two on the list—also helped shape the landscape of post-war America.[285] The primary reason for this policy's inclusion was "the unprecedented suburban growth facilitated by its practices."[286]

The FHA regulations fostered massive suburban growth by implementing lending standards and changing existing land subdivision policies in ways that incentivized new home construction.[287] Indeed, as one researcher concluded,

> FHA-insured mortgages in the two decades after World War II were limited to . . . housing on the suburban fringe; the FHA refused to insure mortgages on older houses in typical urban neighborhoods. This meant that a . . . home buyer who wished to stay in his old neighborhood had to seek old-style conventional mortgages with high rates and short terms. The same purchaser who opted for a new suburban house could get an FHA-insured mortgage with lower interest rates, longer terms, a lower down payment, and a lower monthly payment.[288]

While the housing laws that created sprawl have not been as high-profile as the tax[289] and transportation[290] laws, the cumulative effect was the same: a

---

280. *See id.*

281. *Id.*

282. Robert Fishman, Current Issues, *The American Metropolis at Century's End: Past and Future Influences*, 11 HOUSING POL'Y DEBATE 199, 199–200 (2000).

283. *See id.* at 200.

284. *Id.* at 201.

285. *Id.* at 200, 202.

286. Lang & Sohmer, *supra* note 263, at 292.

287. *See* Fishman, *supra* note 282, at 202 (describing the impact of the new FHA policies).

288. *Id.*

289. *See supra* Part III.

290. *See supra* Part IV.

national housing policy geared toward low-density, single-use, new suburban construction.[291] The housing policy was premised on the consumption of cheaper, peripherally located land, without which the extensive federal benefits directed toward new construction could not be fully realized.[292]

Ultimately, the triumvirate of federal tax, transportation, and housing laws enacted during the early 1900s would stamp the imprint of sprawl on nearly every corner of this country. Yet, while extensive damage has been done to American urbanism, this does not mean we have no other option but to continue down this destructive path. The next section offers several possible policy changes that could begin the slow, but crucial, move toward deconstructing sprawl.

VI. REMEDIES TO SPRAWL-ORIENTED FEDERAL LAWS AND REGULATIONS

With extensive evidence demonstrating how federal laws and regulations have historically facilitated sprawl, the next obvious question is, "How can these harms be remedied?" While a comprehensive set of solutions is beyond the scope of this Article, creative responses to this complex problem can serve as a blueprint for mitigating the federal laws that have made sprawl. The following ideas may serve as one basis for beginning that conversation.

A. Solutions to the Federal Tax Laws that Created Sprawl

1. Commercial Sprawl

As discussed above, the federal regulation that has most facilitated unsustainable patterns of commercial sprawl development has been the depreciation tax deduction, with the deduction's accelerated feature being the most culpable.[293] Therefore, in order to avert continued commercial sprawl growth, Congress should consider regulatory changes that would mitigate the effects of the deduction.

One way Congress could reduce the sprawl-inducing effects of accelerated depreciation would be to amend section 168 of the Internal Revenue Code[294] and mandate longer useful-life periods for newly-constructed buildings than for existing buildings. As a result, businesses would still be able to use the deduction for new construction but at a less accelerated pace than if the businesses simply reused existing structures. From a purely logical standpoint,

---

291. *See* Fishman, *supra* note 282, at 202 (noting the general impact of the FHA's policies on post-war suburbia).

292. *See id.* (relating the FHA's policies to new home construction).

293. *See supra* Part III.A–B.

294. 26 U.S.C.A. § 168 (Supp. 2007).

this proposal makes at least theoretical sense—newly constructed buildings should last longer than existing buildings given that the quality of materials and construction techniques will improve over time. At the same time, this change would incentivize companies to renovate existing buildings in order to realize the full benefit of the accelerated depreciation deduction.

Another option could be for Congress to offer extensive tax credits for developers who build Transit-Oriented Development projects and for the residents or businesses who occupy such new developments.[295] Not only would tax credits provide an economic incentive for mixed-use redevelopment around existing urban transit stations, but the credits would also facilitate the redevelopment of suburban transit stations into more compact and walkable developments.[296] Indeed, such changes are already being realized in several suburban transit stations along the Washington, D.C. Metro lines.[297] Congress should do all it can to facilitate the expansion of this successful strategy along all of the country's existing rail lines.

2. Residential Sprawl

The key federal tax driver behind residential sprawl, the mortgage interest deduction,[298] can also be quickly modified in ways that would reduce its negative effects.

For example, if Congress wanted to act as extensively as possible, it could limit the mortgage interest deduction to homes already in existence or homes built as replacement housing on existing home sites. This would essentially bar those who constructed homes on previously undeveloped land from taking the deduction. This would result in a strong incentive to renovate existing homes or, at the very least, to build new homes within an infill setting.

---

295. *See generally* TRANSP. RESEARCH BD., NAT'L ACADS., TRANSIT-ORIENTED DEVELOPMENT IN THE UNITED STATES: EXPERIENCES, CHALLENGES, AND PROSPECTS (2004), *available at* http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_102.pdf (describing the concept of transit-oriented development).

296. *See generally* NE. ILL. PLANNING COMM'N, TRANSIT-ORIENTED DEVELOPMENT: BUILDING A REGIONAL FRAMEWORK (2001), *available at* http://www.nipc.org/planning/pdf/nipc_transit/pdf (describing the goals of transit-oriented development).

297. *See* Robert Cervero, Christopher Ferrell & Steven Murphy, *Transit-Oriented Development and Joint Development in the United States: A Literature Review*, RES. RESULTS DIG., Oct. 2002, at 1, 22. *See generally* SEATTLE DEP'T OF TRANSP., *Washington D.C. Metro*, *in* STATION AREA PLANNING: TRANSIT-ORIENTED DEVELOPMENT CASE STUDIES, http://www.seattle.gov/transportation/SAP/TOD_Case_Studies/Washington_Metro.pdf (last visited Apr. 6, 2008) (presenting a case study of how Transit-Oriented Development projects have facilitated a more sustainable development pattern for suburban stops along the Washington, D.C. Metro).

298. *See supra* Part III.C–D.

As discussed earlier, the mortgage interest deduction is most often used by taxpayers whose home value well exceeds the national average.[299] This means that the deduction serves as a subsidy for generally larger homes that are disproportionately located on the suburban and exurban fringes.[300] In response to this reality, Congress could limit the use of the deduction to homes with values at or below the regional or state median,[301] similarly to how Congress caps other deductions and tax credits based on adjusted gross income. While this alone would not eliminate exurban sprawl, it could help stall residential sprawl by, at the very least, eliminating its de facto favorable tax treatment.

Either independently or in combination with state and local tax reform proposals,[302] these ideas—while certainly requiring continued research and development—can serve as starting points for modifying the federal tax laws and regulations that have facilitated commercial and residential sprawl.

B. Solutions to the Federal Transportation Laws that Created Sprawl

The federal system for funding transportation is on the cusp of a crisis, one in which spending will outstrip incoming revenue by an increasing amount every year. According to one research group,

> It now appears that the tipping point is expected to hit in FY 2009. Based on the information provided in the Treasury Department's Midsession [B]udget Review forecast for FY 2008, the highway program faces a serious funding crisis beginning in fiscal year 2009. Current Highway Account revenue projections for 2009 show a shortfall of $4.3 billion in revenue. That shortfall will require an obligation reduction in the highway program of about $16 billion . . . .[303]

The reality of this scenario is that either additional funding will need to be generated or dramatic cuts will need to be made to federal transportation spending. Yet the silver lining of this situation is that it provides a unique historical opportunity to re-evaluate how the federal government generates and

---

299. *See supra* Part III.D.

300. *See supra* Part III.D.

301. Using the national median would likely not work well because of the disparity in housing costs among different parts of the country. *See* News Release, Nat'l Ass'n of Realtors, Metro Home Prices Transition in Second Quarter (Aug. 15, 2006), *available at* http://www.realtor.org/press_room/news_releases/2006/2ndqtrmetros06.html.

302. *See generally* Richard W. England, *Property Tax Reform and Smart Growth: Connecting Some of the Dots*, LAND LINE (Lincoln Inst. of Land Policy, Cambridge, Mass.), Jan. 2004, at 8, *available at* http://www.lincolninst.edu/pubs/download.asp?doc_id=377&pub_id=867 (proposing several state and local tax policy controls).

303. AM. ASS'N OF STATE HIGHWAY & TRANSP. OFFICIALS, TRANSPORTATION: INVEST IN OUR FUTURE—REVENUE SOURCES TO FUND TRANSPORTATION NEEDS 21 (2007), *available at* http://www.transportation1.org/tif4report/TIF4-1.pdf.

allocates transportation revenues. The federal government should seize the opportunity to allocate transportation funds more equitably between urban and rural areas while also placing the fiscal burden of unsustainable transportation squarely upon those who insist upon its continued existence. The following recommendations offer several "outside the box" ideas to bring about this goal.

1. Federal Transportation Funding

As discussed earlier, Congress's most recent updates to transportation funding laws allocate some spending toward non-road-building expenditures.[304] Unfortunately, in the overall scheme of things, this is a very small amount.[305] To make meaningful this now merely token treatment, Congress could adopt the following measures.

Congress could allow individuals to claim a tax deduction or tax credit for mass transit fares. This tax break would encourage use of mass transit and could either be applied equally to all mass transit riders or could be graduated by income level.

Congress could also level the playing field between rail and road transportation by removing some of the implicit subsidies realized by the current federal road transportation policies.[306] As one researcher has noted,

> USA railroads have pointed to property taxes as the reason that they have not electrified (no taxes on their diesel, property taxes on electrification infrastructure). Exempting any rail line that electrifies from property taxes under the Interstate Commerce clause would promote the rapid electrification of many rail lines. Expanding capacity would then be more economically attractive without the burden of property taxes. . . . Trucks pay no property taxes, directly or indirectly, on their right-of-way. Trains do.[307]

For an even more innovative approach, Congress could require the collection of tolls from suburban and exurban commuters. These tolls could be graduated based on the distance traveled. Such a tiered system would most heavily charge travel between exurban and urban locations within a region, while charging less for travel from region to region and charging the least for intra-urban travel. This scale would place the highest revenue burden on those who use the least sustainable transportation methods—namely, those who commute every day between the fringes and the city center. The technology for such a system is fairly straightforward. Moreover, the government could either

---

304. *See supra* Part IV.A.5.

305. *See supra* Part IV.A.5.

306. *See supra* Part IV.B.2.

307. Alan Drake, Guest Commentary, A 10% Reduction in America's Oil Use in Ten to Twelve Years (July 9, 2007), http://www.aspo-usa.com/index.php?option=com_content&task=view&id=168&Itemid=93.

lease the toll road, so that its operation is privately funded and managed, or use a funding system similar to those in place on existing toll roads.

Ultimately, removing the funding inequity will be critical to leveling the playing field between vehicular and mass transit travel. Equalizing funding would allow more precise comparisons between the costs and usefulness of these travel systems and enable more sustainability-conscious transportation planning.

2. Federal Gas Tax

While the federal gas tax represents the clichéd 500 pound gorilla in the area of transportation funding, Congress can still take several steps to distribute more equitably the massive amounts of revenue generated from these tax receipts.

For example, Congress could require federal gas tax receipts from urbanized areas to be allocated more evenly between road construction and transit construction. This would recognize that mass transit is a much more viable transportation option in compact urban areas. Cities would thus have the opportunity to fund more sustainable transit options rather than essentially being forced to spend most of their monies on road building.

Congress also could incentivize both the development and consumption of more sustainable transportation patterns by increasing the tax on non-renewable fuels such as gasoline while decreasing consumption taxes on renewable energy sources, such as solar technology and renewable fuels.

C. Solutions to the Federal Housing Laws that Created Sprawl

While the federal government has taken steps to remedy national housing laws that have historically induced sprawl,[308] additional steps should be taken to further address the issue.

The Taxpayer Relief Act of 1997[309] provides that most single homeowners can avoid paying capital gains tax on up to $250,000 of the profit from the sale of a home, while married homeowners can avoid capital gains tax on profits up to $500,000.[310] One of the few requirements is that, prior to the sale, the house must have served as the homeowner's primary residence for two of the last five years.[311] If that small hurdle is cleared, then the buying and selling of homes makes for an excellent investment option—one that implicitly encourages

---

308. *See, e.g.*, Lang & Sohmer, *supra* note 263, at 297 ("Fallout from the [1949 Housing Act's] failures sparked a variety of subsequent policy reforms.").

309. Pub. L. No. 105-34, 111 Stat. 788 (1997).

310. Blanche Evans, *Taxpayer Relief Act Of 1997 Still Fueling Housing Boom*, REALTY TIMES, Aug. 9, 2005, http://realtytimes.com/rtapages/20050809_taxpayerrelief.htm.

311. *Id.*

homeowners to regularly buy and sell houses in order to realize a massive capital gains tax break.[312] In most instances, these newly purchased homes are newly constructed ones.[313] Thus, the tax break ultimately increases the construction of new homes, a clear indicator of sprawl.[314] To counter this, Congress should allow a homeowner to take the tax break only if she replaces the home she sells with the purchase or rental of a previously existing residence. This would still allow homeowners to shield the appreciation in value of their homes from capital gains taxes, but only if they reinvested in the existing housing stock.

Congress could also limit sprawl by eliminating individuals' ability to deduct mortgage interest and property taxes on second homes. Indeed, one might wonder why, if a homeowner can afford two houses, he even needs this type of deduction.[315] Moreover, second homes are especially prone to requiring long car trips to reach.[316] According to the National Association of Realtors, "[t]he typical vacation-home owner . . . purchase[s] a property that is 220 miles from their primary residence."[317] Ultimately, there is no legitimate reason for the federal government to help subsidize second homes, many of which are newly constructed, and thereby contribute to sprawl.

Congress could also promote infill development by providing tax credits for buyers who renovate all buildings, not just historic ones, within existing neighborhoods. Additional tax credits for developers and home buyers who renovate existing buildings in accordance with environmentally sound "green building" practices could further sweeten the pot.[318] Such policies could be patterned after the existing federal program wherein Congress provides tax credits to purchasers of environmentally friendly hybrid vehicles.[319]

---

312. *See id.* (comparing the tax break received on home investments and stock investments).

313. *See* Lamer, *supra* note 1, at 398 (describing tax policies that contribute to sprawl and new home construction).

314. *See id.*

315. According to the National Association of Realtors,

> The median size of a vacation home is 1,480 square feet, 29 percent were new when purchased, and owners estimated the current value to be a median of $300,000—68 percent said the value of that property was lower than their primary residence. Sixty-five percent of owners said their vacation property was a better investment than stocks.

News Release, Nat'l Ass'n of Realtors, Second-Home Owner Survey Shows Solid Market, Appetite For More (May 11, 2006), *available at* http://www.realtor.org/press_room/news_releases/2006/2ndhomesurvey06.html.

316. *Id.*

317. *Id.*

318. *See generally* U.S. Green Building Council, http://www.usgbc.org/ (last visited Apr. 6, 2008) (offering information related to green building practices).

319. *See generally* U.S. Environmental Protection Agency & U.S. Department of Energy, New Energy Tax Credits for Hybrids, http://www.fueleconomy.gov/feg/tax_hybrid.shtml (last visited Apr. 6, 2008) (providing information related to the federal government's hybrid vehicle

Finally, the Department of Commerce could address the sprawl-friendly Standard State Zoning Enabling Act and the Standard State Planning Enabling Act, promulgated as model laws in the 1920s,[320] by developing a Twenty-first Century Standard State Zoning and Planning Enabling Act. This tool could be used by states as a model for replacing their current, use-based enabling acts with ones that prioritize form-based coding and planning.[321] Such a move by the federal government would represent an explicit renunciation of single, separated-use-based zoning and planning as the predominant type of land use regulation and replace it with a more sustainable planning scheme.

VII. CONCLUSION

With the harmful costs of sprawl well-documented, understanding the root causes of this unsustainable phenomenon is imperative. While a variety of non-legal factors have facilitated sprawl, the largest drivers of unsustainable land development patterns have been laws and regulations that promoted, if not mandated, these results.

Although state and local policies have played a major role in the proliferation of sprawl, the triumvirate of federal regulatory areas discussed herein has most systematically fostered this trend. This Article has focused on analyzing how federal tax, transportation, and housing regulations have entrenched sprawl as the dominant land development pattern in the United States. Hopefully, by understanding the root causes of sprawl and beginning to discuss novel solutions to the deeply entrenched problem, we can avoid repeating history and counter the continued negative effects of sprawl on this country's built environment.

---

tax credit program).

320. *See generally* Chad D. Emerson, *Making Main Street Legal Again: The SmartCode Solution to Sprawl*, 71 MO. L. REV. 637, 652–54 (2006) (discussing the background, purpose, and structure of the Standard State Zoning Enabling Act).

321. *See id.* at 641.

EXPLAINING THE SPREAD OF AT-WILL EMPLOYMENT AS AN INTERJURISDICTIONAL RACE TO THE BOTTOM OF EMPLOYMENT STANDARDS

RICHARD A. BALES*

I. INTRODUCTION

The employment-at-will rule[1] is a default rule that courts apply when parties in an employment relationship have failed to specify the duration of that relationship. Courts in forty-nine of fifty states apply this rule.[2] The rule, succinctly stated, is that absent a prior agreement, either the employer or the employee may terminate the employment relationship at any time for any lawful reason without notice.[3]

In the late 1800s and early 1900s, the at-will rule replaced the English rule, described by William Blackstone, that employment should be presumed to be for a year, or "throughout all the revolutions of the respective seasons."[4] The

---

* Professor of Law and Associate Dean of Faculty Development, Salmon P. Chase College of Law, Northern Kentucky University. B.A., Trinity University, 1990; J.D., Cornell Law School, 1993. Special thanks to James Atleson, John Bickers, Jeff Hirsch, Sanford Jacoby, Gillian Lester, Paul Secunda, research assistant Anna Maki, and faculty assistant Judith Brun. This article benefited tremendously from comments received at the 2007 University of Cincinnati College of Law Summer Scholarship Series and the 2007 Second Annual Colloquium on Current Scholarship in Labor and Employment Law.

1. For discussions of the at-will rule and its provenance, see generally Deborah A. Ballam, *Exploding the Original Myth Regarding Employment-at-Will: The True Origins of the Doctrine*, 17 BERKELEY J. EMP. & LAB. L. 91 (1996) [hereinafter Ballam, *Exploding the Original Myth*]; Deborah A. Ballam, *The Traditional View on the Origins of the Employment-at-Will Doctrine: Myth or Reality?*, 33 AM. BUS. L.J. 1 (1995) [hereinafter Ballam, *Traditional View*]; Jay M. Feinman, *The Development of the Employment at Will Rule*, 20 AM. J. LEGAL HIST. 118 (1976); Mayer G. Freed & Daniel D. Polsby, *The Doubtful Provenance of "Wood's Rule" Revisited*, 22 ARIZ. ST. L.J. 551 (1990); Sanford M. Jacoby, *The Duration of Indefinite Employment Contracts in the United States and England: An Historical Analysis*, 5 COMP. LAB. L. 85 (1982); Andrew P. Morriss, *Exploding Myths: An Empirical and Economic Reassessment of the Rise of Employment At-Will*, 59 MO. L. REV. 679 (1994); Theodore J. St. Antoine, *You're Fired!*, 10 HUM. RTS. 32 (1982); and Clyde Summers, *Individual Protection Against Unjust Dismissal: Time for a Statute*, 62 VA. L. REV. 481 (1976).

2. Montana is the outlier. *See* Wrongful Discharge From Employment Act, MONT. CODE ANN. §§ 39-2-901 to -915 (2007).

3. *See, e.g.*, Feinman, *supra* note 1, at 118.

4. 1 WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND 413 (Univ. of Chi. Press 1979) (1765).

at-will rule often is attributed[5] to Horace Gay Wood, who described it in an 1877 treatise.[6] Over the next forty years, judges in most American states adopted the rule.[7] How and why the rule spread, however, has been the subject of considerable academic debate. This Essay argues that the at-will rule spread because states were economically pressured to keep their labor markets competitive with other states, thus precipitating a "race to the bottom" in employment standards.

The prevailing wisdom is that the at-will rule spread because of a judiciary fixated, from about 1890 to 1930, on laissez-faire reasoning and freedom of contract.[8] However, during this same time period, courts often imposed at-will terms on parties who clearly had intended to contract for something other than at-will employment.[9] As Professor Sanford Jacoby has explained, this refusal to consider party intent was inconsistent with the then-prevailing contractualist judicial philosophy.[10] For this reason, Samuel Williston, the leading contract scholar of the time, considered the at-will rule a deviation from general contract principles.[11]

Professor Jacoby proposes that weak American trade unions allowed the at-will rule to be imposed on manual laborers, and the rule then spread to white collar workers.[12] Professor Jay Feinman, on the other hand, suggests that the at-will rule actually spread in the opposite direction—from salaried mid-level managers to manual workers.[13] Feinman concludes from this that the spread of at-will employment must have been a product of industrialization.[14] As the American economy grew from local to national and the means of production shifted from artisanal labor to large factories, mid-level managers became more important, prevalent, and potentially powerful.[15] The emergence of the at-will

---

5. *See, e.g.*, St. Antoine, *supra* note 1, at 33 (describing the at-will rule as having "spr[u]ng full-blown . . . from [Wood's] busy and perhaps careless pen").

6. H.G. WOOD, A TREATISE ON THE LAW OF MASTER AND SERVANT 280–83 (1877).

7. *See, e.g.*, Feinman, *supra* note 1, at 126–27; Morriss, *supra* note 1, at 688.

8. *See, e.g.*, Phung v. Waste Mgmt., Inc., 491 N.E.2d 1114, 1118 (Ohio 1986) (Brown, J., dissenting); Mary Ann Glendon & Edward R. Lev, *Changes in the Bonding of the Employment Relationship: An Essay on the New Property*, 20 B.C. L. REV. 457, 458 (1979); Jacoby, *supra* note 1, at 116.

9. *See, e.g.*, Odom v. Bush, 53 S.E. 1013, 1014, 1016 (Ga. 1906) (finding that plaintiff was employed at-will notwithstanding facts that plaintiff had given up a job, sold his house and possessions, invested the proceeds in the defendant's business, moved hundreds of miles to establish the new plant, and existed on a subsistence salary while the new plant was being established).

10. Jacoby, *supra* note 1, at 116–18.

11. *See* 1 SAMUEL WILLISTON, THE LAW OF CONTRACTS 60–63 (1920).

12. Jacoby, *supra* note 1, at 85–86.

13. *See* Feinman, *supra* note 1, at 130–31.

14. *See id.* at 131–32.

15. *See id.* at 132–33.

rule gave capitalists the ability to re-assert control over the workplace and entitlement to profits.[16]

Professor Andrew Morriss "explodes the myth" that at-will employment was linked to industrialization by demonstrating that the first states to adopt at-will employment tended to be western populist states and southern states, not the northeastern states presumably dominated by capital and big business.[17] Morriss's observation is valid, but his conclusion that this necessarily disproves any linkage between at-will employment and industrialization[18] should be rejected. Instead, it may have been precisely *because* the western and southern states were underindustrialized that they were among the first to adopt the at-will rule. Underindustrialized states needed a way to attract capital, and one of their options was to offer attractive employment rules to capitalists deciding where to build their next factory.

In any event—and more importantly—once the first underindustrialized states adopted the rule, other underindustrialized states would have been compelled to follow suit to remain economically competitive with the early adopters. Industrialized states would then have been compelled to adopt the rule, as well, to maintain their competitive advantage in the labor market. The adoption of the at-will rule by a handful of underindustrialized states, therefore, precipitated an interjurisdictional race to the bottom[19] in employment standards, culminating in the universal adoption of the at-will rule. This competitive labor market-based explanation for the spread of the at-will rule is consistent with contemporaneous attempts to regulate child labor, to establish an unemployment insurance program, and to set a minimum wage.

Although the focus of this Essay is on labor market conditions that existed as the United States was transitioning from a local to a national economy, the implications resonate today, as the United States and other developed countries transition from national to international economies. Just as underindustrialized states once obtained a competitive advantage in the national labor market by adopting at-will employment, so are underindustrialized countries today obtaining a competitive advantage in the global labor market through low wages and un- or underregulated working conditions. Likewise, just as industrialized states once sought to regain their competitive advantage in the national labor market by standardizing many employment terms (originally by

16. *Id.* at 133.

17. Morriss, *supra* note 1, at 681, 753.

18. *See id.* at 753, 763.

19. "Top" and "bottom" are necessarily relative terms. A free-market economist like Richard Epstein might argue, for example, that at-will employment is at the "top" and that overregulation of the labor market is at the "bottom." *Cf.* Richard A. Epstein, *In Defense of the Contract at Will*, 51 U. CHI. L. REV. 947, 951 (1984) (arguing in favor of employment contract constructions that "advance individual autonomy and . . . promote the efficient operation of labor markets"). This Essay, however, adopts the terminology used in most of the race-to-the-bottom literature and assumes that the "top" represents worker (investor, environmental, etc.) protection.

adopting at-will employment and later by setting wage rates through the Fair Labor Standards Act), the United States today is attempting to brake the race to the bottom in the global labor market by proposing an international trade framework that would standardize many employment terms such as child labor, forced labor, collective bargaining, and discrimination. Other aspects of the American experience with labor market competition may also prove instructive for developed countries addressing the international labor market.

II. THE EMPLOYMENT-AT-WILL RULE

In 1765, William Blackstone wrote in his Commentaries that the length of employment should be presumed to be a year, or "throughout all the revolutions of the respective seasons; as well when there is work to be done, as when there is not."[20] Blackstone's rule, predicated on the presumption of a largely agricultural workforce, was designed in part to avoid opportunism.[21] Because farmworkers worked long hours during the growing season but little during winter, they were at risk of being discharged in late fall after the harvest.[22] Likewise, the employer was at risk of workers leaving or demanding unreasonable wages during the summer and early fall.[23] Blackstone's rule also was designed in the shadow of earlier English statutes, such as the Statute of Artificers, which forbade hirings for less than a year as a means of restricting labor mobility,[24] and the Poor Laws, which used a test of residence and employment to determine which community was responsible for supporting a pauper.[25]

Scholars disagree about the extent to which Blackstone's rule was transported to the American colonies and later followed in American states. Professor Jacoby asserts that "[t]he colonies followed English usage" of employment for a set term.[26] In contrast, Deborah Ballam argues that at-will employment was common in practice, and at least to some extent reified in law, starting in the colonial period.[27] Professor Feinman describes American law from colonial days to the mid-1800s as "a confusion of principles and rules."[28]

---

20. BLACKSTONE, *supra* note 4, at 413.

21. *See id.*; Ballam, *Traditional View*, *supra* note 1, at 1–2.

22. Ballam, *Traditional View*, *supra* note 1, at 2.

23. *Id.*; *cf.* Stewart J. Schwab, *Life-Cycle Justice: Accommodating Just Cause and Employment at Will*, 92 MICH. L. REV. 8, 10–11 (1993) (arguing that modern employees and employers face similarly shifting risks of opportunism throughout the employment life cycle).

24. *See generally* Donald Woodward, *The Background to the Statute of Artificers: The Genesis of Labour Policy, 1558–63*, 33 ECON. HIST. REV. 32 (1980) (describing the policy goals motivating the Statute of Artificers and other Tudor-era labor laws).

25. *See* Ballam, *Traditional View*, *supra* note 1, at 2; *see also* 1 C.B. LABATT, COMMENTARIES ON THE LAW OF MASTER AND SERVANT 39 (1913) (commenting on the statutory relationship between paupers and guardians).

26. Jacoby, *supra* note 1, at 91.

27. Ballam, *Exploding the Original Myth*, *supra* note 1, at 98.

28. Feinman, *supra* note 1, at 122.

In the colonies, he explains, day laborers tended to be employed at will, while agricultural laborers and domestic servants were usually hired on a yearly basis.[29] During the 1800s, though, he finds that "whatever consensus [had] existed about the state of the law dissolved."[30]

In any event, scholars seem to agree that most American states did not formally begin to adopt the at-will rule until Horace Gay Wood, a New York State treatise writer, published Master and Servant in 1877.[31] Wood wrote:

> [T]he rule is inflexible, that a general or indefinite hiring is *prima facie* a hiring at will, and if the servant seeks to make it out a yearly hiring, the burden is upon him to establish it by proof. . . . It is competent for either party to show what the mutual understanding of the parties was in reference to the matter; but unless their understanding was mutual that the service was to extend for a certain fixed and definite period, it is an indefinite hiring and is determinable at the will of either party . . . .[32]

Wood was not claiming to have invented a new rule. Instead, he thought that he was accurately summarizing an existing majority rule. Scholars have since questioned whether Wood correctly interpreted the law. Theodore J. St. Antoine, for example, has described the at-will rule as having "spr[u]ng full-blown . . . from [Wood's] busy and perhaps careless pen."[33]

The controversy centers on the four cases cited in Wood's Footnote 4.[34] None of the holdings of these cases are directly on point. Two of the cases were employee victories[35]—weak support for a supposed legal rule giving employers an unfettered right to fire. One of the cases did not involve an employment contract at all, but rather a contract between the U.S. Army and a private company for the transportation of goods.[36] The final case involved a

---

29. *Id.*

30. *Id.*

31. *See* Ballam, *Exploding the Original Myth*, *supra* note 1, at 92–93 (discussing scholarship on Wood's statement of the at-will rule); *cf.* Lawrence E. Blades, *Employment at Will vs. Individual Freedom: On Limiting the Abusive Exercise of Employer Power*, 67 COLUM. L. REV. 1404, 1405 (1967) (citing a post-1877 case, *Payne v. Western & Atlantic Railroad Co.*, 81 Tenn. 507 (1884), for the "traditional rule" that an employer has absolute power of discharge and can terminate employment "'even for cause morally wrong'"). As of 1876, only seven states applied some form of an at-will rule. *See* Morriss, *supra* note 1, at 700 tbl.II.

32. Wood, *supra* note 6, at 272.

33. St. Antoine, *supra* note 1, at 33.

34. For a thorough discussion of these cases and evaluation of the quality of Wood's scholarship, see generally STEVEN L. WILLBORN ET AL., EMPLOYMENT LAW: CASES AND MATERIALS 53–54 (2007); Feinman, *supra* note 1; and Freed & Polsby, *supra* note 1.

35. Tatterson v. Suffolk Mfg. Co., 106 Mass. 56, 59 (1870); Franklin Mining Co. v. Harris, 23 Mich. 115, 116–17 (1871).

36. Wilder v. United States (Wilder's Case), 5 Ct. Cl. 462, 462 (1869), *rev'd*, 80 U.S. 254 (1871).

bartender who lived on the employer's premises; when he was fired, he challenged his eviction but not his discharge.[37]

Regardless of the rule's provenance, it quickly became the national norm.[38] As Feinman points out,

> Whatever its origin and the inadequacies of its explanation, Wood's rule spread across the nation until it was generally adopted. New York, for example, . . . adopted the rule in 1895, the Court of Appeals noting that the rule was "correct" and by then widely in use. Charles Labatt, author of the next great master and servant treatise and proponent of a different rule, in 1913 conceded that Wood's rule was the law in a "great majority of the states," and Williston, bemoaning the failure of the courts to adhere to true contract principles, admitted the universal application of Wood's employment at will rule.[39]

Thus, by the end of the 1930s, nearly every state had formally adopted the at-will rule.[40]

The at-will rule that courts adopted, however, was not identical to the at-will rule that Wood had described. Wood had described the at-will rule as a default rule and noted that parties to an employment arrangement were free to contract for a fixed term of employment.[41] Early cases, however, show that courts interpreted the rule as making it very difficult for employees to demonstrate that the parties intended anything other than at-will employment.[42] As Jacoby points out, courts invariably concluded that "permanent" or "lifetime" employment contracts were at-will, even in cases where the employees had bargained for long-term employment by agreeing to drop injury claims against their employers.[43]

Such decisions were common through the 1950s. One example is *Skagerberg v. Blandin Paper Co.*,[44] in which an employer promised an employee "permanent" employment in return for the employee's acceptance of a job offer, rejection of a competing offer, and purchase of a supervisor's home.[45] When the employee challenged his subsequent discharge, the court dismissed the case, stating that "[i]n case the parties to a contract of service expressly agree that the employment shall be 'permanent[,]' the law implies,

---

37. DeBriar v. Minturn, 1 Cal. 450, 451 (1851).

38. *See* Morriss, *supra* note 1, at 688.

39. Feinman, *supra* note 1, at 126–27 (citing Martin v. N.Y. Life Ins. Co., 148 N.Y. 117 (1895); LABATT, *supra* note 25, at § 159 n.2; and WILLISTON, *supra* note 11, at § 39).

40. *See* Morriss, *supra* note 1, at 700 tbl.II.

41. Wood, *supra* note 6, at 272.

42. *See* Morriss, *supra* note 1, at 684.

43. Jacoby, *supra* note 1, at 117–18.

44. 266 N.W. 872 (Minn. 1936).

45. *Id.* at 873.

not that the engagement shall be continuous or for any definite period, but that the term being indefinite the hiring is merely at will."[46]

In 1959, the at-will pendulum began to swing in the opposite direction. In *Petermann v. International Brotherhood of Teamsters, Local 396*,[47] a California appellate court held that the employer's right to fire an at-will employee could be limited by considerations of public policy in a situation where the employee was terminated for refusing to commit perjury.[48] Since the mid-seventies, courts have eroded the at-will rule by applying contract and tort principles to restrict employers' ability to fire employees.[49] Courts, both federal and state, are far more likely today than they were forty years ago to enforce employer promises made orally[50] or in employee handbooks[51] and to apply the covenant of good faith and fair dealing[52] to employee discharges. Courts also are more likely to apply tort doctrines like wrongful discharge in violation of public policy[53] and intentional infliction of emotional distress[54] to employee discharges and other adverse employment actions. Similarly, Congress and state legislatures have passed statutes prohibiting discrimination (including discharge) on such bases as "race, color, religion, sex,"[55] pregnancy,[56] national origin,[57] age,[58] disability,[59] and now, perhaps, an employee's genetic information.[60]

Today, the at-will rule remains the default employment rule in every state but Montana[61] but is subject to myriad exceptions. Beginning with Lawrence

---

46. *Id.* at 873–74 (citation omitted).

47. 344 P.2d 25 (Cal. Dist. Ct. App. 1959).

48. *See id.* at 27.

49. *See* Clyde W. Summers, *Labor Law as the Century Turns: A Changing of the Guard*, 67 NEB. L. REV. 7, 12–13 (1988).

50. *See, e.g.*, Toussaint v. Blue Cross & Blue Shield of Mich., 292 N.W.2d 880, 885 (Mich. 1980).

51. *See, e.g.*, Woolley v. Hoffmann-La Roche, Inc., 491 A.2d 1257, 1258, *judgment modified*, 499 A.2d 515 (N.J. 1985).

52. *See, e.g.*, Metcalf v. Intermountain Gas Co., 778 P.2d 744, 748 (Idaho 1989), *modified*, Sorensen v. Comm Tek, Inc., 799 P.2d 70 (Idaho 1990).

53. *See, e.g.*, Nees v. Hocks, 536 P.2d 512, 515–16 (Or. 1975) (in banc).

54. *See, e.g.*, Agis v. Howard Johnson Co., 355 N.E.2d 315, 317–18 (Mass. 1976).

55. 42 U.S.C. § 2000e-2(a) (2000); *e.g.*, 43 PA. CONS. STAT. ANN. § 955 (West 2007).

56. *E.g.*, 42 U.S.C. §§ 2000e(k), 2000e-2(a); *see, e.g.*, Dallastown Area Sch. Dist. v. Commonwealth, Pa. Human Relations Comm'n, 460 A.2d 878, 880 (Pa. Commw. Ct. 1983).

57. *E.g.*, 42 U.S.C. § 2000e-2(a); 43 PA. CONS. STAT. ANN. § 955.

58. *E.g.*, 29 U.S.C. § 623 (2000); 43 PA. CONS. STAT. ANN. § 955.

59. *E.g.*, 42 U.S.C. § 12112 (2000); 43 PA. CONS. STAT. ANN. § 955.

60. *See* Karen L. Werner, *Discrimination: Genetic Nondiscrimination Measure Approved by House Panel Despite Scope, Impact Issues*, 57 DAILY LAB. REP. (BNA) A-1 (Mar. 26, 2007), *available at* 57 DLR A-1, 2007 (Westlaw) (discussing proposed legislation "that would prohibit employers and health insurers from discriminating against individuals on the basis of genetic information").

61. *See* Wrongful Discharge From Employment Act, MONT. CODE ANN. §§ 39-2-901 to -

Blades's trailblazing 1967 article,[62] scholars have widely criticized the rule as an anachronism.[63]

III. EXPLANATIONS FOR THE SPREAD OF THE AT-WILL RULE

All but two of the states that adopted the at-will employment rule in the forty years after Wood wrote his treatise did so by judicial decision rather than by statute.[64] These courts made "little attempt to support the adoption of the rule with arguments or analysis" and often provided little or no citation to authority.[65] For this reason, it is probably impossible to know for certain why a given court or judge decided to adopt the rule. The best that researchers can do is examine other social, political, and legal trends and infer causation from juxtaposition. Commentators have proposed several possible explanations.

The first proffered explanation is that the at-will rule is the product of a judiciary fixated on laissez-faire reasoning and freedom of contract. At-will employment became firmly entrenched in the United States at about the same time the Supreme Court was invalidating, on due process grounds, multiple legislative attempts to regulate the employment relationship.[66] Therefore, many courts and commentators have assumed that courts readily adopted the rule of at-will employment because that rule was consistent with the prevailing judicial orthodoxy of laissez-faire contractualism.[67]

---

915 (2007).

62. Blades, *supra* note 31, at 1404–05.

63. *See, e.g.*, Pauline T. Kim, *Bargaining with Imperfect Information: A Study of Worker Perceptions of Legal Protection in an At-Will World*, 83 CORNELL L. REV. 105, 106–08 (1997); Arthur S. Leonard, *A New Common Law of Employment Termination*, 66 N.C. L. REV. 631, 675 (1988); Daniel J. Libenson, *Leasing Human Capital: Toward a New Foundation for Employment Termination Law*, 27 BERKELEY J. EMP. & LAB. L. 111, 118 (2006); Stewart J. Schwab, *Life-Cycle Justice: Accommodating Just Cause and Employment at Will*, 92 MICH. L. REV. 8, 9–10 (1993); Theodore J. St. Antoine, *A Seed Germinates: Unjust Discharge Reform Heads Toward Full Flower*, 67 NEB. L. REV. 56, 67 (1988); Summers, *supra* note 1, at 484.

64. *See* Morriss, *supra* note 1, at 700 tbl.II. The two states adopting the at-will rule by statute during that period were the Dakota Territory (now North and South Dakota) and Montana. *Id.*

65. *Id.* at 697.

66. *See, e.g.*, Adair v. United States, 208 U.S. 161, 178, 180 (1908), *overruled in part by* Phelps Dodge Corp. v. NLRB, 313 U.S. 177, 186–87 (1941); Lochner v. New York, 198 U.S. 45, 53, 64 (1905), *overruled in part by* Day-Brite Lighting, Inc. v. Missouri, 342 U.S. 421, 422–24 (1952); *cf.* Muller v. Oregon, 208 U.S. 412, 421 (1908) (acknowledging the "general right to [freedom of] contract" under the Fourteenth Amendment but upholding protective legislation for women as an exception to that rule, on the basis that their "physical structure and the performance of maternal functions place [them] at a disadvantage in the struggle for subsistence . . . especially . . . when the burdens of motherhood are upon" them).

67. *See, e.g.*, Morriss v. Coleman Co., 738 P.2d 841, 852 (Kan. 1987) (Herd, J., concurring) ("The doctrine of employment-at-will and termination without cause is a carry-over from 19th Century laissez faire economic philosophy . . . ."); Pierce v. Ortho Pharm. Corp., 417

As described in Part II, though, the at-will rule adopted by courts at the turn of the nineteenth century was different from the at-will rule Wood had articulated in 1877.[68] Wood's rule was a default rule, while the judicial version of the rule was virtually impermeable.[69] Williston, and later Jacoby, explained that a strict contractarian would have considered rate of pay (such as "income of $1000 per month") to be evidence "that the parties intended the employment to last at least for one such period" of payment.[70] Most courts, however, refused to do so.[71] Moreover, a strict contractarian would have considered the parties' intent in construing contracts for "permanent" or "lifetime" employment.[72] Courts instead construed such contracts reflexively as at-will employment contracts.[73] Thus, while Wood's default rule was largely contractualist because an employee could rebut the presumption with evidence that the parties intended otherwise,[74] the judicial version of the rule was anti-contractualist because it made party intent essentially irrelevant.[75] Therefore, the at-will rule adopted by courts at the turn of the nineteenth century represented not laissez-faire contractualism, but rather a doctrine whereby an employer has nearly absolute control over employment terms.[76]

A.2d 505, 509 (N.J. 1980); Kurt H. Decker, *Pennsylvania's Whistleblower Law's Extension to Private Sector Employees: Has the Time Finally Come to Broaden Statutory Protection for All At-Will Employees?*, 38 DUQ. L. REV. 723, 731 (2000); Matthew W. Finkin, *Shoring Up the Citadel (At-Will Employment)*, 24 HOFSTRA LAB. & EMP. L.J. 1, 24–25 (2006) (noting "the judiciary's . . . wholehearted embrace of *laissez-faire* in the last quarter of the nineteenth century, captured by the at-will rule . . ."); Matthew C. Palmer, Note, *Where Have You Gone, Law and Economics Judges? Economic Analysis Advice to Courts Considering the Enforceability of Covenants Not to Compete Signed After At-Will Employment Has Commenced*, 66 OHIO ST. L.J. 1105, 1112 (2005) ("Coupled with the laissez faire economic philosophy that reigned supreme at the time, the employment at will doctrine became established as a fundamental aspect of labor relations."); Matthew White, Comment, *Conscience Clauses for Pharmacists: The Struggle to Balance Conscience Rights with the Rights of Patients and Institutions*, 2005 WIS. L. REV. 1611, 1628 ("The employment-at-will doctrine is an extension of the venerable laissez-faire notion that, in most cases, it is better to keep the law away from how private individuals manage their capital.").

68. *See supra* text accompanying notes 41–43.

69. *See supra* text accompanying notes 41–43.

70. WILLISTON, *supra* note 11, at 62; Jacoby, *supra* note 1, at 117 (quoting *id.*).

71. WILLISTON, *supra* note 11, at 61–62; Jacoby, *supra* note 1, at 117.

72. Jacoby, *supra* note 1, at 117; *see* WILLISTON, *supra* note 11, at 62.

73. WILLISTON, *supra* note 11, at 61; Jacoby, *supra* note 1, at 117.

74. *Cf.* Feinman, *supra* note 1, at 130 (arguing that Wood's original rule was not purely contractualist because "an artificial presumption of terminability was introduced; the parties' intentions were secondary, to be considered only rarely to rebut the presumption").

75. WILLISTON, *supra* note 11, at 61–62; Jacoby, *supra* note 1, at 116–18.

76. *See* Morriss, *supra* note 1, at 690 (noting that the at-will rule constituted "an active state intervention in favor of employers rather than [the] state neutrality" implied by the term *laissez faire*).

Jacoby points out that a variety of socio-legal developments preceded the formal adoption of the at-will rule by courts in the late 1800s and early 1900s.[77] These developments, including changes in laws that had permitted settlement by hiring, the lack of criminal enforcement of long-term employment contracts, and the absence of a rule requiring notice before termination of employment, facilitated the shift in presumption from Blackstone's annual hiring to at-will employment.[78] Jacoby demonstrates that the formation of trade unions in the United States precipitated the formal adoption of the at-will rule.[79] American courts were hostile toward unions, so the unions relied on legislation, such as laws prohibiting discriminatory dismissals and yellow-dog contracts, and collective bargaining agreements to protect workers.[80] Courts reacted by asserting that employers nonetheless had the "'fundamental right' to dismiss trade unionists at will"[81]—an assertion readily adapted to salaried employees as well.[82]

Feinman builds on this observation by pointing out that salaried workers and mid-level managers, rather than manual workers, brought the cases in which the at-will rule was originally adopted.[83] From this, Feinman offers a third explanation for the spread of the at-will employment rule: "[T]he rule was an adjunct to the development of advanced capitalism in America."[84] As the American economy grew from local to national and the means of production shifted from artisanal labor to large factories, professionals and mid-level managers became more important, prevalent, and potentially powerful.[85] Thus, they "might have been expected to seek a greater share in the profits" and degree of control over the companies for which they worked.[86] The at-will rule made the employment of these mid-level employees precarious and permitted owners to reassert absolute control over the workplace and entitlement to profits.[87] At-will employment, then, was "the ultimate guarantor of the capitalist's authority over the worker."[88]

Andrew Morriss, on the other hand, argues that the spread of at-will employment was unrelated to industrialization.[89] He compares the dates that

77. *See* Jacoby, *supra* note 1, at 103–08.

78. *See id.*

79. *See id.* at 120–22.

80. *Id.* at 121–22.

81. *Id.* at 122, 126 (quoting Adair v. United States, 208 U.S. 161, 180 (1908), *overruled in part by* Phelps Dodge Corp. v. NLRB, 313 U.S. 177, 186–87 (1941)); *see also, e.g.*, *Adair*, 208 U.S. at 172–76 (concluding that the Fifth Amendment prohibited Congress from interfering with employers' freedom of contract).

82. Jacoby, *supra* note 1, at 126.

83. Feinman, *supra* note 1, at 130–31.

84. *Id.* at 131.

85. *Id.* at 131–32.

86. *Id.* at 133.

87. *Id.*

88. *Id.* at 132–33.

89. Morriss, *supra* note 1, at 681.

various states formally adopted the at-will rule[90] with state-by-state patterns of industrialization[91] and concludes that the correlation was both negative and not particularly strong.[92] For example, the first states to adopt the rule through the common law were Maine (1851) and Mississippi (1858),[93] while Pennsylvania (1891) and New York (1895)—which together accounted for approximately one-fifth to one-third of the national economy—were relative latecomers.[94] The pattern of adoption was from the underindustrialized West and South to the industrialized East.[95] Morriss therefore proposes a fourth explanation for the spread of the at-will rule. Noting that the rule was adopted sooner in states where supreme court judges were popularly elected than in states with appointed judges,[96] Morriss concludes that courts adopted the rule in an effort to keep difficult employment contract disputes out of the courts.[97]

Morriss's conclusion, however, does not correspond with his evidence. He may be correct that employment cases were difficult for judges and juries to evaluate, and that judges in the South and West, where judges were few and distances large, would have been particularly attracted to a legal rule that effectively kept these cases out of the courts.[98] But many types of cases present difficult issues of proof; Morriss does not explain why employment cases were singled out for a particularly restrictive legal rule. More importantly, Morriss provides no link between his observation that the at-will rule was adopted sooner in southern and western states with popularly elected judges and his conclusion that judges desired to keep employment cases out of the courts. If anything, his conclusion is counterintuitive. One would expect that judges elected by a population "influenced by Populist and Granger rhetoric against large employers"[99] would be less, not more, likely to adopt the decidedly pro-employer rule of at-will employment.

IV. AT-WILL EMPLOYMENT AND THE RACE TO THE BOTTOM OF EMPLOYMENT STANDARDS

Morriss's observation that early adoption of the at-will employment rule was negatively correlated to the degree of industrialization is valid, but his conclusion that this disproves any linkage between the two is unsatisfactory. Instead, it was precisely *because* the underindustrialized states were underindustrialized that they were among the first to adopt the at-will rule.

---

90. *Id.* at 700 tbl.II.
91. *Id.* at 724–36.
92. *Id.* at 736.
93. *See id.* at 704.
94. *Id.* at 703.
95. *Id.* at 703.
96. *Id.* at 745.
97. *Id.* at 753.
98. *See id.*
99. *Id.*

Industrialized states then joined the underindustrialized states in an interjurisdictional race to the bottom of employment standards.

The concept that interjurisdictional competition can create a race to the bottom in law was pioneered by William Cary, who observed that various states, particularly Delaware, were competing for corporate charters by reducing shareholder protection.[100] In recent years, legal scholars have observed similar phenomena in other areas of law such as trust law,[101] property law,[102] "banking law, environmental law, income taxation, local-government law, bankruptcy, and family law."[103] Similarly, the adoption of employment at-will was a race to the bottom in employment law.

A. The Race to the Bottom in At-Will Employment

By the late 1800s, industrialization and advances in transportation had transformed the American economy from local to national.[104] Capital, however, was still heavily concentrated along the East Coast, particularly in New York and Pennsylvania.[105] Western and southern states, by contrast, remained largely agrarian.[106] A major challenge of these states, therefore, was attracting capital investment and creating jobs.

One way that underindustrialized states may have attempted to attract capital and foster job creation was by offering employment rules that would

---

100. William L. Cary, *Federalism and Corporate Law: Reflections Upon Delaware*, 83 YALE L.J. 663, 663, 705 (1974); *cf.* Ralph K. Winter, Jr., *State Law, Shareholder Protection, and the Theory of the Corporation*, 6 J. LEGAL STUD. 251, 254–62 (1977) (describing this race to the bottom and its effect on the capital market).

101. Stewart E. Sterk, *Asset Protection Trusts: Trust Law's Race to the Bottom?*, 85 CORNELL L. REV. 1035, 1039 (2000).

102. *See generally, e.g.*, John V. Orth, *"The Race to the Bottom": Competition in the Law of Property*, 9 GREEN BAG 47 (2005) (discussing this phenomenon).

103. Sterk, *supra* note 101, at 1038–39. *See generally, e.g.*, Vicki Been, *"Exit" as a Constraint on Land Use Exactions: Rethinking the Unconstitutional Conditions Doctrine*, 91 COLUM. L. REV. 473 (1991) (local government); Jennifer Gerarda Brown, *Competitive Federalism and the Legislative Incentives to Recognize Same-Sex Marriage*, 68 S. CAL. L. REV. 745 (1995) (family law); Henry N. Butler & Jonathan R. Macey, *The Myth of Competition in the Dual Banking System*, 73 CORNELL L. REV. 677 (1988) (banking law); Louis Kaplow, *Fiscal Federalism and the Deductibility of State and Local Taxes Under the Federal Income Tax*, 82 VA. L. REV. 413 (1996) (income taxation); Richard L. Revesz, *Rehabilitating Interstate Competition: Rethinking the "Race-to-the-Bottom" Rationale for Federal Environmental Regulation*, 67 N.Y.U. L. REV. 1210 (1992) (environmental law); David A. Skeel, Jr., *Rethinking the Line Between Corporate Law and Corporate Bankruptcy*, 72 TEX. L. REV. 471 (1994) (bankruptcy law).

104. *See* HAROLD G. VATTER, THE DRIVE TO INDUSTRIAL MATURITY: THE U.S. ECONOMY, 1860–1914, at 61 (1975).

105. *See* Morriss, *supra* note 1, at 703.

106. *See* VATTER, *supra* note 104, at 89; *see also* Morriss, *supra* note 1, at 701 (noting that the Northeast was more heavily industrialized than the West and South during the late 1800s).

have been attractive to capitalists. Other things being equal, capitalists would have preferred to build manufacturing facilities in the Northeast, where they could take advantage of proximity to product markets, access to international shipping lanes, and a larger workforce (fed by immigration) that was more accustomed to factory work than were their rural contemporaries. The advent of at-will employment in southern and western states might have served as an inducement for capital investment in those states. This inducement would have seemed particularly attractive because it came just as manufacturers were beginning to wrest control over the production process from artisanal craft unions, leading to significant labor unrest in the Northeast.[107]

This argument is strengthened by the fact that the South and West at this time actively sought to attract capital investment from the Northeast. In the South, for example, agricultural employment declined significantly during this period; southern leaders sought to offset this decline by attracting cotton textile factories and promoting logging, lumber production, and mining.[108] Similarly, the West and Midwest experienced a growing interest in manufacturing as a way to offset a slow decline in the relative importance of farming and to take advantage of the availability of raw materials, water power, and a seasonable climate.[109]

It would be a stretch to claim that Maine or Mississippi, two of the earliest adopters of the at-will rule,[110] adopted the rule specifically to attract capital investment. Indeed, the first courts to adopt the at-will rule may have done so by mere happenstance; no one can definitively explain why the rule was first adopted. The main point here is that once the rule was adopted in a handful of underindustrialized states, other such states would have felt economic pressure to follow suit to avoid being left behind in attracting capital. The industrialized states would then have felt similar pressure to adopt the rule to maintain their competitive advantage in the labor market.

B. Contemporaneous Labor Market Races to the Bottom

Contemporaneous attempts to regulate child labor, to set a minimum wage, and to create unemployment insurance programs illuminate the market forces that created this economic pressure.

By 1916, when the last of the states were formally adopting the at-will rule,[111] some progressive states had also passed restrictive child labor legislation.[112] Though the Supreme Court had upheld state legislation

107. *See* KATHERINE V.W. STONE, FROM WIDGETS TO DIGITS: EMPLOYMENT REGULATION FOR THE CHANGING WORKPLACE 24–25 (2004).

108. VATTER, *supra* note 104, at 89–90.

109. *See id.* at 99, 102–05.

110. Morriss, *supra* note 1, at 704.

111. *Id.* at 700 tbl.II.

112. *See* Hammer v. Dagenhart, 247 U.S. 251, 273 (1918), *overruled by* United States v. Darby, 312 U.S. 100 (1941). I thank my colleague, John Bickers, for his insights on this case.

protecting women and children from the Lochnerian labor marketplace,[113] these states were nevertheless at an economic disadvantage compared to states that permitted child labor.[114] Because the dormant commerce clause prohibited states from erecting barriers to interstate commerce,[115] the states with restrictive standards could not stop products made with child labor at the border. Further, Congress could not directly regulate child labor at the national level on account of the Supreme Court's then-restrictive interpretation of Congressional power under the interstate commerce clause.[116] Thus, Congress attempted to level the playing field among the states in 1916 by prohibiting, for one month, the shipment in interstate commerce of goods created by child labor.[117] The idea behind this restriction was that increasing the storage costs of the goods would offset their competitive advantage, thereby protecting states with restrictive child labor legislation from competitive pressure by the other states.[118]

When the federal statute came before the Supreme Court in *Hammer v. Dagenhart*, the Court explicitly recognized Congress's intent to ameliorate the interstate competitive pressure that was retarding the progress of meaningful child labor laws: "[T]hat the unfair competition, thus engendered, . . . be controlled by closing the channels of interstate commerce to manufacturers in those [s]tates where the local laws do not meet what Congress deems to be the more just standard of other [s]tates."[119]

Nonetheless, the Court struck down the statute as unconstitutional.[120] This early effort at national child labor legislation illustrates vividly the competitive labor market pressure pushing states toward a race to the bottom in employment standards—a pressure that was not relieved until Congress passed (and the Supreme Court upheld[121]) the child labor provisions in the Fair Labor Standards Act of 1938 (FLSA).[122]

---

113. *See, e.g.*, Muller v. Oregon, 208 U.S. 412, 421 (1908) (upholding protective legislation for women).

114. *See Hammer*, 247 U.S. at 273.

115. *See* Gibbons v. Ogden, 9 U.S. 1, 199–200 (1824) ("[W]hen a State proceeds to regulate commerce . . . among the several States, it is exercising the very power that is granted to Congress, and is doing the very thing which Congress is authorized to do.").

116. *See generally* United States v. E.C. Knight Co., 156 U.S. 1 (1895) (discussing the commerce power); ERWIN CHEMERINSKY, CONSTITUTIONAL LAW: PRINCIPLES AND POLICIES 243 (2d ed. 2002) ("Between the late nineteenth century and 1937, the Court was controlled by conservative Justices deeply committed to laissez-faire economics and strongly opposed to government economic regulations. Many federal laws were invalidated as exceeding the scope of Congress's commerce power . . . .").

117. Act of Sept. 1, 1916, Pub. L. No. 64-249, 39 Stat. 675.

118. *See Hammer*, 247 U.S. at 273.

119. *Id.*

120. *Id.* at 277.

121. *See* United States v. Darby, 312 U.S. 100, 125 (1941) (upholding the entire Act in the context of a dispute over the wage and hour provisions).

122. Pub. L. No. 75-718, 52 Stat. 1060 (codified as amended at 29 U.S.C. §§ 201–219 (2000)).

A second example of labor market forces pressuring states toward minimalist employment standards is unemployment insurance. Early state attempts to create such a program failed because states were afraid that enacting a payroll tax would put their employers at a competitive disadvantage vis-à-vis employers in other states.[123] Only one state, Wisconsin, was able to enact an unemployment insurance statute[124] prior to passage of the 1935 federal Social Security Act (SSA).[125] Unsuccessful state efforts to create unemployment insurance programs prior to 1935 do not represent a race to the bottom, as those states were already at the bottom, but they do illustrate how labor market forces tend to impede any effort to reverse course and create a "race to the top."

A third example of labor market pressure, this time coming a few decades after most states had formally adopted the at-will rule, is the set of minimum wage provisions of the FLSA.[126] The FLSA was most strongly opposed by legislators in southern states who believed it would eliminate the competitive advantage those states enjoyed in the form of low wage rates.[127] Likewise, the FLSA was most strongly supported by legislators in northern states who believed the statute would slow the rapid flow of capital and jobs from, for example, northern textile mills to southern textile mills, where labor was cheaper.[128] The FLSA therefore represented an effort by the industrialized states to avert a race to the bottom in wage rates by federally fixing a rate that was relatively high by southern standards.

C. Reacting to a Race to the Bottom

In a national economy, no state wants to put its employers at a competitive disadvantage by adopting a major new employment law that is radically different from, and more costly than, the norm. Both the FLSA and the SSA represent one way that states "on the top" can react to competing states that are

---

123. *See Economic Security Act: Hearings Before the Comm. On Ways and Means*, H.R., 74th Cong. 32 (1935) (Report of the Comm. on Economic Security) [hereinafter *ESA Hearings*]; Willborn et al., *supra* note 34, at 617.

124. *ESA Hearings*, *supra* note 123, at 27; WILLBORN ET AL., *supra* note 34, at 617.

125. Social Security Act, Pub. L. No. 74-271, 49 Stat. 620 (1935) (codified as amended at 42 U.S.C. §§ 301–1397jj (2000)).

126. 29 U.S.C § 206 (2000).

127. *See* Robert K. Fleck, *Democratic Opposition to the Fair Labor Standards Act of 1938*, 62 J. ECON. HIST. 25, 26 (2002); Andrew J. Seltzer, *The Political Economy of the Fair Labor Standards Act of 1938*, 103 J. POL. ECON. 1302, 1315 (1995); *see also* GAVIN WRIGHT, OLD SOUTH, NEW SOUTH: REVOLUTIONS IN THE SOUTHERN ECONOMY SINCE THE CIVIL WAR 224–25 (1986) (suggesting that "the South succumbed to the 'Yankee plot' to impose northern wages" as a means to eliminate black jobs).

128. *See* Fleck, *supra* note 127, at 26; Seltzer, *supra* note 127, at 1315 ("[T]here is strong evidence that northern legislators sought to impose national wage standards in order to prevent the flow of capital to the South."); *see also* Henry C. Simons, *Some Reflections on Syndicalism*, 52 J. POL. ECON. 1, 10–11 (1944) (describing this as the motivation for northern workers' and employers' lobbying efforts).

"at the bottom" or, having significantly changed their laws, are threatening a race to the bottom. They can overcome market forces and stay on top by using a federal statute to impose new labor market terms on a national scale. This pattern has repeated several times over in the adoption of major workplace legislation such as the prohibition of discrimination[129] and the regulation of workplace safety.[130]

The race to the bottom sparked by the early adopters of employment at-will represents a second approach states can take to an impending race to the bottom: "If you can't change 'em, join 'em." While the at-will rule was spreading, *Hammer*,[131] *Lochner*,[132] and their progeny would have made a federal statute regulating the term of employment unthinkable. States that had not yet adopted the at-will rule thus had either to adopt the at-will rule or to put their employers at a competitive disadvantage in the labor market.

States facing an impending race to the bottom of employment standards also have a third possible approach—offset the competitive disadvantage by creating a competitive advantage elsewhere in the labor market. New York, for example, might have ceded the southern flight of textile jobs, but attempted to offset that loss by creating a highly educated workforce and concentrating on the creation of jobs in the financial sector.

In the meantime, an explanation must be proffered for the "outlier" states. If labor market forces create a race to the bottom in employment standards, why was Wisconsin able to pass an unemployment statute in 1932? What explains the contract and tort inroads in employment at-will that have occurred in the last fifty years? And how is it that Montana deviates from the at-will-rule norm?

The answer to these questions, in part, is that only *major* changes in law will shift the competitive landscape sufficiently to create a race to the bottom. Minor tinkering at the margins—such as, arguably, the contract and tort inroads[133]—is not likely to put one state at a significant competitive advantage or disadvantage. The same is true of Montana's statute, which was designed less as an employee windfall and more as a mechanism for relieving employers of high damage awards.[134] The Montana statute gives employees just-cause

---

129. *See, e.g.*, 42 U.S.C. §§ 2000e-2 to 2000e-3 (2000).

130. *See* Occupational Safety and Health Act of 1970, 29 U.S.C. §§ 651–78 (2000).

131. Hammer v. Dagenhart, 247 U.S. 251 (1918).

132. Lochner v. New York, 198 U.S. 45 (1905).

133. One could certainly argue, of course, that the contract and tort inroads were, and are, far more than "minor tinkering." Nonetheless, because all states adopted contract and tort inroads to some degree, and did so incrementally over a period of years, these changes likely would not have affected the labor market to the same degree as an abrupt change in the standard of baseline employment or the imposition of a significant payroll tax.

134. Libenson, *supra* note 63, at 130–31. State workers' compensation statutes similarly were designed in part to relieve employers of high tort damage awards. Richard A. Epstein, *The Historical Origins and Economic Structure of Workers' Compensation Law*, 16 GA. L. REV. 775, 800 (1982).

employment only after a period of probationary at-will employment;[135] recent decisions concerning the statute even suggest that employers may be able to extend the probationary period indefinitely.[136] Similarly, Wisconsin's unemployment insurance statute was significantly less onerous for employers than was the later-passed federal unemployment statute, and therefore may not have created the type of overwhelming labor market pressure that precipitates a race to the bottom (or prevents states at the bottom from climbing back up).[137]

On the other hand, Wisconsin may be an example of a fourth way that states can react to race-to-the-bottom labor market forces. It was probably no accident that Wisconsin—home of Robert LaFollette, the most powerful progressive in the country[138]—was the only state to pass an unemployment statute prior to the New Deal. Strong political willpower, therefore, may be an antidote to race-to-the-bottom labor market forces, at least in the short term.[139]

V. RACING TO THE BOTTOM IN A GLOBAL LABOR MARKET

Though the focus of this Essay is on labor market conditions that existed as the United States transitioned from a local to a national economy, the implications resonate today as the United States and other developed countries transition from national to international economies. The developed-country labor markets today are in much the same economic position as the northeastern states were in the late 1800s:[140] Relative to most of the developing world, they have an educated, highly paid workforce.[141] The labor market in developing

---

135. MONT. CODE ANN. § 39-2-904 (2007); Richard A. Lord, *The At-Will Relationship in the 21st Century: A Consideration of Consideration*, 58 BAYLOR L. REV. 707, 712 n.6 (2006).

136. Lord, *supra* note 135, at 713–14 n.6.

137. The Wisconsin statute capped an employer's payroll tax at 2%, Harold M. Groves & Elizabeth Brandeis, *Economic Bases of the Wisconsin Unemployment Reserves Act*, 24 AM. ECON. REV. 38, 44 (1934); George Wheeler & Eleanor Wheeler, *Individual Employer Reserves in Unemployment Insurance*, 43 J. POL. ECON. 246, 250 (1935), while the federal statute capped the tax at 3%, *see ESA Hearings*, *supra* note 123, at 33. Moreover, the Wisconsin statute created employer-financed company reserve plans that allocated the cost of maintaining the unemployed within specific industries, *see* Groves & Brandeis, *supra*, at 38–40; Wheeler & Wheeler, *supra*, at 246–47, whereas the federal statute created a general unemployment fund, *see ESA Hearings*, *supra* note 123, at 32–33.

138. *See* David P. Thelen, *Author's Preface* to ROBERT M. LA FOLLETTE AND THE INSURGENT SPIRIT, at vii, vii (Oscar Handlin ed., 1976).

139. I am indebted to my colleague, John Bickers, for suggesting this argument.

140. The analogy is not a perfect one. For example, capital markets are far more liquid today than they were in the late 1800s, and the common currency among American states made it impossible for southern states to amass capital by artificially deflating their exchange rates, as China is doing today. Nevertheless, the labor market dynamics are similar.

141. *Cf.* David A. Gantz, *The United States and the Expansion of Western Hemisphere Free Trade: Participant or Observer?*, 14 ARIZ. J. INT'L & COMP. L. 381, 393 (1997) (noting the apparent demand in Mexican factories for "high-tech U.S.-source parts" and the concomitant increase in jobs for "well-trained U.S. workers").

countries today is similar to the labor market that existed in southern and western American states in the late 1800s—largely agrarian, with aspirations to industrialization and a surfeit of low-skilled, inexpensive labor.[142] Just as the South and West once obtained a competitive advantage in the national labor market by adopting at-will employment, so developing countries today are obtaining a competitive advantage in the global labor market through low wages and unregulated or underregulated working conditions.[143]

Of course, important differences exist between the nineteenth century United States labor market and the twenty-first century global labor market, such as radically different political, economic, and social systems. Nonetheless, existing wage disparities and increasing global trade disparities set the stage for another possible race to the bottom in employment standards, this time on a global scale. As the American experience with at-will employment, child labor, unemployment insurance, and the minimum wage illustrate, the United States and other developed countries can respond in four non-mutually-exclusive ways.

First, developed countries could attempt to regain their competitive advantage in the global labor market by standardizing employment terms on a global scale, much as northeastern states did by passing a national minimum-wage law.[144] In fact, the United States recently proposed an international trade framework that would standardize many employment terms such as child labor, forced labor, collective bargaining, and discrimination.[145] The framework would require signatories to free trade agreements "to 'adopt, maintain, and enforce' . . . International Labor Organization language on core labor standards."[146] This approach to standardizing employment terms has two obvious limitations. First, it would not standardize all employment terms; it would not, for example, impose an international minimum wage. Second, the terms that are standardized would be binding only on signatory countries.

The second method by which developed countries could respond is the "if you can't change 'em, join 'em" approach. Perhaps this is one explanation for the relaxation—some might say evisceration—in recent decades of federal laws governing workplace safety and union membership, as well as the historically low minimum wage.[147] Or perhaps these trends are merely a product of the current political winds, subject to change with the next election.

---

142. *Cf., e.g., id.* at 383 ("Mexico . . . desired . . . that the NAFTA would serve as a tool for medium- and long-term economic development and industrialization.").

143. *Cf., e.g., id.* at 392–93 (noting that "NAFTA-related job losses" in the U.S., though ultimately negligible, are largely attributable to the move of low-skilled U.S. factory jobs to Mexico).

144. *See supra* text accompanying notes 126–128.

145. *Administration, Democrats Reach Deal on Labor Standards in Free Trade Pacts*, 75 U.S.L.W. 2680, 2680 (2007).

146. *Id.*

147. *See* Stephen Labaton, *Congress Passes Increase in the Minimum Wage*, N.Y. TIMES, May 25, 2007, at A12. In May 2007, more than a decade after the last increase, Congress

Third, developed countries could attempt to offset the high price of labor by adding value. One such attempt would require these countries to substantially increase their investment in the enhancement of worker productivity through education, worker training, and technology.

Fourth, developed countries could demonstrate strong political willpower by keeping their wages high and maintaining safe working conditions despite the countervailing market forces. Perhaps if developed countries set a positive example, the rest of the world will follow, much as the American federal government followed Wisconsin's lead in creating unemployment insurance. This approach, however, may come with a short-term cost as employers in developed countries continue to send jobs overseas to countries where labor costs are lower. The alternative, though, may be to join the developing world in a race to the bottom in employment standards.

VI. CONCLUSION

Scholars have proposed several theories to explain why states rapidly adopted the at-will employment rule after Horace Gay Wood described it in his 1877 treatise. A new theory may also apply: that the spread of the at-will rule can be explained as a race to the bottom in employment standards created by interjurisdictional competition for investment capital and jobs. Today, the United States and other developed countries face a similar scenario, this time on a global scale, as developing countries gain an advantage in the global labor market by offering low wages and unregulated or underregulated working conditions. Developed countries may learn a lesson from America's experience with the at-will rule and use creative lawmaking to halt this race to the bottom of international employment standards.

---

enacted legislation (as part of an Iraq War spending bill) that will increase the national minimum wage by $2.10 over a two-year period. *Id.*; *see* Fair Minimum Wage Act of 2007, 29 U.S.C.A. §§ 201, 206 (Supp. 2007).

ELECTION AS APPOINTMENT: THE TENNESSEE PLAN RECONSIDERED

BRIAN T. FITZPATRICK*

In 1971, the Tennessee legislature followed the lead of a number of other states and replaced the direct election of appellate judges with a selection method called "merit selection."[1] Tennessee's merit selection system—fittingly referred to as the "Tennessee Plan"—calls for the governor to appoint all appellate judges in Tennessee, including state supreme court justices, from a list of three nominees submitted by a commission predominately comprised of lawyers.[2] After a period of time on the bench, the judges appointed by the governor have their names put on the ballot in uncontested retention referenda in which voters are asked whether they wish the judges to remain in office.[3] In light of activity in the most recent legislative session, the Tennessee Plan is now scheduled to expire on June 30, 2009.[4] This is, therefore, an opportune time to consider whether Tennessee should continue to use the Plan to select appellate judges.

The Tennessee Plan has been controversial ever since it was enacted in 1971 to replace contested elections. Many people doubt, for example, whether the Plan has actually accomplished any of its intended purposes. The Plan's principal purposes are to select better qualified judges, to take the politics out of judicial selection, and to bring more racial and gender diversity to the bench.[5] Scholars, however, have found little evidence that any of these purposes are furthered by merit selection plans in general or the Tennessee Plan in particular.[6]

---

* Assistant Professor of Law, Vanderbilt University Law School. J.D., 2000, Harvard Law School. I am grateful to the many people who provided helpful comments on earlier drafts of this paper. I am also indebted to Sybil Dunlop for her helpful research assistance. Research for this paper was supported by a grant from the Federalist Society for Law and Public Policy. The views expressed herein are my own.

1. *See* 1971 Tenn. Pub. Acts, ch. 198.

2. *See* TENN. CODE ANN. §§ 14-4-101, -102 (1994 & Supp. 2007).

3. *See id.* §§ 17-4-114 to -116.

4. *See infra* notes 109–11 and accompanying text.

5. *See* TENN. CODE ANN. § 17-4-101(a) ("It is the declared purpose and intent of the general assembly by the passage of this chapter to assist the governor in finding and appointing the best qualified persons available for service on the appellate courts of Tennessee . . . and . . . to insulate the judges of the courts from political influence and pressure . . . and . . . to make the courts 'nonpolitical.'"); *id.* § 17-4-102(b)(3) (requiring the speakers of the legislature to appoint to the judicial nominating commission "persons who approximate the population of the state with respect to race . . . and gender"); *id.* § 17-4-102(d) (requiring lawyers' organizations to submit nominees for the judicial nominating commission "with a conscious intention of selecting a body which reflects a diverse mixture with respect to race . . . and gender").

6. With respect to the claim that merit selection leads to better qualified judges, scholars

Nonetheless, perhaps the greatest controversy surrounding the Tennessee Plan is whether it is even constitutional. The Tennessee constitution states, as it has since 1853, that all judges "shall be elected by the qualified voters" of the

---

have found that "the credentials of merit selection judges are not superior to nor substantially different from those of other judges." Henry R. Glick & Craig F. Emmert, *Selection Systems and Judicial Characteristics: The Recruitment of State Supreme Court Judges*, 70 JUDICATURE 228, 235 (1987). Moreover, scholars have found that "[j]udges in *more* partisan systems are more productive than judges in less partisan systems [such as merit selection]." Stephen J. Choi, G. Mitu Gulati & Eric A. Posner, *Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather than Appointed Judiciary* 16 (Univ. of Chi. Sch. of Law, John M. Olin Law & Econ. Working Paper No. 357, 2007), *available at* http://ssrn.com/abstract=1008989.

With respect to the claim that merit selection takes the politics out of judicial selection, scholars have concluded that "[o]f course [it does] not." Herbert M. Kritzer, *Law is the Mere Continuation of Politics by Different Means: American Judicial Selection in the Twenty-First Century*, 56 DEPAUL L. REV. 423, 466 (2007). "The politics come into play in determining who actually gets appointed to the commission . . . and in how the commission chooses to weigh various criteria in making both initial nominations and in doing the periodic evaluations." *Id.* In other words, "[t]he system is not nonpolitical; it is simply differently political." *Id.*; *see also* HARRY P. STUMPF, AMERICAN JUDICIAL POLITICS 167 (1988) (noting that "far from taking judicial selection out of politics, [merit selection] actually tend[s] to replace [electoral] [p]olitics, wherein the judge faces popular election . . ., with a somewhat subterranean process of bar and bench politics, in which there is little popular control" and "raw political considerations masquerade[] as professionalism via attorney representation of the socioeconomic interests of their clients"); Harry O. Lawson, *Methods of Judicial Selection*, 75 MICH. BUS. L.J. 20, 24 (1996) ("Merit selection does not take politics out of the judicial selection process. It merely changes the nature of the political process involved. It substitutes bar and elitist politics for those of the electorate as a whole.").

Finally, with respect to the claim that merit selection leads to a more diverse bench, nationwide studies have proved inconclusive. *See* Sherrilyn A. Ifill, *Through the Lens of Diversity: The Fight for Judicial Elections After* Republican Party of Minnesota v. White, 10 MICH. J. RACE & L. 55, 85 (2004) ("Studies that have examined the effect of appointment versus election of judges on diversity have produced conflicting results."). In Tennessee, the evidence likewise is conflicting. In 2007, appellate judges in Tennessee—those selected by the Tennessee Plan—were more diverse in both race and gender than were trial judges. *See* Diversity of the Bench, American Judicature Society, http://www.judicialselection.us/judicial_selection/bench_diversity/index.cfm?state= (providing data on race and gender of judges for each state). The opposite was true in both 2004 and 2001. *See* American Bar Association, National Database on Judicial Diversity in State Courts, http://www.abanet.org/judind/diversity/tennessee.html (reporting data for 2004 on racial diversity only); American Judicature Society, Judicial Selection in the States, http://www.ajs.org/JSremoved 3.3.08/js/TN_diversity.htm (citing AMERICAN BAR ASSOCIATION, THE DIRECTORY OF MINORITY JUDGES OF THE UNITED STATES (3d ed. 2001) (reporting data for race and gender in 2001)). Moreover, complaints by the governor that the Tennessee Plan was not producing sufficient racial diversity led him to sue the judicial nominating commission in a case that eventually reached the Tennessee Supreme Court. *See* Bredesen v. Tenn. Judicial Selection Comm'n, 214 S.W.3d 419 (Tenn. 2007).

state.[7] The Tennessee Plan, relying as it does on initial appointment by the governor and retention in an uncontested referendum, would seem to be in some tension with that language. Given this tension, it is not surprising that the Tennessee Plan has been mired in litigation ever since its inception, with several cases challenging the constitutionality of the Plan heard by the Tennessee Supreme Court.[8]

In this Essay, I examine the constitutional questions surrounding the Tennessee Plan. Although the Tennessee Supreme Court has upheld the constitutionality of the Tennessee Plan on two occasions—once in 1973[9] and again in 1996[10]—in neither case did the court's decision command a majority of regular supreme court justices,[11] and, in the 1996 case, the opinion was not published and does not constitute binding precedent.[12] Moreover, and more importantly, neither of these decisions even attempted to address three of the most serious constitutional questions raised by the Plan. As I explain, these questions are not easily answered, and, in my view, suggest that the Tennessee Plan is unconstitutional in many of its applications.

The first question that the Tennessee Supreme Court has never addressed is how the constitution permits the governor to appoint all appellate judges in the first place. Although a provision of the constitution permits the governor to appoint judges to fill "vacancies," it appears that the constitution uses the word "vacancies" to refer only to interim vacancies—i.e., where judges leave in the middle of their terms—rather than to positions that are vacant simply because judges choose not to run for reelection.[13] It would seem, then, that, to the extent the Tennessee Plan permits the governor to make appointments to fill vacancies created by judges who leave office at the end of their terms, the Plan is unconstitutional.[14] This issue has never come before the Tennessee Supreme Court because both the 1973 and 1996 cases involved judges appointed to fill interim vacancies.[15]

The second question that the Tennessee Supreme Court has never addressed is how retention referenda can be squared with the original understanding and purposes of the constitutional requirement of an "election." In 1870, when the current constitutional provision was enacted, the idea of retention referenda for public officials was unknown in the United States.[16] As

---

7. TENN. CONST. art. VI, §§ 3, 4.

8. *See, e.g.*, State *ex rel.* Hooker v. Thompson, No. 01S01-9605-CH-00106, 1996 WL 570090 (Tenn. Oct. 2, 1996); State *ex rel.* Higgins v. Dunn, 496 S.W.2d 480 (Tenn. 1973).

9. *See Dunn*, 496 S.W.2d 480.

10. *See Thompson*, 1996 WL 570090.

11. *See infra* notes 150–53 and accompanying text.

12. *See infra* note 143 and accompanying text.

13. *See infra* text accompanying notes 162–68.

14. *See infra* text accompanying notes 161–67.

15. *See infra* text accompanying note 167.

16. *See infra* text accompanying notes 182–183. The idea was first conceived in 1914. *See id.*

such, it would have been impossible for the authors of this provision to have intended such a device when they used the word "election." Although many scholars believe it should not be necessary to amend the constitution to permit the legislature to take advantage of every new way of doing things, it is doubtful whether retention referenda even serve the democratic purposes of the 1870 constitution.[17] As an historical matter, retention referenda were originally designed not to facilitate democratic accountability, but, rather, to insulate judges from such accountability.[18] It is therefore not surprising that, in Tennessee and elsewhere, judges who run in retention referenda are virtually never defeated.[19]

Finally, the Tennessee Supreme Court has never explained how the Tennessee Plan can be constitutional in light of the fact that, in 1977, the voters in Tennessee *rejected* a constitutional amendment that would have repealed the constitutional provision requiring elected judges in favor of provisions permitting the Tennessee Plan.[20] This is in stark contrast to each of the sixteen other states that select judges through a method of initial appointment by the governor followed by a retention referendum; each of these states has amended its constitution to change provisions requiring elected judges in favor of provisions permitting the appointment-retention method of selection.[21]

None of these questions is easily answered, and, together, they comprise a compelling case for the view that many appellate judges in Tennessee have been selected in an unconstitutional manner for the better part of four decades. For this reason, I argue that the Tennessee legislature should allow the Tennessee Plan to expire next year and, in doing so, return the selection of appellate judges to contested elections.[22] In my view, the legislature should employ contested elections to select all judges in the state at least until the voters of Tennessee have been given another opportunity to amend the Tennessee constitution—and perhaps beyond that point if the voters again reject the amendment.

In Part I of this Essay, I briefly recount the history of judicial selection in Tennessee. Like most states that entered the Union in the founding era, Tennessee originally appointed all of its judges, but then switched to elections as the populism of home-grown "Jacksonian Democracy" spread across America.[23] Tennessee only turned to merit selection—a Progressive Era reform seeking to place greater control over government in the hands of "experts"—late in the twentieth century.[24] In Part II, I describe the provisions of the Tennessee Plan. Although many of the provisions have been revised over the

17. *See infra* text accompanying notes 184–202.
18. *See id.*
19. *See id.*
20. *See infra* text accompanying notes 204–11.
21. *See infra* notes 212–13 and accompanying text.
22. *See infra* notes 109–11 and accompanying text.
23. *See infra* text accompanying notes 27–39.
24. *See infra* text accompanying notes 52–70.

years, the core of the Tennessee Plan—appointment by the governor from a list of names supplied by a lawyer-dominated commission followed by a retention referendum—remains the same.[25] In Part III, I recount the litigation over the constitutionality of the Tennessee Plan, including the two Tennessee Supreme Court decisions upholding it.[26] In Part IV, I explain why these two decisions have left a series of important constitutional questions unanswered. In Part V, I conclude that the constitutional case against the Tennessee Plan is strong, and I argue that the legislature should allow it to expire and return the state to contested elections until Tennessee voters decide to amend the state constitution.

I. THE HISTORY OF JUDICIAL SELECTION IN TENNESSEE

At the time of the founding, judges throughout the new United States came to the bench either by executive or legislative appointment, and they often held their positions for life.[27] The first Tennessee constitution, ratified in 1796 when Tennessee became the nation's sixteenth state, granted judges life tenure (so long as they exhibited "good behavior") and placed the power to select those judges exclusively in the hands of the state legislature.[28]

While the federal judicial system has stayed the same over the ensuing two hundred years, the state judicial systems changed radically in the first half of the nineteenth century.[29] By the time of the Civil War, the vast majority of states had changed their method of judicial selection from executive or legislative appointment to direct election by the people.[30]

In some ways, this dramatic shift in the states was a phenomenon with its roots in Tennessee: Most historians attribute the change in judicial selection to a shift in this country's attitude about democracy that was inspired by Tennessean Andrew Jackson. At the time of the founding, democracy was an ideal embraced only tentatively by the political elite.[31] It was not until the nineteenth century, during the populist movement led by Andrew Jackson, that the country began to emphatically embrace the notion that ordinary citizens

---

25. *See infra* text accompanying notes 76–104.

26. *See infra* text accompanying notes 115–53.

27. *See* EVAN HAYNES, THE SELECTION AND TENURE OF JUDGES 98 (1944); Larry C. Berkson, *Judicial Selection in the United States: A Special Report*, 64 JUDICATURE 176, 176 (1980). Before it became a state, Vermont briefly selected judges by election. *See* HAYNES, *supra*, at 99.

28. *See* TENN. CONST. art. V, § 2 (1976) ("The general assembly shall by joint ballot of both houses appoint judges of the several courts of law and equity . . . who shall hold their respective offices during their good behavior.").

29. *See* Berkson, *supra* note 27, at 176.

30. *See id.* ("By the time of the Civil War, 24 of 34 states had established an elected judiciary with seven states adopting the system in 1850 alone.").

31. *See* ALEXANDER KEYSSAR, THE RIGHT TO VOTE: THE CONTESTED HISTORY OF DEMOCRACY IN THE UNITED STATES 21–25 (2000) (explaining that the American Revolution "produced modest, but only modest, gains in the formal democratization of politics").

were fully capable of making decisions about their government.[32] According to historians, this populist movement (dubbed "Jacksonian Democracy") that restructured so many American institutions was also responsible for the tide of elected judiciaries that washed across America in the middle of the nineteenth century.[33]

The tide did not begin in Tennessee, however. The first state to select any of its judges by election was Georgia in 1812,[34] and the first state to select all of them in that manner was Mississippi in 1832.[35] Tennessee first considered proposals to select its judges by election at its second Constitutional Convention in 1834.[36] The proposals failed to pass, and the Convention ultimately voted to continue selecting judges by legislative appointment.[37] Tennessee did not make the change to an elected judiciary until 1853.[38] In that year, the people of Tennessee approved a constitutional amendment providing that all judges in the state "shall be elected by the qualified voters" to terms of eight years.[39]

After the Civil War, Tennessee held another Constitutional Convention to bring its constitution into compliance with the requirements demanded by the federal Reconstruction Congress.[40] The Convention of 1870 maintained the provision requiring the election of all judges,[41] and the 1870 language has not been changed since then.[42] Thus, the Tennessee constitution still declares that

---

32. *See id.* at 33–42, 74 (outlining the expansion of suffrage).

33. *See* Republican Party of Minn. v. White, 536 U.S. 765, 785 (2002) ("Starting with Georgia in 1812, States began to provide for judicial election, a development rapidly accelerated by Jacksonian democracy."); CHARLES H. SHELDON & LINDA S. MAULE, CHOOSING JUSTICE: THE RECRUITMENT OF STATE AND FEDERAL JUDGES 4 (1997) (noting that "the Jacksonian movement . . . encouraged more popular control of judges"); Berkson, *supra* note 27, at 176 (noting that people "were determined to end [the] privilege of the upper class and to ensure the popular sovereignty we describe as Jacksonian Democracy"); Rachel Paine Caufield, *How the Pickers Pick: Finding A Set of Best Practices for Judicial Nominating Commissions*, 34 FORDHAM URB. L.J. 163, 167 (2007) ("States began to move away from appointive selection methods in the mid-1800s with the rise of Jacksonian democracy and its emphasis on democratic accountability, individual equality, and direct voter participation in governmental decision-making.").

34. *See* HAYNES, *supra* note 27, at 99–100.

35. *See id.*

36. *See* Timothy S. Huebner, *Judicial Independence in an Age of Democracy, Sectionalism, and War*, *in* A HISTORY OF THE TENNESSEE SUPREME COURT 66 (James W. Ely Jr. ed., 2002); N. Houston Parks, *Judicial Selection—The Tennessee Experience*, 7 MEM. ST. U. L. REV. 615, 624 (1977).

37. *See id.* The 1834 Convention did, however, eliminate life tenure for judges in favor of twelve- and eight-year terms. *See id.*

38. *See* Huebner, *supra* note 36, at 87; Parks, *supra* note 36, at 626–28.

39. *Id.*

40. *See* Parks, *supra* note 36, at 630–31.

41. *See id.*

42. *See id.*

all judges—whether on the "supreme court" or "inferior courts"—"shall be elected by the qualified voters" of the state to a term of eight years.[43]

Throughout the next one hundred years, judges in Tennessee were selected, at least in theory, by voters in contested elections similar to those held for other public offices.[44] But the actual practice of judicial elections did not necessarily comport with the theory. As one commentator has noted, "[e]lection campaigns generally were not very partisan. In fact, incumbent judges usually ran with no, or only nominal opposition."[45] (After all, for much of the post-Civil War era, Tennessee was a one-party state; thus, whichever candidate was nominated by the Democratic Party was all but certain to win a judgeship.[46]) Moreover, most judges in Tennessee were elevated to the bench after 1853 not by election, but by gubernatorial appointment to fill interim vacancies.[47] Since 1834, the Tennessee constitution has permitted the legislature to direct how such interim vacancies should be filled,[48] and, from the very beginning, the legislature vested that power with the governor.[49] Consequently, one commentator has reported that, in the first one hundred years of judicial elections in Tennessee, "nearly 60 percent of the regular judges who . . . served on [the] Supreme Court [were] appointed by the Governor in the first instance."[50] This conflict between the theory and reality of judicial elections was not a phenomenon unique to Tennessee; many judges in states with elected judiciaries also were elevated by appointment to unexpired terms.[51]

---

43. TENN. CONST. art. VI, §§ 3, 4.

44. *See* Parks, *supra* note 36, at 628–29.

45. *Id.* at 629.

46. *See id.* at 630 ("Since after the Civil War [Tennessee] was generally controlled by the Democratic Party, nomination by the Democrats to a seat on the bench was tantamount to election.").

47. *See id.* at 629 ("[T]hose elected most often had reached the bench initially though gubernatorial appointment.").

48. The 1796 constitution made no provision for the filling of vacancies. Thus, all vacancies had to be filled by the manner set forth for initial appointment, i.e., by appointment of both houses of the legislature. *See* Smith v. Normant, 13 Tenn. 271, 272–73 (Tenn. 1833) (holding that, in the case of vacancies, the "constitution has made no exception in favor of the legislature giving authority by law to an agent to appoint judges" and "[t]he two houses acting jointly, and voting by ballot, is the only appointing power under the constitution"). By 1834, the constitution permitted the legislature to prescribe the manner of filling vacancies that arose by reason of "death, resignation, or removal." TENN. CONST. art. VII, § 4 (1834). The current (1870) constitution likewise permits the legislature to prescribe the manner of filling all vacancies. *See id.* ("[The] filling of all vacancies not otherwise directed or provided by this Constitution, shall be made in such manner as the Legislature shall direct.").

49. *See* Parks, *supra* note 36, at 629.

50. *Id.* at 629 (quoting William H. Wicker, *Constitutional Revision and the Courts*, *in* PROCEEDINGS OF THE SIXTH ANNUAL SOUTHERN INSTITUTE OF LOCAL GOVERNMENT 12, 14 (Bureau of Public Administration, University of Tennessee - Knoxville, 1947)).

51. *See* SUSAN B. CARBON & LARRY C. BERKSON, JUDICIAL RETENTION ELECTIONS IN THE UNITED STATES 14 (1980) (noting that, "in the 30 states which employ partisan and nonpartisan

Despite the states' limited experience with contested judicial elections—or perhaps because of it—the trend in favor of elected judiciaries began to wane in America in the early twentieth century. During the Progressive Era, professional lawyers' organizations across the country began to advocate for a new method of judicial selection.[52] The new method was intended to take selection out of the political process, whether that process was political appointment or popular election.[53] The proponents of this new method believed that judges should be selected by "experts";[54] in particular, they thought that the lawyers' organizations themselves should make the selections. These organizations called the method whereby they would select judges "merit selection." In 1937, the nation's largest organization of lawyers, the American Bar Association, formally endorsed merit selection plans,[55] and in 1940, the state of Missouri became the first of many states to change its method of judicial selection from popular election to merit selection.[56] With the heavy support of lawyers' organizations in the state,[57] Tennessee first adopted a merit selection plan in 1971.[58]

The merit selection plans adopted by these states did not turn judicial selection entirely over to local lawyers' organizations. Rather, the plans typically charged the state's governor with appointing judges from a list of names submitted by a nominating commission comprised largely of members of local lawyers' organizations.[59] Moreover, although many of the architects of merit selection favored life tenure for judges appointed in this manner, they suspected the public would balk at being entirely excluded from a role in choosing such important public officials.[60] Thus, the architects of merit selection designed a mechanism that they thought would result in life tenure but without the appearance of life tenure: the retention referendum. In a retention referendum, a judge runs unopposed and the electorate is simply asked whether the judge should remain on the bench.[61] That is, the public votes on retention

---

elections to fill most of their judiciaries, a substantial number of judges actually reach the bench by appointment").

52. *See id.* at 3–6.

53. *See id.*

54. *See, e.g.*, Luke Bierman, *Judicial Independence: Beyond Merit Selection*, 29 FORDHAM URB. L.J. 851, 854 (2002) (noting that the reform movement in the Progressive Era was based on the hope that "experts, rather than voters, would be responsible for selecting judges").

55. *See* CARBON & BERKSON, *supra* note 51, at 4.

56. *See id.* at 11.

57. *See* John R. Vile, *The Tennessee Supreme Court, 1946–1974: Tranquility Amid a National Judicial Revolution*, *in* A HISTORY OF THE TENNESSEE SUPREME COURT 268 (James W. Ely Jr. ed., 2002).

58. *See* Parks, *supra* note 36, at 615 & n.1.

59. *See* STUMPF, *supra* note 6, at 163 (describing the chief features of merit-selection plans).

60. *See* CARBON & BERKSON, *supra* note 51, at 6–8.

61. *See id.*

without any knowledge of who might replace the judge if he or she is voted out of office.[62] Under these circumstances, the public nearly always votes in favor of retention.[63] Again, this was not a surprise to the architects of merit selection. As historians have explained, "many proponents of the commission plan would have preferred good behavior tenure in lieu of retention elections"; "[t]hey perceived retention as a 'sop' to those committed to electoral control over the judiciary."[64]

As explained in more detail in Part II of this Essay, the merit selection plan adopted by Tennessee in 1971—fittingly referred to as the "Tennessee Plan"—is much like the plans in other states. Similar to other plans, judges are initially appointed by the governor from a list of names submitted by a judicial nominating commission.[65] These judges must then run in retention referenda some period of time thereafter.[66] The 1971 Tennessee Plan applied to judges on both the intermediate appellate courts and the state supreme court.[67] In 1974, the Plan for the supreme court was repealed,[68] but it was reenacted in 1994.[69] The Plan has never been adopted for the selection of trial judges.[70]

Unlike every other state that has adopted a method of judicial selection that relies on initial appointment followed by a retention referendum,[71] Tennessee has never amended its constitution to replace the provision requiring that all state judges shall be elected.[72] Indeed, not only has the constitution never been amended, but the voters of Tennessee rejected such an amendment in 1977. In that year, a limited Constitutional Convention was called to make several changes to the 1870 constitution.[73] The Convention proposed thirteen different

---

62. *See id.*

63. *See* Larry Aspin, *Trends in Judicial Retention Elections, 1964–1998*, 83 JUDICATURE 79, 79 & n.1 (1999) (finding that in 4,588 retention referenda in a sample of ten states over thirty-four years, only fifty-two judges were not retained).

64. CARBON & BERKSON, *supra* note 51, at 6–8. Although Missouri was the first state to adopt a merit selection plan, California was the first state to use retention referenda. *See id.* at 11. California began using the referenda in 1934 when they were proposed by a group of citizens that included Earl Warren, who would eventually become Chief Justice of the United States. *See* Gerald F. Uelmen, *Supreme Court Retention Elections in California*, 28 SANTA CLARA L. REV. 333, 339 (1988).

65. *See* TENN. CODE ANN. § 17-4-112(a) (1994 & Supp. 2007).

66. *See id.* §§17-4-114 to -116.

67. *See* 1971 Tenn. Pub. Acts, ch. 198.

68. *See* 1974 Tenn. Pub. Acts, ch. 433, § 1.

69. *See generally* 1994 Tenn. Pub. Acts, ch. 942.

70. The one exception is when the governor fills interim vacancies in the trial courts. Since 1994, the governor has been required to fill interim vacancies using the judicial nominating commission. Even so, all trial judges must still run for reelection in contested elections. *See infra* note 85.

71. *See infra* note 213.

72. *See* TENN. CONST. art. VI, §§ 3, 4.

73. *See* LEWIS L. LASKA, THE TENNESSEE STATE CONSTITUTION 23–27 (1990) (outlining the history and proceedings of Tennessee's 1977 limited Constitutional Convention).

amendments to the people of Tennessee on a variety of topics, including one that would have, among other things, replaced the language guaranteeing an elected judiciary with language providing for the Tennessee Plan.[74] The voters approved every one of the thirteen amendments *except* the one that would have replaced the language on elected judges with the provisions of the Tennessee Plan; this amendment failed by a margin of 55% to 45%.[75]

II. THE TENNESSEE PLAN

As originally enacted by the legislature in 1971, the Tennessee Plan called for all "vacancies" on the intermediate appellate courts and supreme court to be filled by the governor.[76] The Plan described "vacancies" not only as interim vacancies—i.e., instances where a judge left in the middle of an eight-year term—but also as instances where the judge *completed* an eight-year term and did not run for reelection.[77] That is, the Tennessee Plan required the governor to initially appoint *all* judges on the intermediate appellate courts and the supreme court.

In making the appointments, the governor was required to select one of three persons submitted by a judicial nominating commission.[78] Under the 1971 legislation, the nominating commission was comprised of nine members: three members of the legislature, three attorneys elected by their peers, and three others appointed by the governor, only one of whom could be a lawyer.[79] The judges appointed by the governor were permitted to serve until the next biennial general election, at which time they would face referenda where voters

---

74. The proposal would have amended Article VI of the Tennessee constitution by deleting Sections 1–15 and substituting language stating, among other things, that "Justices of the Supreme Court and judges of the Court of Appeals shall be appointed by the Governor from three nominees recommended . . . by the Appellate Court Nominating Commission," and that "[t]he name of each justice and judge seeking retention shall be submitted to the qualified voters for retention or rejection . . . ." Governor Ray Blanton, *Proclamation by the Governor*, *in* THE LIMITED CONSTITUTIONAL CONVENTION OF 1977, STATE OF TENNESSEE, THE JOURNAL OF THE DEBATES OF THE LIMITED CONSTITUTIONAL CONVENTION OF 1977 (see Proposal 13, § 4).

75. *See id.* (noting that the amendment received 157,581 votes in favor and 190,421 votes against).

76. *See* TENN. CODE ANN. § 17-712 (1972).

77. *See id.* §§ 17-712, -716.

78. *See id.* § 17-712. As originally enacted, the statute permitted the governor to reject names from the commission indefinitely. *See id.* The statute now permits the governor to reject only one list of three names; the governor is required to select someone from the second list submitted by the commission. *See* TENN. CODE ANN. § 17-4-112 (1994). This requirement was the subject of recent litigation between the governor and the judicial nominating commission that ultimately reached the Tennessee Supreme Court. *See* Bredesen v. Tenn. Judicial Selection Comm'n, 214 S.W.3d 419 (Tenn. 2007) (holding, *inter alia*, that the commission could not include a person on the second list of names sent to the governor if that person had been on the first list as well).

79. *See* TENN. CODE ANN. § 17-702 (1972).

would be asked only: "Shall (Name of Candidate) be elected and retained in office as (name of Office)? Vote Yes or No."[80] If a majority of voters voted to retain the judge, the judge would serve for the remainder of an eight-year term, at which time the judge would face another retention referendum.[81] If the judge was not retained, then the governor would appoint a new judge from a list of three names submitted by the nominating commission.[82]

Much of the 1971 legislation remains intact today, but there have been several important changes to the Tennessee Plan since then. First, in 1974, the legislature amended the Plan to revoke its applicability to vacancies on the supreme court.[83] The legislature would not add the supreme court back until 1994.[84] Thus, for twenty years, the Plan applied only to the intermediate appellate courts. Today, the Tennessee Plan applies to both the intermediate appellate courts and the supreme court. It has never been extended to trial courts.[85]

Second, the legislature has significantly reworked the nominating commission that supplies the list of names from which the governor must appoint judges. In 2001, the nominating commission was expanded to its present size of seventeen members.[86] Although legislators no longer serve on the commission, the two speakers of the legislature select all seventeen members.[87] Fourteen members must be lawyers, leaving only three non-lawyers.[88] Twelve of the fourteen lawyer members must come from names supplied by five special lawyers' organizations.[89] Two members must be taken from names submitted by the Tennessee Bar Association, one from the Tennessee Defense Lawyers Association, three from the Tennessee Trial Lawyers Association, three from the Tennessee District Attorneys General Conference, and three from the Tennessee Association of Criminal Defense Lawyers.[90] The two remaining lawyer members need not be taken from one of these groups.[91] Each lawyers' organization is required to compose these lists "with a conscious intention of selecting a body which reflects a diverse mixture

---

80. *Id.* §§ 17-714 to -716.

81. *See id.*

82. *See id.* §§ 17-714 to -715 (stating that such a situation would create a "vacancy," and that, per § 17-712, the governor would fill that vacancy).

83. *See* 1974 Tenn. Pub. Acts, ch. 433, § 1.

84. *See generally* 1994 Tenn. Pub. Acts, ch. 942.

85. The one exception is interim appointments to fill unexpired terms, which, since 1994, the legislature has required the governor to fill through the judicial nominating commission. *See* TENN. CODE ANN. § 17-4-118(a) (1994). Unlike interim appellate appointments, however, all trial judges must run for reelection in contested elections. *See id.* § 17-4-118(e).

86. *See id.* § 17-4-102(a) (Supp. 2007).

87. *See id.* § 17-4-102(b).

88. *See id.* § 17-4-102(a)(5) (noting that "[t]hree (3) members . . . shall not be lawyers").

89. *See id.* § 17-4-102(a)(1)–(4).

90. *See id.*

91. *See id.* § 17-4-102(a)(6).

with respect to race . . . and gender";[92] the speakers are likewise required to appoint from these lists "persons who approximate the population of the state with respect to race . . . and gender."[93] Each commission member serves a term of six years.[94]

Third, in 1994, the legislature created a new "judicial evaluation commission" to publish an evaluation of all judges before they run in their required retention referenda.[95] If the evaluation commission recommends that the public retain a judge, then the judge runs in a retention referendum.[96] If the commission does not recommend that the public retain a judge, however, then the general election laws apply and the judge runs in a contested, partisan election.[97] Given that judges who run in retention referenda virtually never lose,[98] the evaluation commission can make a big difference as to whether a judge stays on the bench. The evaluation commission is comprised of twelve members, only four of whom are non-lawyers.[99] The members are selected by the speakers and the Tennessee Judicial Council,[100] an advisory body created to advise the legislature on judicial administration.[101] Four of the members must be selected from lists proposed by many of the same special lawyers' organizations that propose names for the judicial nominating commission.[102] As with the nominating commission, those selecting the evaluation commission "shall endeavor to make appointments and submit nominees . . . that approximate the population of the state with respect to race and gender."[103] Evaluation commission members serve six-year terms.[104]

Since the judicial evaluation commission was created in 1994, the commission has evaluated sixty-six judges. In every single one of these sixty-six evaluations, the commission recommended that the judge be retained.[105]

---

92. *Id.* § 17-4-102(d).

93. *Id.* § 17-4-102(b)(3). These statutory provisions, which appear to set forth racial and gender quotas for service on the judicial nominating commission, themselves raise constitutional questions. *See, e.g.*, City of Richmond v. J.A. Croson Co., 488 U.S. 469, 486 (1989) (striking down racial set aside in government contracting); Regents of the Univ. of Cal. v. Bakke, 438 U.S. 265, 320 (1978) (opinion of Powell, J.) (striking down racial quota in medical school admissions).

94. *See* TENN. CODE ANN. § 17-4-106(a) (Supp. 2007).

95. *Id.* § 17-4-201.

96. *See id.* §§ 17-4-114(c), -115(c).

97. *See id.*

98. *See infra* text accompanying notes 107–08, 189–93.

99. *See* TENN. CODE ANN. § 17-4-201(b)(1)–(4) (Supp. 2007).

100. *See id.* § 17-4-201(b)(2)–(4).

101. *See id.* § 16-21-101.

102. *See id.* §§ 17-4-102(a)–(b), -201(b)(3)–(4).

103. *Id.* § 17-4-201(b)(7). These apparent racial and gender quotas may themselves be unconstitutional. *See supra* note 93.

104. *See* TENN. CODE ANN. § 17-4-201(b)(8) (Supp. 2007).

105. *See* TENN. JUDICIAL EVALUATION COMM'N, TENNESSEE APPELLATE JUDGES EVALUATION REPORT (1998) (on file with Tennessee Law Review) (recommending the retention

Since the Tennessee Plan was created in 1971, there have been 146 retention referenda.[106] In 145 of the 146 referenda, the public voted in favor of retention, a retention rate of 99.3%.[107] The only exception was in 1996, when 55% of the public voted against retaining a supreme court justice, Penny White.[108]

The Tennessee legislature permitted the statutes creating the judicial nominating and judicial evaluation commissions to expire on June 30, 2008.[109] Nonetheless, the commissions continue to operate until June 2009 under a provision of the law that allows them to wind down their activities for one year.[110] Unless the legislature acts to save the commissions next year, it appears that the Tennessee Plan will terminate at that time. If the legislature

---

of four supreme court justices, ten court of appeals judges, and twelve court of criminal appeals judges); TENN. JUDICIAL EVALUATION COMM'N, TENNESSEE APPELLATE JUDGES EVALUATION REPORT (2000), *available at* http://www.tsc.state.tn.us/geninfo/JudEval/judeval2000.pdf (last visited May 13, 2008) (recommending the retention of two court of appeals judges and four court of criminal appeals judges); TENN. JUDICIAL EVALUATION COMM'N, TENNESSEE APPELLATE JUDGES EVALUATION REPORT (2004) (on file with Tennessee Law Review) (recommending the retention of two court of appeals judges); TENN. JUDICIAL EVALUATION COMM'N, TENNESSEE APPELLATE JUDGES EVALUATION REPORT (2006), *available at* http://www.tncourts.gov/geninfo/Publications/judeval/judEval2006b.pdf (last visited May 13, 2008) (recommending the retention of three supreme court justices, twelve court of appeals judges, and twelve court of criminal appeals judges); TENN. JUDICIAL EVALUATION COMM'N, TENNESSEE APPELLATE JUDGES EVALUATION REPORT (2008), *available at* http://www.tncourts.gov/geninfo/Publications/JudicialEvaluation/2008%20Final%20Report.pdf (last visited May 13, 2008) (recommending the retention of two supreme court justices, two court of appeals judges, and one court of criminal appeals judge).

106. Telephone interview with Tim Gregory, Tennessee Division of Elections (Dec. 11, 2007) (transcript on file with Tennessee Law Review).

107. *Id.*

108. *See* SECRETARY OF STATE, TENNESSEE BLUE BOOK 1996–1997, at 543 (listing results for the August 1, 1996, general election). For an account of the controversial ruling that led to Justice White's defeat, see Carl A. Pierce, *The Tennessee Supreme Court and the Struggle for Independence, Accountability, and Modernization, 1974–1998*, *in* A HISTORY OF THE TENNESSEE SUPREME COURT 308–11 (James W. Ely Jr. ed., 2002).

109. The commissions were scheduled to expire on June 30, 2008, *see* TENN. CODE ANN. § 4-29-229(a)(46)–(47) (Supp. 2007), and after much hand wringing, the legislature refused to extend them. *See, e.g.*, John Rodgers, *Wilder's Last Gasp on State Judges Falls Short*, NASHVILLE CITY PAPER, May 21, 2008, available at http://www.nashvillecitypaper.com/news.php?viewStory=60360; John Rodgers, *Wilder's Judicial Plan on Rocks as Senate Tempers Ignite*, NASHVILLE CITY PAPER, May 14, 2008, available at http://www/nashvillecitypaper.com/news.php?viewStory=60187.

110. *See* TENN. CODE ANN. § 4-29-112 (2005) ("Upon the termination of any governmental entity under the provisions of this chapter, it shall continue in existence until June 30 of the next succeeding calendar year for the purpose of winding up its affairs. During that period, termination shall not diminish, reduce, or limit the powers or authorities of each respective governmental entity.").

does not enact a new system in the meantime, the selection of appellate judges will most likely return by default to the prior system of contested elections.[111]

III. LITIGATION AGAINST THE TENNESSEE PLAN

Although the Tennessee Plan has been in operation since 1971, the language from the 1870 Tennessee constitution that requires all judges in the state to be "elected by the qualified voters" has never been changed.[112] (Indeed, a proposed amendment that would have changed this language in favor of language providing for the Tennessee Plan was rejected by voters in 1977.[113]) For this reason, the Tennessee Plan has always operated under a cloud of legal uncertainty. Indeed, on three occasions since 1971, the Tennessee Plan's constitutionality has been tested in litigation.[114]

The earliest and most important litigation was *State ex rel. Higgins v. Dunn*.[115] In *Dunn*, a supreme court justice, Larry Creson, passed away in June 1972, some two years before his term was set to expire on August 31, 1974.[116] Governor Winfield Dunn appointed Thomas Turley, Jr., to fill the position from a list of names submitted by the judicial nominating commission, but the governor did not make the appointment effective until September.[117] In the

111. This is the case because the old statutory provisions requiring appellate judges to be selected by election are still on the books. *See, e.g.*, TENN. CODE ANN. § 17-1-103 (1994 & Supp. 2007) ("The judges of the supreme court, court of appeals, and court of criminal appeals are elected by the qualified voters of the state at large . . . ."); *see also, e.g.*, *id.* §§ 16-3-101, 16-4-102, 16-5-103. Although these provisions were repealed to the extent they conflict with the Tennessee Plan, *see* 1971 Tenn. Pub. Acts, ch. 198, § 17, the Tennessee Plan instructs the courts to return to contested elections if any provision of the Plan is held "invalid," *see* 1994 Tenn. Pub. Acts, ch. 942, § 23. It is true that allowing part of the Tennessee Plan to expire is not the same thing as a court holding part of the Plan "invalid," but it does suggest that the legislative intent behind the Tennessee Plan was to have all of it or none at all. This was also the assumption of one of the special courts that was asked to rule on the constitutionality of the Tennessee Plan; the special supreme court in *DeLaney* noted that, if the Tennessee Plan was by its terms inapplicable to a particular appellate vacancy, then the vacancy would be filled with a contested election. *See* DeLaney v. Thompson, 982 S.W.2d 857, 858 (Tenn. 1998) ("[T]he failure of the Commission to recommend the retention of any judge would render the Tennessee Plan inapplicable to the election to fill that judge's seat, and the election therefore would be conducted as any other election (rather than as a 'retention election').").

112. *See* TENN. CONST. art. VI, §§3, 4.

113. *See Constitutional Amendments Proposed by the Convention for Submission to the People*, *in* THE LIMITED CONSTITUTIONAL CONVENTION OF 1977, STATE OF TENNESSEE, THE JOURNAL OF THE LIMITED CONSTITUTIONAL CONVENTION of 1977 (see Article VI, § 4).

114. *See* DeLaney v. Thompson, 982 S.W.2d 857 (Tenn. 1998); State *ex. rel.* Hooker v. Thompson, No. 01S01-9605-CH-00106, 1996 WL 570090 (Tenn. Oct. 2, 1996); State *ex. rel.* Higgins v. Dunn, 496 S.W.2d 480 (Tenn. 1973).

115. 496 S.W.2d 480 (Tenn. 1973).

116. *See id.* at 482, 491.

117. *See id.* at 482.

meantime, there was an August general election, and, despite the fact that there was no ballot question for the vacant supreme court position, Robert Taylor ran a write-in campaign for the seat.[118] The secretary of state certified Taylor to the position, the governor certified Turley, and the entire matter went to the Tennessee Supreme Court for resolution.[119] The court held both that the governor's appointment was invalid (because the governor could not appoint someone to a vacancy beyond the time for the next general election) and that the write-in election was invalid (because the supreme court position had not been put on the ballot).[120]

Although it did not appear necessary to its decision, the *Dunn* court also considered the constitutionality of the Tennessee Plan.[121] The court found the Plan constitutional for two reasons. First, the court found that it was constitutional for the governor to initially appoint judges—despite the language of the constitution requiring their election—because the constitution elsewhere gives the legislature the power to prescribe how "all vacancies not otherwise directed or provided by this Constitution" shall be filled.[122] In the court's view, when Justice Creson passed away, a vacancy was created, and the broad powers of this provision kicked in.[123] The court noted that governors had been filling interim vacancies for over one hundred years.[124]

Second, the court found that the "yes or no" retention referendum that takes place under the Tennessee Plan at the next scheduled election qualifies as an "election" under the constitutional provision requiring all judges to be "elected by the qualified voters."[125] Although contested judicial elections had always been used under the 1870 constitution before the advent of the Tennessee Plan, the court noted that that the word "elected" in the constitution was not specifically defined, and, therefore, was ambiguous.[126] The court further noted that three other provisions of the constitution use the word "election" to refer to other ballot matters where voters are asked only a "yes or no" question;[127] in these provisions, voters are asked ballot questions such as whether to approve amendments to the constitution[128] or to authorize municipalities to lend

118. *See id.*

119. *See id.* at 482–83.

120. *See id.* at 487, 491.

121. *See id.* at 487.

122. *See id.* at 487–88 (quoting TENN. CONST. art. VII, § 4).

123. *See id.* at 488 ("[T]he Legislature as authorized by Article 7, Section 4, exercised the authority vested in it to make provision for 'the filling of all vacancies not otherwise directed or provided for by this Constitution.'").

124. *See id.* at 487–88.

125. *See id.* at 488 (quoting TENN. CONST. art. VI, § 3).

126. *See id.* at 489 ("The Constitution of Tennessee does not define the words, 'elect', 'election', or 'elected' and we have not found nor have we been referred to any provision of the Constitution or of a statute or to any decision of one of our appellate courts defining these words.").

127. *See id.*

128. *See* TENN. CONST. art. XI, § 3.

credit.[129] In light of these other provisions, the court thought that the word "election" could encompass a "yes or no" vote for a public official as well.[130] This was especially the case in light of another provision of the constitution giving the legislature the power to direct the "manner" of "election of all officers . . . not otherwise directed or provided by this Constitution."[131] The court concluded that, to the extent the legislature was given discretion in the constitution over prescribing the format of elections, the legislature was within its rights to choose retention referenda.[132]

One justice dissented in *Dunn*. Justice Humphreys argued that "the part of the Plan that does away with the popular election of judges, and substitutes a recall election, is so obviously contrary to the arrangement in our Constitution . . . for the people to have the right both to *nominate* and *elect* their constitutional officers" that the unconstitutionality of the Tennessee Plan was "obvious."[133] Justice Humphreys came to this view because the constitution requires the election not only of judges, but of other civil officers, including members of the legislature.[134] He argued that, if members of the legislature can abolish contested elections for judicial positions, then presumably they could do so for other positions, including their own, a result that he thought was clearly inconsistent with the constitution.[135]

After *Dunn*, the Tennessee legislature repealed the Tennessee Plan insofar as it applied to "vacancies" on the supreme court.[136] The legislature would not reauthorize the Plan for supreme court vacancies until 1994,[137] and when it did, it inspired a new round of litigation over the Plan's constitutionality. In 1996, the suits in *State ex rel. Hooker v. Thompson*[138] were filed by Lewis Laska and John Jay Hooker, two lawyers who wished to run for a seat then occupied by Justice Penny White (who, under the Tennessee Plan, would run only in a retention referendum).[139] The litigation went up to the Tennessee Supreme Court and was heard by a special panel of judges appointed by the governor because all of the regular justices recused themselves.[140] The special court held the Tennessee Plan constitutional on the authority of *Dunn*.[141] The *Thompson*

129. *See id.* art. II, § 29.

130. *See Dunn*, 496 S.W.2d at 489 ("It seems to us that if the Constitution itself denominates these methods of ratification as elections, it cannot be that Chapter 198 is unconstitutional because the elections therein provided for are limited to approval or disapproval.").

131. TENN. CONST. art. VII, § 4.

132. *See Dunn*, 496 S.W.2d at 489.

133. *Id.* at 493 (Humphreys, J., dissenting).

134. *See id.*

135. *See id.*

136. *See* 1974 Tenn. Pub. Acts, ch. 433, § 1.

137. *See generally* 1994 Tenn. Pub. Acts, ch. 942.

138. No. 01S01-9605-CH-00106, 1996 WL 570090 (Tenn. Oct. 2, 1996).

139. *See id.* at *1.

140. *See id.* at *1 n.6.

141. *See Thompson*, 1996 WL 570090, at *3 ("The issue of whether yes/no retention

opinion has never, however, been published in the official Tennessee Supreme Court reporter.[142] As a result, it is not considered binding precedent.[143]

The final piece of significant litigation challenging the constitutionality of the Tennessee Plan came in 1998, in *DeLaney v. Thompson*.[144] In this case, a court of appeals judge planned to retire at the end of his term, and the plaintiff, Robert DeLaney, sought to run for his seat.[145] The state coordinator of elections denied his application for the seat, and DeLaney sued.[146] The trial court held the Tennessee Plan unconstitutional, not because it denied the voters an election, but because it restricted the candidates who could seek a position on an appellate court to those selected by the judicial nominating commission.[147] The court of appeals, sitting as a special court in light of the recusals of the regular members, reversed and upheld the Tennessee Plan on the authority and arguments of *Dunn* and *Thompson*.[148] But the Tennessee Supreme Court, sitting as a special court as well, reversed the court of appeals on other grounds, finding it unnecessary to reach the constitutional question.[149]

It is interesting to note that, despite all this litigation, a majority of the regularly constituted Supreme Court has never upheld the constitutionality of the Tennessee Plan. In both *Thompson* and *Dunn*, the justices that upheld the constitutionality of the Plan were comprised largely of special justices appointed to hear only those particular cases.[150] In *Thompson*, as I noted, all

---

elections violate the Constitution of Tennessee has previously been decided by the Tennessee Supreme Court in the case of *State ex rel. Higgins v. Dunn*, and no compelling reason has been given to persuade this Court that it should disturb that ruling." (citation omitted)).

142. *See generally Thompson*, 1996 WL 570090.

143. *See* TENN. SUP. CT. RULE 4(A)(1), (G)(1) (designating opinions not published in the Southwestern Reporter as "persuasive" and not "controlling" authority to all persons other than those who were parties to the case). Interestingly, although the *Thompson* opinion has never been published in the Southwestern Reporter, the special supreme court that decided the case designated it as "for publication." I brought this discrepancy to the attention of the office of the Tennessee Attorney General (which is responsible for reporting supreme court decisions), and the office suggested that it may have committed an error by not publishing the opinion. The office also indicated that it might seek to correct the error by publishing the opinion now. If the *Thompson* opinion is eventually published in the Southwestern Reporter, it would presumably become binding precedent at that time.

144. 982 S.W.2d 857 (Tenn. 1998).

145. *See id.* at 858.

146. *See id.* at 859.

147. *See* DeLaney v. Thompson, No. 01A01-9806-CH-00304, 1998 WL 397363, at *1 (Tenn. Ct. App. July 16, 1998) (noting that the Chancery Court found the Tennessee Plan unconstitutional because "it drastically limits the group of persons who can become appellate judges" and "virtually insures the name of the incumbent on the ballot").

148. *See id.* at *5–8.

149. *See DeLaney*, 982 S.W.2d at 861 (holding that "the Tennessee Plan was inapplicable to the election to fill [the appellate judge's] seat").

150. Although it is beyond the scope of this paper to address this question, it is interesting to ask whether it is comports with the Due Process Clause of the U.S. Constitution to permit the

five justices recused themselves and the governor named special justices to replace them.[151] In *Dunn*, two of the five justices who heard the case were special justices, including two of the four justices who comprised the majority that upheld the Plan.[152] Although, as a formal legal matter, decisions by special justices are just as binding as those rendered by regular justices,[153] the fact that the Plan has never been upheld by a regular court has only added to its controversy.

IV. IS THE TENNESSEE PLAN CONSTITUTIONAL?

As noted above, under the 1870 Tennessee constitution, all judges in the state must be "elected by the qualified voters."[154] For most of Tennessee's history, that meant judges were initially placed into new terms and retained for subsequent terms through contested elections.[155] Under the Tennessee Plan, however, judges are initially placed into new terms by gubernatorial appointment, and judges are retained for subsequent terms by retention referenda.[156] The question is whether these two devices—initial appointment by the governor and the retention referendum—are consistent with the constitutional requirement that all judges be "elected."

As explained below, it is hard to see how these devices are consistent with the constitution. Moreover, to the extent any uncertainty existed over the meaning of the Tennessee constitution, that uncertainty was arguably resolved by the people of Tennessee in 1977 when they rejected an amendment to the constitution that would have replaced the provision requiring elected judges with one that would have permitted the Tennessee Plan.[157] Indeed, of the seventeen states that select judges by some mechanism of appointment followed by a retention referendum, Tennessee is the only one that has not revised its constitution to change a provision requiring elections in favor of a provision setting forth the appointment-retention mechanism.[158]

---

governor, as he did in the *Thompson* case, to appoint an entire supreme court to hear a single case *after* the issue in that case is already known. In these circumstances, the governor has an unusual power to control the outcome of the case by appointing judges sympathetic to his views.

151. *See supra* text accompanying note 140.

152. *See Dunn*, 496 S.W.2d. at 491 (noting that Justices McAmis and Wilson joined the majority as special justices).

153. *See* Ridout v. State, 30 S.W.2d 255, 257 (1930).

154. TENN. CONST. art. IV, §§ 3, 4.

155. *See supra* text accompanying notes 44.

156. *See supra* text accompanying notes 76–104.

157. *See infra* text accompanying notes 204–11.

158. *See infra* notes 212–13 and accompanying text.

A. Are Judges "Elected" if They are Initially Appointed by the Governor?

Under the Tennessee Plan, judges are initially placed on the bench through an appointment by the governor,[159] and they can serve for as long as two years before they are put before the people in retention referenda.[160] Yet, Article VI of the 1870 Tennessee constitution requires that all state judges be "elected by the qualified voters."[161] How can the two be reconciled?

The answer given by the *Dunn* court[162] refers to another part of the 1870 Tennessee constitution, Article VII, which states that "the filling of all vacancies not otherwise directed or provided by this Constitution . . . shall be made in such manner as the legislature shall direct."[163] But if the constitution permits the legislature to fill judicial "vacancies" however it wishes, then what effect would be left for the provision of the constitution requiring judicial elections? That is, if any time a judge left office and a position became open the legislature could empower the governor to appoint a replacement, then the provision regarding vacancies would nullify the provision requiring an elected judiciary.

The solution to this puzzle is that the authors of the 1870 constitution did not intend the word "vacancies" in Article VII to include a judicial position that becomes available because a judge has served his or her entire term and chooses not to run for reelection. Rather, the authors of the 1870 constitution intended "vacancies" to mean interim judicial positions that became available in the middle of a term, such as by the death or resignation of a judge. Appointment is a common mechanism by which to fill interim vacancies in states that otherwise elect office holders; it is often thought too expensive and too cumbersome to hold a special election every time a someone leaves office early.[164] Indeed, the Tennessee constitution explicitly prohibits special elections for judges.[165]

---

159. *See* TENN. CODE ANN. §17-4-112(a) (2004 & Supp. 2007) ("When a vacancy occurs in the office of an appellate court . . . by death, resignation or otherwise, the governor shall fill the vacancy by appointing one (1) of three (3) persons nominated by the judicial selection committee . . . .").

160. *See* TENN. CODE. ANN. § 17-4-112(b) (2004 & Supp. 2007) ("The term of a judge appointed under this section shall expire on August 31 after the next regular August election occurring more than thirty (30) days after the vacancy occurs.").

161. TENN. CONST. art. VI, §§ 3, 4.

162. *See* Shriver *ex rel* Higgins v. Dunn, 496 S.W.2d 480, 487 (Tenn. 1973).

163. TENN. CONST. art. VII, § 4.

164. *See* Joseph A. Colquitt, *Rethinking Judicial Nominating Commissions*, 34 FORDHAM URB. L. J. 73, 77 (2007) ("The death or resignation of a judge from the active bench seriously disrupts the work of the court, and the speedy selection of a replacement is important to the litigants and the public. Most, perhaps virtually all, of these interim vacancies are filled by gubernatorial appointment. In a few states, the legislature makes the appointment. Alternatively, a state could choose a special election, but that method entails uncertainty, delay, and costs. Appointment is the better method of filling vacancies."); Daniel R. Deja, *How Judges are Selected*, 75 MICH. B.J. 904, 906 (1996) ("Judges die or resign from office on

It is apparent from a neighboring provision in Article VII that the authors of the 1870 constitution used the word "vacancies" there to refer only to interim vacancies. The neighboring provision states that "[n]o appointment . . . to fill a vacancy shall be made for a period extending beyond *the unexpired term*."[166] By limiting the legislature's ability to fill vacancies only for the rest of an "unexpired term," the authors of the 1870 constitution indicated that they intended for the legislature to fill only those vacancies *with* unexpired terms—i.e., only those that occur in the middle of a term (such as by death or resignation) and not those that occur when a judge serves his or her entire term but chooses not to run for reelection (in which case there is no "unexpired term" remaining).

Thus, to the extent the Tennessee Plan permits the governor to appoint a new judge to a position created when the previous judge served his or her full term, the Plan would appear unconstitutional.

None of the courts that have considered the constitutionality of the Tennessee Plan have addressed this point. Indeed, not only has this point never been addressed, but the two Tennessee Supreme Court opinions that upheld the Tennessee Plan are not even necessarily to the contrary. In both *Dunn* and *Thompson*, the vacancy occurred in the middle of a term.[167] There is no doubt that this is the kind of vacancy that the Tennessee constitution permits the legislature to fill in whatever manner it chooses.[168] With respect to other vacancies, however—those that occur when a judge completes his or her term and does not run for reelection—it is hard to see how the initial appointment device of the Tennessee Plan is constitutional.

B. Are Retention Referenda " Elections"?

As the supreme court in *Dunn* noted, the 1870 constitution does not explicitly say whether a retention referendum qualifies as an "election."[169] The court thought, however, that the constitution answered this question elsewhere. The court found three provisions in the constitution where the word "election" is used to describe a vote that, much like a retention referendum, poses only a

---

schedules that are determined by factors other than the dates of general elections. The cost and time needed to schedule special elections is generally prohibitive. The presumption is that gubernatorial appointment is a more expedient and cost-efficient means of filling judicial vacancies.")

165. *See* TENN. CONST. art. VII, § 5 ("No special election shall be held to fill a vacancy in the office of Judge . . . .").

166. *Id.* (emphasis added).

167. *See* State *ex rel* Higgins v. Dunn, 496 S.W.2d 480, 482, 491 (Tenn. 1973); State *ex rel* Hooker v. Thompson, 1996 WL 570090, at *1 n.1 (Tenn. Oct. 2, 1996) (noting that Justice White had been initially appointed to fill the unexpired term of Justice O'Brien).

168. *See* TENN. CONST. art. VII, § 4 ("[T]he filling of all vacancies not otherwise directed or provided by the Constitution, shall be made in such manner as the Legislature shall direct.").

169. *Dunn*, 496 S.W.2d at 487.

yes-or-no question to the voters.[170] One of these provisions requires an "election" to authorize a municipal government to loan its credit to others;[171] another requires amendments to the constitution to be approved "at an election";[172] and another requires a variety of other municipal acts to be ratified "in an election."[173] In each of these instances, the constitution refers not to a vote that is contested between two people, but to one that asks for an up or down decision by the voters. The court extrapolated from these three provisions to conclude that the retention referendum mechanism in the Tennessee Plan qualified as an "election" as well.[174]

There are several difficulties with extrapolating from these three examples to a conclusion that the word "election" in Article VI must include uncontested, yes-or-no votes on the tenure of public officials. The first difficulty is the one raised by Justice Humphreys in *Dunn*: If a retention referendum can be an "election" for judges, why not also for other public officials, such as legislators or even the governor?[175] The majority did not respond to this argument, and there is good reason for that: The argument is hard to answer. One might be able to distinguish the constitutional provision requiring the election of legislators from that requiring the election of judges: The former says that the legislature shall be "dependent on the people,"[176] whereas the latter says that judges "shall be elected,"[177] and one might argue that the former implies a different, more democratic form of election than the latter. It is quite difficult, however, to distinguish the provision requiring the election of the governor. Like the provision for judges, the provision for the governor says simply that the "governor shall be elected."[178] Thus, if *Dunn* is correct, then the legislature might permit governors to win second terms in uncontested retention referenda—a proposition few would believe is consistent with the democratic guarantees of the Tennessee constitution.

There are other difficulties with the *Dunn* analysis. For example, two of the three examples relied upon by the court were not even part of the 1870 constitution; they were added many decades later, in 1953.[179] These two examples are, therefore, of little probative value in discerning what the authors of the 1870 constitution meant when they used the word "elected." In addition, all three examples relied upon in *Dunn* involved votes on ballot propositions as opposed to votes on public officials. Voting on ballot propositions has almost always taken place in the form of yes-or-no votes—the proposition is either

170. *See id.* at 489.
171. TENN. CONST. art. II, § 29.
172. *Id.* art. XI, § 3.
173. *Id.* art. XI, § 9.
174. *See Dunn*, 496 S.W.2d at 489.
175. *See id.* at 493 (Humphreys, J., dissenting).
176. TENN. CONST. art. II, § 3.
177. *Id.* art VI, §§ 3, 4.
178. *Id.* art. III, § 4.
179. *See id.* art. XI, §§ 3, 9.

agreed to or not.[180] By contrast, voting for public officials has rarely taken place—and, for most of American history, had never taken place—in the yes-or-no form.[181] The fact that the 1870 constitution once uses the word "election" to refer to a yes-or-no vote in the ballot proposition context, where such votes have almost always taken place in the yes-or-no form, does not answer the question whether the word "election" means the same thing in the different context of public officials, where such votes have almost never taken place in the yes-or-no form.

But perhaps the greatest difficulty with the conclusion that the authors of the 1870 constitution intended the word "election" to include retention referenda is that such referenda appear to have been unknown in the United States at that time. The first retention referendum was adopted in the United States in 1934,[182] and the very idea of a retention referendum for public officials was not even conceived until 1914, when it was first proposed by a law professor at Northwestern University.[183] It is, obviously, impossible for the authors of the 1870 constitution to have intended that document to encompass something that did not yet exist. No court considering the constitutionality of the Tennessee Plan has addressed this point.

Of course, the authors of the 1870 constitution did not know many of the things that we know today. Many scholars believe it would be cumbersome and impractical to force legislatures to amend their constitutions every time they wanted to take advantage of a new idea or a new technology; these scholars believe that the meanings of constitutional provisions should change over time to encompass new ideas so long as the new ideas serve the old purposes.[184] This reasoning is especially appropriate in this case because, as

---

180. The only possible exception of which I am aware is the dilemma that is occasionally created by "conflicting ballot propositions." *See* Philip L. Dubois & Floyd Feeney, LAWMAKING BY INITIATIVE 158–163 (1998). In a small handful of states, voters can be given a choice between two competing ballot questions. *See id.* at 160–61 (listing the states of Washington, Maine, Mississippi, and Massachusetts). This practice has never been followed in Tennessee, *see id.* at 158–163, and, even in the states in which it is practiced, it arose during the Progressive Era and well after the 1870 constitution was written. *See* ME. CONST. art. IV, pt. 3, § 18 (approved in 1909); MASS. CONST. amend. Art. XLVIII, Init., pt. 6 (approved in 1918); MISS. CONST. art. 15, § 273 (approved 1912); WASH. CONST. art. II, § 1 (amended in 1911).

181. *See infra* notes 182–83 and accompanying text.

182. *See* CARBON & BERKSON, *supra* note 51, at 11.

183. *See id.* at 2. Of course, other mechanisms of removing public officials from office were well known in 1870, including impeachment and recall. Until the Progressive Era, however, it appears that neither of these mechanisms had ever been placed directly in the hands of the electorate. Thus, even the closest analogue to the retention referendum—the recall election—post-dated the 1870 Convention. *See* Rod Farmer, *Power to the People: The Progressive Movement for the Recall 1890s–1920*, 57 NEW ENG. J. HIST. 59, 62, 64 (2001); Joshua Spivak, *California's Recall: Adoption of the "Grand Bounce" for the Elected Officials*, CAL. HIST., Mar. 22, 2004, at 22.

184. *See, e.g.*, Michael J. Klarman, *Antifidelity*, 70 S. CAL. L. REV. 381, 395 (1997) (noting that many scholars and judges believe that the Constitution should be interpreted "by

the *Dunn* court noted, the 1870 constitution explicitly confers flexibility on the legislature in deciding the "manner" in which judicial elections should take place where the constitution does not otherwise provide.[185] Thus, even though the retention referendum was unknown in 1870, the device nonetheless may be constitutional because it serves the democratic purposes of the 1870 constitution just as well as contested elections do. There are a number of reasons, however, to doubt that retention referenda do a very good job of facilitating democratic accountability.

First among these reasons is the fact that retention referenda were originally designed to *insulate* judges from public accountability. The architects of merit selection in the early nineteenth century favored life tenure for judges, but feared that the post-Jacksonian public would no longer accept this as they once had.[186] Thus, the architects of merit selection came up with what some scholars have concluded was a "sop" to the public: the retention referendum.[187] That is, the retention referendum was designed to make the public feel as though they had a role in selecting their judges but make it unlikely they would exercise that role by voting a judge off the bench.[188]

The experience with retention referenda has vindicated its design. Scholars have found that judges virtually never lose retention referenda. In the most comprehensive study, which examined over thirty years of data in ten states, judges running in retention referenda were returned to office 98.9% of the time.[189] Even that incredibly high number is misleading, however, because over half of the defeats were from Illinois, a state that requires judges to win 60% of the vote rather than a mere majority (as do Tennessee and most other states) in order to stay on the bench.[190] Removing the Illinois defeats from the data where the judges won more than 50% but less than 60% of the vote yields a retention rate of 99.5%.[191] By contrast, judges who run for reelection in states that use contested elections are defeated much more often. One comprehensive study of state supreme court races between 1980 and 2000 showed that justices running for reelection in states that use partisan elections were defeated nearly 23% of the time—a full *thirteen times* as often as justices running in retention

---

'translating' the Framers' concepts into modern circumstances").

185. TENN. CONST. art. VII, § 4 ("The election of all officers . . . not otherwise directed or provided by this Constitution, shall be made in such manner as the legislature shall direct.").

186. *See* CARBON & BERKSON, *supra* note 51, at 6, 8.

187. *See, e.g., id.* at 8, 10 (noting that the architects "perceived retention as a 'sop' to those committed to electoral control over the judiciary").

188. *See, e.g.*, Michael R. Dimino, *The Futile Quest for a System of Judicial "Merit" Selection*, 67 ALB. L. REV. 803, 806 (2004) ("Merit selection uses the public as participants in what is predetermined to be a useless exercise designed to ensure the retention of the incumbent.").

189. *See* Aspin, *supra* note 63, at 79 (finding that only fifty two out of 4,588 judges were not retained).

190. *See id.*

191. *See id.*

referenda over the same period.[192] As the author of that study has noted, in states that use contested elections, "supreme court justices face competition that is, by two or three measures, equivalent if not higher to that for the U.S. House."[193]

The experience in Tennessee is in line with these studies. As noted above, there have been 146 retention referenda in Tennessee, and in every single referendum but one (99.3%), the voters retained the incumbent.[194]

It is unclear why the public so infrequently votes against retention. One possible theory is that, without another candidate in the race, there is no one with an interest in providing information to the public about the incumbent.[195] Another possible theory is that, in this atmosphere of inadequate information, the absence of a political trademark—affiliation with a political party—makes it especially hard for voters to assess whether to retain a public official.[196] Finally, some commentators believe that voters are reluctant to vote against an incumbent if they have no idea who will replace the incumbent—"the devil you know is preferable to the devil you don't."[197] Regardless of the reason for the high rates of retention, scholars have concluded that, in light of the fact that these judges are a virtual lock to keep their seats, "those who maintain that retention elections serve to insulate judges from popular control seem to be correct."[198]

It should be noted that the Tennessee Plan is a bit different from many of the merit selection plans used in other states insofar as judges appointed under the Plan do not automatically run in retention referenda. Rather, they do so only if the judicial evaluation commission recommends that the public retain them; if the commission votes the other way, they must run in a contested election.[199] Thus, in assessing the accountability offered by the Tennessee Plan, the fact that the commission might not grant some judges the security of retention referenda should be considered. It appears, however, that this feature of the Tennessee Plan has not transformed it into a device of democratic

192. *See* Melinda Gann Hall, *Competition as Accountability in State Supreme Court Elections*, *in* RUNNING FOR JUDGE 177 (Matthew Streb ed., 2007) (finding that 22.9% of state supreme court incumbents were defeated in partisan elections while only 1.8% of incumbents were defeated in retention referenda between 1980 and 2000).

193. Melinda Gann Hall, *State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform*, 95 AM. POL. SCI. REV. 315, 319 (2001).

194. *See supra* text accompanying notes 107–08.

195. *See, e.g.*, Dimino, *supra* note 188, at 805 ("By removing challengers from the ballot, retention races eliminate the public figures most likely to motivate and organize opposition to the incumbent.").

196. Political scientists believe "that the most important cue for voters is political party affiliation. Party labels are signals . . . and voters rely heavily on them." Kritzer, *supra* note 6, at 433 (footnote omitted).

197. *See* STUMPF, *supra* note 6, at 170 ("You can't beat somebody with nobody.").

198. William K. Hall & Larry T. Aspin, *What Twenty Years of Judicial Retention Elections Have Told Us*, 70 JUDICATURE 340, 347 (1987).

199. *See* TENN. CODE ANN. §§ 17-4-114(c), -115(c) (Supp. 2007).

accountability. Since the commission was created in 1994, it has rendered sixty-six evaluations.[200] In every single one, the commission recommended that the judge be retained.[201]

Despite the limitations of retention referenda, there are some reasons to think that they are no less democratic than the contested elections that preceded them. Although judges who run in referenda are virtually guaranteed to win, they nonetheless report on surveys that the prospect of running in the referenda influences their decisions on the bench.[202] Thus, it is possible that retention referenda produce judges that are accountable to the public even though they do not produce judges who get defeated. Moreover, it bears reiterating that, even when contested elections were used to select appellate judges in Tennessee, the races were often not very spirited. As noted above most judges still came to the bench through gubernatorial appointment, and, in a state that was for a long time controlled by one political party, even the reelection campaigns often were not contested.[203] Thus, even if retention referenda are largely coronations, it is not entirely clear that, at least as an historical matter, contested elections were much different. For this reason, the case against the constitutionality of the Tennessee Plan's provision for retention referenda is not as strong as it is against the provision calling for gubernatorial appointment of all appellate judges in the first instance.

C. What About the Failed Amendment of 1977?

Much of the uncertainty over the constitutionality of the Tennessee Plan might have been resolved in 1977. That year, a limited Constitutional Convention proposed to the people of Tennessee thirteen separate amendments to the Tennessee constitution.[204] The thirteen amendments covered topics as diverse as repealing the 1870 constitution's ban on interracial marriage to repealing the 1870 constitution's prohibition on charging interest rates of more than 10%.[205] One of the amendments would have made several changes to the judiciary, including repeal of the 1870 constitution's requirement that all judges "shall be elected" in favor of a provision stating that "Justices of the Supreme Court and judges of the Court of Appeals shall be appointed by the Governor from three nominees recommended . . . by the Appellate Court Nominating Commission" and that "[t]he name of each justice and judge seeking retention shall be submitted to the qualified voters for retention or rejection . . . at the expiration of each six year term."[206] In other words, the proposed amendment

---

200. *See supra* note 105 and accompanying text.

201. *See id.*

202. *See* Larry T. Aspin & William K. Hall, *Retention Elections and Judicial Behavior*, 77 JUDICATURE 306, 312–13 (1994).

203. *See* Parks, *supra* note 36, at 629–30.

204. *See* LASKA, *supra* note 73, at 23–25.

205. *See id.* at 24–25.

206. *See* Governor Ray Blanton, *Proclamation by the Governor*, *in* THE LIMITED

would have replaced the constitution's requirement of an elected judiciary with the Tennessee Plan.

Voters approved each of the thirteen proposed amendments submitted to the public from the 1977 Convention *except* the amendment that would have inserted the Tennessee Plan into the constitution.[207] As one historian has noted, this amendment "became the first amendment ever offered by a limited convention to face voter rejection."[208]

The fact that voters rejected putting the Tennessee Plan into the constitution when given the chance is a powerful, but not conclusive, point in favor of the view that the Tennessee Plan is unconstitutional. The amendment containing the Tennessee Plan would have made many other significant changes to the judicial branch, including the designation of a uniform jurisdiction for all trial courts and the creation of a statewide public defender program.[209] It is possible that the voters favored the Tennessee Plan but rejected the amendment for the other changes it would have made to the judicial branch. Indeed, the Tennessee Plan does not appear to have been the most controversial part of the proposed amendment.[210] It is also possible that the voters favored the Tennessee Plan but rejected the amendment because they thought the constitution *already* permitted the Plan.

Despite the ambiguous meaning of rejected constitutional amendments, it is certainly not uncommon to use them to interpret the meaning of a constitution.[211] Nonetheless, the Tennessee Supreme Court has never addressed the events of 1977 in any of its opinions regarding the constitutionality of the Tennessee Plan. This is even more surprising in light of the constitutional experience of other states with similar methods of judicial selection. Of the seventeen states that rely upon appointment by the governor followed by a retention referendum,[212] Tennessee is the *only one* that has not

---

CONSTITUTIONAL CONVENTION OF 1977, STATE OF TENNESSEE, THE JOURNAL OF THE DEBATES OF THE LIMITED CONSTITUTIONAL CONVENTION OF 1977 (see Proposal 13, § 4).

207. *See* LASKA, *supra* note 73, at 25–26.

208. *Id.* at 26.

209. *See id.* ("All trial courts were to have uniform jurisdiction, and the legislature was restricted in creating new types of courts; the Missouri Plan was approved for appellate judges. Provision was made for a chief court administrator. The legislature was required to set up a statewide public defender program.").

210. *See id.* at 24–25.

211. *See, e.g.*, Seminole Tribe of Florida v. Florida, 517 U.S. 44, 111 (1996) (Souter, J., dissenting) ("If the Framers had meant the Amendment to bar federal-question suits as well, they could not only have made their intentions clearer very easily, but could simply have adopted the first post-*Chisholm* proposal, introduced in the House of Representatives by Theodore Sedgwick . . . ."); David B. Kopel, *It Isn't About Duck Hunting: The British Origins of the Right to Arms*, 93 MICH. L. REV. 1333, 1357 (1995) (arguing that "the 'National Guard' interpretation of the Second Amendment amounts to an Orwellian reversal [by] treating the enacted Amendment that guarantees the right of the people as having a meaning identical to a proposed but rejected amendment dealing with the rights of states").

212. The seventeen states are: Alaska, Arizona, California, Colorado, Florida, Indiana,

revised its constitution to change a requirement of an elected judiciary in favor of a provision setting forth the appointment-retention mechanism.[213] Courts often interpret constitutions in light of how neighboring jurisdictions have treated similar provisions in their own constitutions.[214] When the voters' decision in 1977 is juxtaposed against the experience in every other state, it becomes even harder to conclude that the Tennessee Plan is consistent with the constitution. Again, although this point might not be dispositive on its own, when combined with the other serious doubts about the Tennessee Plan, the case against its constitutionality becomes close to compelling.

V. CONCLUSION

Ever since it was enacted in 1971, the Tennessee Plan has been controversial, and its greatest controversy has always been whether it is constitutional. Although the Tennessee Supreme Court has twice said that it is, neither of these decisions was supported by a majority of regular supreme court justices, one decision was unpublished and therefore is not even binding, and, most importantly of all, both of these decisions left serious constitutional questions unanswered. The most serious of these questions is how the constitution's requirement that all judges be "elected" can be squared with the Plan's requirement that all judges be appointed by the governor. Although the constitution permits the governor to appoint judges to "vacancies," it seems rather clear that the constitution means only interim vacancies. Thus, to the extent the Tennessee Plan calls for the governor to appoint judges to positions where their predecessors completed their full terms, it is hard to come to any conclusion other than that the Tennessee Plan is unconstitutional.

There are also serious doubts that the feature of the Tennessee Plan requiring judges to run for reelection in uncontested retention referenda is consistent with the constitution. Although these doubts are not as strong as those surrounding the feature of the Plan requiring gubernatorial appointment, when combined with the fact that the voters of Tennessee rejected a

---

Iowa, Kansas, Maryland, Missouri, Nebraska, New Mexico, Oklahoma, South Dakota, Tennessee, Utah, and Wyoming. *See* American Judicature Society, Methods of Judicial Selection: Selection of Judges State-by-State Report, *available at* http://www.judicialselection.us/judicial_selection/methods/selection_of_judges.cfm?state= (last visited Feb. 25, 2008).

213. *See* ALASKA CONST. art. IV, §§ 5, 6; ARIZ. CONST. art. VI, §§ 37, 38; CAL. CONST. art. VI, § 16(d); COLO. CONST. art. VI, §§ 20, 25; FLA. CONST. art. V, §§ 10, 11; IND. CONST. art. VII, §§ 10, 11; IOWA CONST. art. V, §§ 15, 17; KAN. CONST. art. III, § 5; MD. CONST. art. IV, § 5A; MO. CONST. art. V, § 25; NEB. CONST. art. V, § 21; N.M. CONST. art. VI, §§ 33, 35; OKLA. CONST. art. VII, § 3; S.D. CONST. art. V, § 7; UTAH CONST. art. VIII, §§ 8, 9; WYO. CONST. art. V, § 4.

214. *See, e.g.*, Richardson v. Tenn. Bd. of Dentistry, 913 S.W.2d 446, 453 (Tenn. 1995) ("careful[ly] examin[ing] cases . . . in other states" to resolve a question of Tennessee constitutional law).

constitutional amendment that would have explicitly adopted the Tennessee Plan, these doubts, too, become compelling. It is rare that we have explicit instructions from the body politic as to what a constitutional provision means; the public's rejection of the 1977 amendment is often as close as we ever come.

For these reasons, I am persuaded that the best reading of the Tennessee constitution is one that holds the Tennessee Plan unconstitutional. If we are to have merit selection in Tennessee, then the proponents of merit selection should do what they could not do in 1977: persuade the voters to pass a constitutional amendment. Until then, the legislature—duty bound as it is to uphold the Tennessee constitution[215]—should do what it needs to do to return the selection of appellate judges to contested elections. The easiest way to accomplish this task would be to take no action in June 2009 when the operative pieces of the Tennessee Plan—the judicial nominating and evaluation commissions—are scheduled to terminate.[216] If the legislature allows the commissions to terminate—and does not adopt a new system in the meantime—then appellate judges in Tennessee should revert by default to initial selection and retention in contested elections.[217] In short, as the Wall Street Journal recently put it, "the best thing [the legislature] can do is nothing at all."[218] +

215. *See* TENN. CONST. art. X, § 2 ("Each member of the Senate and House of Representatives, shall before they proceed to business take an oath or affirmation to support the Constitution of this State . . . .").

216. *See supra* notes 109–10 and accompanying text.

217. *See supra* note 111.

218. *Tennessee's Trial Run*, WALL ST. J., May 10, 2008, at A10.

+ [EDITOR'S NOTE: Professor Penny White and Malia Reddick, Ph.D., have written a response to this Essay. It may be found at 75 TENN. L. REV. 501 (2008). In addition, Professor Fitzpatrick has written a reply to their response, which is posted at http://papers.ssrn.com/abstract=1152413.]

A RESPONSE TO PROFESSOR FITZPATRICK: THE REST OF THE STORY

PENNY J. WHITE* & MALIA REDDICK+

In his essay *Election as Appointment: The Tennessee Plan Reconsidered*, Professor Brian T. Fitzpatrick contends that Tennessee's selection and retention method for appellate court judges is both unconstitutional and unmeritorious.[1] This Essay responds to those claims. Part I will respond to Professor Fitzpatrick's claim that the Tennessee Plan is unconstitutional; Part II will respond to his claim that the Plan is not fulfilling the purposes which led the Tennessee legislature, in its wisdom, to adopt it.

It is impossible, or at least disingenuous, to respond to Professor Fitzpatrick's essay without highlighting a multitude of significant omissions that must be considered to fairly evaluate either the constitutionality or the merit of the Tennessee Plan. The essay exhibits a cherry-picking tendency[2] throughout that prompts memories of Paul Harvey's favorite line: "And now you know the rest of the story." This response will complete the story, mindful that "history is the witness that testifies to the passing of time; it illumines reality, vitalizes memory, provides guidance in daily life, and brings us tidings of antiquity."[3]

---

* Professor of Law and Director, Center for Advocacy and Dispute Resolution, University of Tennessee College of Law; former Co-Chair of the Tennessee Judicial Performance Guidelines Committee; author of Part I of this Essay.

+ Ph.D., Michigan State University; Director of Research and Programs at the American Judicature Society, a national nonpartisan organization dedicated to maintaining the independence and integrity of the courts and increasing public understanding of the justice system; author of Part II of this Essay.

1. The Tennessee General Assembly failed to pass the legislation necessary to continue the Tennessee Plan before it adjourned in May 2008.

2. For example, Professor Fitzpatrick begins his essay by suggesting that the Tennessee General Assembly's decision to move to a merit selection system for appellate judges was a response to changes in other states. Brian T. Fitzpatrick, *Election as Appointment: The Tennessee Plan Reconsidered*, 75 TENN. L. REV. 473, 473 (2008). This suggestion ignores the myriad of circumstances that led to the 1971 legislation. *See infra* text accompanying notes 66–77. Before the close of the second paragraph, the essay misinforms the reader on the mechanics of the Tennessee Plan, describing it as marred in controversy, and stating, without attribution, that "many people doubt" whether it has accomplished its purposes. Fitzpatrick, *supra*, at 473. These propositions, even when addressed in more detail in the body of the essay, create an incomplete and, unfortunately, misleading description of the issues that the Professor undertakes to address.

3. CICERO, PRO PUBLIO SESTIO.

I. AN ANALYSIS OF TENNESSEE HISTORY, FUNDAMENTAL PRINCIPLES, AND THE CONSTITUTIONALITY OF THE TENNESSEE PLAN

Before responding to the essay's three specific challenges to the constitutionality of the Tennessee Plan, we will address two essential cornerstones absent from its analysis. The first is the historic role of the Tennessee legislature in judicial selection; the second are the fundamental principles of statutory construction and constitutional interpretation.

A. The Tennessee Legislature's Historic Role in Judicial Selection

Professor Fitzpatrick's essay begins by suggesting that Tennessee simply fell in line with other states in the 1970s to move from an elective to an appointive system of judicial selection for appellate court judges.[4] While the article does briefly acknowledge that the legislature elected judges for the first half of Tennessee's history,[5] it does not recognize the continued integral role that the Tennessee legislature would play in judicial matters. Described as "preeminent," the first Tennessee legislature was granted the power to elect most state officers.[6] That the legislature would also control the creation of the courts and the selection of judges was never doubted.

1. The Constitution of 1796

As historians have noted, any discussion of Tennessee's constitutional history must begin with a discussion of North Carolina's constitutional history.[7] This is because Tennessee "as the daughter of North Carolina, quite naturally adopted the judicial system of the Mother State."[8] Similar to most of the original states, the North Carolina legislature controlled the state, choosing both the governor (described as "little more than a dependency of the legislature")[9] and the judges.[10] The early constitutions of North Carolina and Tennessee

---

4. Fitzpatrick, *supra* note 2, at 473.

5. *Id.* at 478–79.

6. N. Houston Parks, *Judicial Selection—The Tennessee Experience*, 7 MEM. ST. U. L. REV. 615, 619 (1977).

7. LEWIS L. LASKA, THE TENNESSEE STATE CONSTITUTION: A REFERENCE GUIDE 2 (1990).

8. SAMUEL C. WILLIAMS, PHASES OF THE HISTORY OF THE SUPREME COURT OF TENNESSEE 5 (1944). It is surmised that Tennessee's frontier leaders chose to follow the North Carolina model in order to add "respectability" to their separatist movement. Parks, *supra* note 6, at 619.

9. WALLACE MCCLURE, STATE CONSTITUTION-MAKING WITH ESPECIAL REFERENCE TO TENNESSEE 33–35 (1916); *see* TENN. CONST. art. II, § 2 (1796) ("The governor shall be chosen by the electors of the members of the general assembly . . . .").

10. MCCLURE, *supra* note 9, at 35.

therefore provided for the legislative election of judges,[11] and left "the establishment of courts entirely to the legislature."[12] Both constitutions contained sections which were entitled "Election of Judges," and both provided for these judicial "elections" by joint ballot of the two houses of the General Assembly.[13] From its initial use in Tennessee's first constitution, the word "elect" has maintained a broad and generic meaning.[14]

Tennessee's first constitution, adopted in 1796, granted judicial power to the courts, but retained for the legislature all power to establish courts, set their jurisdiction, and determine the methods for the selection of judges.[15] This legislative preeminence was consistent with the model of the times in which most governmental power was entrusted to a legislative body.[16] It follows that the legislature would be entrusted to elect the judiciary.[17]

---

11. The original draft of the 1796 constitution included the creation of a constitutional superior court comprised of three judges. This significant departure from the North Carolina model was not adopted. JOSHUA W. CALDWELL, STUDIES IN THE CONSTITUTIONAL HISTORY OF TENNESSEE 149 (2d ed. 1907). Rather, the 1796 Tennessee Constitution provided that "[t]he judicial power of the state shall be vested in such superior and inferior courts of law and equity, as the legislature shall, from time to time, direct and establish." TENN. CONST. art. V, § 1 (1796).

12. CALDWELL, *supra* note 11, at 149.

13. TENN. CONST. art. V, § 2 (1796); N.C. CONST. art. XIII (1776); *see* MCCLURE, *supra* note 9, at 424–25.

14. *See infra* text accompanying notes 26–50.

15. *See* TENN. CONST. art. V, §§ 1–12 (1796); Lewis L. Laska, *The Tennessee Constitution*, *in* TENNESSEE GOVERNMENT AND POLITICS: DEMOCRACY IN THE VOLUNTEER STATE 7, 8 (John R. Vile & Mark Byrnes eds., 1998); MCCLURE, *supra* note 9, at 48 ("An article of twelve sections defines in considerable detail the judicial system of the state, but leaves the establishment of the courts and the appointment of the judges entirely to the legislature . . . ."). In his book, Joshua Caldwell suggests that the detail contained in the several sections was more a result of oversight than intention. When the initial proposal for the creation of a superior court set forth in article V, section 1 was defeated, Caldwell asserts that the remaining portions were "not carefully recast." CALDWELL, *supra* note 11, at 150.

16. MCCLURE, *supra* note 9, at 138.

17. The state's early history is replete with colorful descriptions of judges elected by the legislature. Because the constitution provided for election by both houses of the legislature, the chore often involved multiple ballots and numerous candidates. One such election, described as one of the "hottest races in judicial annals" was that of Justice Robert J. McKinney, an Irishman. McKinney was elected on the seventh ballot in the legislature despite his having written a letter of recommendation for the other candidate. WILLIAMS, *supra* note 8, at 55 n.18. Justice McKinney was later praised as the supreme court's best opinion writer, noted for his "incisive . . . and logical [opinions] marked by [their] brevity and unusual clarity and exactness." *Id.* at 56. But he was deemed "unelectable" by the people because "he had not the parts or arts of the politician." *Id.*; *see also* Timothy S. Huebner, *Judicial Independence in an Age of Democracy, Sectionalism, and War, 1835–1865*, *in* A HISTORY OF THE TENNESSEE SUPREME COURT 61, 84–85 (James W. Ely ed., 2002).

2. The Creation of the Tennessee Supreme Court

The 1796 Tennessee Constitution referred to superior and inferior courts but neither provided for a court of last resort, nor mandated the existence of any court.[18] In fact, courts only existed *if*, and *when*, and *as long* as the legislature desired.[19] The Tennessee Supreme Court was not created until 1809,[20] was not given appellate jurisdiction until 1819,[21] and did not become exclusively an appellate court until 1834.[22] Even then, the legislature maintained the power to abolish the supreme court since it was not created by the constitution.[23] It was not until 1835, when the constitution of 1834 was adopted, that the supreme court was given constitutional stature sufficient to save it from the control of or abolition by the legislative branch.[24]

3. The Constitution of 1834

Although the constitution of 1834 insured the existence of a state supreme court, by vesting the judicial power of the state in "one Supreme Court [and] in such Inferior Courts as the Legislature shall from time to time ordain and establish,"[25] the legislature retained its power to elect the judges.[26] Using virtually identical language to that used in the constitution of 1796, and under the same heading "Election of Judges," the 1834 constitution provided for judicial election by "joint vote of both Houses."[27] Immediately following that provision, the 1834 constitution provided that "Judges of the Supreme Court shall be *elected* for the term of twelve years."[28] Thus, the State persisted in its generic and broad use of the term "elect." The legislative election of judges to twelve-year terms was viewed as preferable because the judges "did not have to

---

18. TENN. CONST. art. V, § 1 (1796) ("The judicial power of the state shall be vested in such superior and inferior courts . . . as the legislature shall, from time to time, direct and establish."); *see* WILLIAMS, *supra* note 8, at 75.

19. TENN. CONST. art. V, § 1 (1796) ("The judicial power shall be vested in such superior and inferior courts . . . as the legislature shall, from time to time, direct and establish.").

20. Initially the Tennessee Supreme Court included two members who were joined for decision with a circuit judge who had heard the case below. WILLIAMS, *supra* note 8, at 75.

21. *Id.* Originally, the supreme court heard some appeals from circuit court but also maintained original jurisdiction in other cases. *Id.*

22. *Id.* at 75–76.

23. *Id.* at 76.

24. *Id.* at 76–77.

25. TENN. CONST. art. VI, § 1 (1834).

26. *Id.* art. VI, § 3.

27. *Id.* The only difference in the 1796 and 1834 provisions is that the 1796 provision used the phrase "joint ballot of both houses" while the 1834 provision used the phrase "joint vote of both Houses." *See* MCCLURE, *supra* note 9, at 424–25.

28. TENN. CONST. art. VI, § 3 (1834) (emphasis added); *see* MCCLURE, *supra* note 9, at 424.

fear insecurity for a reasonably long period,"[29] nor did they have to "engage in a struggle for official survival [which was described as] always a bitter experience for a judge-like judge."[30]

Although the 1834 constitution corrected what was regarded as the "most conspicuous deficit of the old Constitution"[31] by completing the proper distribution of governmental power into three separate and independent branches, it continued to assert legislative authority over the manner and details of judicial selection. Although five resolutions were offered at the 1834 Constitutional Convention to provide for the popular election of judges, each failed in turn.[32]

The newly created supreme court acknowledged the legislature's control, but exercised independence when cases required it. In 1836, for example, the court avowed that even though the legislature elected the judges, it was not the sovereign of the judiciary: "The fact that the constitution may prescribe that the mode of appointing the judges shall be by the legislature does not constitute the legislature the [courts'] constituent. . . .[T]he legislature is not sovereign; . . . it is not the constituent of the courts, nor are they its agents . . . ."[33]

4. The Constitutional Amendment of 1853

In the late 1840s the issue of judicial selection divided the two prevailing parties, the Democrats and the Whigs. In 1849, Tennessee elected Democratic governor William Trousdale. Trousdale advocated for a popularly elected judiciary based on the encouragement of Andrew Johnson, then a United States Congressman.[34] Johnson's support for a popularly elected judiciary was not principled. Rather, it was purely political, based on his belief that since the Whigs were in control, they would oppose any change that might reduce their power.[35]

---

29. WILLIAMS, *supra* note 8, at 46. Judges were originally elected to "hold their respective offices during their good behavior." TENN. CONST. art. V, § 2 (1796).

30. WILLIAMS, *supra* note 8, at 46. *But see* CALDWELL, *supra* note 11, at 149–50 (suggesting that the legislative control was because most of those in attendance at the 1796 Constitutional Convention were not lawyers and were not aware of the importance of an independent judiciary); Morton J. Horwitz, *The Emergence of an Instrumental Conception of American Law, 1780–1820*, *in* PERSPECTIVES IN AMERICAN HISTORY 287, 297–98 (Donald Fleming & Bernard Bailyn eds., 1971) (expressing yet another viewpoint, that because of the view that common law was static, little attention was paid to concerns regarding judicial independence).

31. JOSHUA W. CALDWELL, STUDIES IN THE CONSTITUTIONAL HISTORY OF TENNESSEE 109 (1st ed. 1895).

32. STATE OF TENNESSEE, JOURNAL OF THE 1834 CONSTITUTIONAL CONVENTION OF THE STATE OF TENNESSEE 27, 34, 39, 53–54, 96–97 (1834).

33. Jones' Heirs v. Perry, 18 Tenn. (10 Yer.) 44, 55 (1836).

34. Parks, *supra* note 6, at 626–28.

35. *Id.* (citing 1 THE PAPERS OF ANDREW JOHNSON 509 (L. Graf & R. Haskins, eds. 1970)); *id.* at 628 ("As an initial promoter of a popularly elected judiciary, Johnson probably recognized

When Johnson co-opted the media into the debate, they reframed the issue as one involving the public's competency to select their own judges.[36] With the issue framed as one of public trust, the Whigs became leery of opposing proposed judicial reforms. The cross-party support and media attention led to the 1851 legislative resolution to amend the constitution.[37] That year, both gubernatorial candidates campaigned in favor of the amendment.[38] In the summer of 1853,[39] the voters approved the amendment which provided that the "[j]udges of the Supreme Court shall be elected by the qualified voters of the State."[40]

5. The Constitution of 1870

Less than two decades later, Tennessee would undertake a complete revision of its constitution occasioned by the aftermath of the Civil War, the election of President Lincoln, and Reconstruction. The leaders at the 1870 Constitutional Convention believed that the changes they made to the existing constitution would be short-lived. The "nestor" of the Convention, Judge A. O. P. Nicholson, cautioned the delegates to only do what was absolutely necessary because "ten years from now all this must be done again."[41]

Paying heed to Judge Nicholson's warnings, it appears that the delegates did very little of consequence to the judicial article in 1870.[42] Proposals made to revise judicial selection, terms of office, and impeachment provisions were all rejected.[43] But inserted between the two sentences of article VI, section 3

---

the change as a potential way to root more Whigs out of public office.").

36. Parks, *supra* note 6, at 627; *see also* Huebner, *supra* note 17, 86–88.

37. For a record of the story of events leading to the amendment, see Huebner, *supra* note 17, at 85–89. The *Nashville Union* railed against legislative appointment of judges, likening the process to "species of log-rolling and bargaining," and argued that an independent judiciary was "necessary only in a monarchy . . . ." *Id.* at 86.

38. Parks, *supra* note 6, at 627. Under the 1834 Tennessee Constitution, governors were elected for two year terms. TENN. CONST. art. III, § 4 (1834).

39. The amendment was not submitted to the voters until 1853 because the constitution required that amendments be passed by two-thirds of the votes of two subsequent legislative sessions before being submitted to the voters. TENN. CONST. art. XI, § 3 (1834).

40. TENN. CONST. art. VI, § 3 (as amended in 1853); Laska, *supra* note 15, at 9.

41. CALDWELL, *supra* note 11, at 300. Some commentators suggest that "many of the [changes] deal with matters which are proper subjects of legislation, and not of constitutional regulation. . . . [T]hey are provisions which are too much dignified by places in the organic law and should be relegated to their proper rank, as stautes." CALDWELL, *supra* note 31, at 152–55 (listing article VI's changes as to the election and ages of judges as among the "unimportant" amendments better left to legislative acts).

42. The 1870 Constitutional Convention has been described as a "political expedient, designed to restore to citizenship and to the mastery of affairs, the majority of the white voters of the State, who had been disenfranchised by a minority party which the war had placed in power." CALDWELL, *supra* note 31, at 147.

43. During the debates, some members suggested that differences between the function

(the 1853 amendment that provided for the election of judges by the qualified voters) was this provision: "The Legislature shall have the power to prescribe such rules as may be necessary to carry out the provisions of section two of this article."[44] The referenced "section two" is the constitutional provision creating the supreme court.[45] Most importantly, the legislature was not given the power to prescribe rules relative to the election of circuit, chancery, or inferior court judges.[46] Apparently the legislature was not prepared to relinquish complete authority over the appellate judiciary.

6. The Legislature and the Courts Today

a. Constitutional and Statutory Provisions

This legislative entanglement with the judiciary, which began in the initial days of statehood, permeates Tennessee's constitutional history.[47] Moreover, the intertwinement remains vibrant today in the applicable constitutional provisions. Three separate constitutional articles contain provisions that relate to the Tennessee court system. Each in turn is linked with the legislature. The first and most basic provision, found in the Declaration of Rights, provides that the legislature may direct the manner and the courts in which suits may be brought.[48] The second set of provisions, set out in article VII, relates to state and county officers.[49] Section 4 of article VII grants the legislature the power to make provisions for "the election of all officers, and the filling of all vacancies not otherwise directed or provided by th[e] Constitution . . . ."[50]

The third and most significant collection of provisions are those set out in article VI, entitled the "Judicial Department."[51] The fifteen sections of the judicial article consign much to the legislature, including the power to create and abolish courts, to alter jurisdiction, and to set salaries and recusal standards.[52] By statute, the legislature has filled much of the void left by the

---

and locations of trial and appellate court judges might be a legitimate basis for differentiation in selection methods. *See* STATE OF TENNESSEE, JOURNAL OF THE 1870 CONSTITUTIONAL CONVENTION OF THE STATE OF TENNESSEE 124 (1870); *see also* CALDWELL, *supra* note 11, at 318–21.

44. TENN. CONST. art. VI, § 3 (1870).

45. *Id.* § 2.

46. *Id.* § 4.

47. As one commentator has noted, "the Tennessee practice of frequent legislative tinkering with the judiciary was begun early in the state's political life." LEWIS L. LASKA, TENNESSEE LEGAL RESEARCH HANDBOOK (1977).

48. TENN. CONST. art. I, § 17 (1870).

49. *Id.* art. VII, §§ 1–4.

50. *Id.* § 4.

51. *Id.* art. VI, §§ 1–15.

52. *See, e.g., id.* at § 1 (legislature may "ordain and establish" inferior courts and may "vest" jurisdiction in Corporation Courts as necessary); *id.* § 2 (legislature may restrict and

constitution. It has enacted legislation that establishes the terms of court,[53] the location of court houses,[54] and the site for appellate judges' chambers;[55] it has set the judges' salaries;[56] it has devised methods for replacing judges upon death, illness, or retirement;[57] and it has even mandated an annual training conference.[58]

Among the most significant of the legislature's enactments pertaining to the judiciary is the legislation creating the intermediate courts of appeal. Prior to their permanent creation, the legislature occasionally created temporary appellate panels to help reduce the supreme court's growing case load.[59] The legislature created the first lasting appellate court, and the predecessor to Tennessee's current Court of Appeals, by statute in 1895.[60] This court, the Court of Chancery Appeals, had purely appellate jurisdiction and its decisions were reviewed only for legal error.[61] In 1907, the number of judges on the intermediate appellate court was increased, its jurisdiction was enlarged, and its name was changed to the Court of Civil Appeals.[62] A subsequent name change and increase in membership in 1925 would create the Court of Appeals, today's intermediate court for appeals for civil cases.[63] More than forty years later, the legislature would follow the same procedure in creating the intermediate appellate court for criminal cases, the Court of Criminal Appeals.[64]

By the time the Tennessee Court of Criminal Appeals was created, the Tennessee Law Revision Commission, in conjunction with the Tennessee Bar

---

regulate the supreme court's appellate jurisdiction); *id.* § 3 (legislature may prescribe rules for the selection of supreme court judges); *id.* § 6 (legislature may remove judges from office); *id.* § 7 (legislature set judges' salaries); *id.* § 8 (legislature may change the jurisdiction of the circuit, chancery, and other inferior courts); *id.* § 11 (legislature shall set standards for relationship recusal and may provide for the appointment of special judges); *id.* § 15 (legislature shall divide the state into judicial districts and may provide for the appointment of justices of the peace).

53. TENN. CODE ANN. § 16-2-103 (1994 & Supp. 2007). For more than fifty years, the legislature also dictated the dates of court in each judicial circuit. *See* TENN. CODE ANN. §§ 16-207 to -255 (Supp. 1979).

54. TENN. CODE ANN. § 16-2-102 (1994 & Supp. 2007).

55. *Id.* § 16-5-113.

56. *Id.* § 8-23-103.

57. *Id.* §§ 17-2-116, 17-3-101.

58. *Id.* §§ 16-3-802, 17-3-105.

59. LASKA, *supra* note 47, at 71. The first such panel was known as the Arbitration Commission. Between 1873 and 1883, this body heard cases at the request of the parties and reported its findings to the supreme court. *Id.* In 1883, the Arbitration Commission was replaced with the Referees Commission which heard cases referred to it by the supreme court and then reported its findings back to that court. *Id.* By legislative dictate, neither Commission was permitted to publish its findings and its holdings were "without precedential value." *Id.*

60. 1895 Tenn. Pub. Acts 113; LASKA, *supra* note 47, at 71–72.

61. LASKA, *supra* note 47, at 72.

62. 1907 Tenn. Pub. Acts 232; LASKA, *supra* note 47, at 72 .

63. 1925 Tenn. Pub. Acts 690; LASKA, *supra* note 47, at 72–73.

64. 1967 Tenn. Pub. Acts 587; LASKA, *supra* note 47, at 73–74.

Association and a lay citizen's organization, was advocating an overhaul of the state's judicial system.[65] Their efforts to call a constitutional convention to institute reform were unsuccessful, but their voices were heard.

b. The 1971 Tennessee Plan

In 1971, a bipartisan Tennessee legislature provided for the merit selection of appellate judges. The legislative intent behind merit selection could not have been clearer. In passionate floor debates and an expressive preamble, the Tennessee legislature articulated the unambiguous purpose of merit selection: to secure a highly qualified, apolitical appellate bench.[66] The introductory section of the new legislation provided:

> It is the declared purpose and intent of the general assembly of Tennessee by the passage of this chapter to assist the governor in finding and appointing the best qualified persons available for service on the appellate courts . . . and to assist the electorate of Tennessee to elect the best qualified persons to said courts; to insulate the judges of said courts from political influence and pressure; to improve the administration of justice; to enhance the prestige of and respect for the said courts by eliminating the necessity of political activities by appellate justices and judges; and to make the said courts "nonpolitical."[67]

In the remaining provisions of the chapter, the legislature dictates the application process,[68] the nomination process,[69] the appointment process,[70] and the subsequent election process.[71] Consistent with the terminology used throughout Tennessee's history, the statute provides that every eight years, and in other years in the case of interim appointments, appellate judges who "seek *election*" must declare their "candidacy for *reelection*" by filing a written declaration of candidacy.[72] When declarations are timely filed, election officials are required to place, on the ballot, the question: "Shall (Name of Candidate) be *elected* and retained in office as (Judge) of the (Name of

---

65. Frank N. Bratton, *Report on Tennessee Citizens' Conference to Improve the Administration of Justice*, TENN. BAR J., May 1966, at 13, 13–16.

66. Parks, *supra* note 6, at 633–34 (citing Senator Edward C. Blank, II, Senate Debate of April 29, 1971, on tape at the Tennessee State Archives).

67. TENN. CODE ANN. § 17-4-102 (1994 & Supp. 2007) (originally enacted on May 12, 1971 at 1971 Tenn. Pub. Acts 510 and later codified as TENN. CODE ANN. § 17-701 (Supp. 1976)) (emphasis added).

68. *Id.* § 17-4-110 (previously codified as TENN. CODE ANN. § 17-710 (Supp. 1976)).

69. *Id.* § 17-4-102 (previously codified as TENN. CODE ANN. § 17-702 (Supp. 1976)).

70. *Id.* § 17-4-112 (previously codified as TENN. CODE ANN. § 17-712 (Supp. 1976)).

71. *Id.* § 17-4-114 (previously codified as TENN. CODE ANN. § 17-714 (Supp. 1976)) (for unexpired term) (emphasis added); *id.* § 17-4-115 (previously codified as TENN. CODE ANN. § 17-715 (Supp. 1976)) (for full term) (emphasis added).

72. *See* sources cited *supra* note 71.

Court)?"[73] If a majority of the voters of Tennessee "vote in favor of *reelecting*" the candidate, the candidate "is duly *elected* to office . . . and given a certificate of *election*."[74]

Nothing about these statutory prescriptions alarmed scholars of Tennessee constitutional history. They had always recognized that the details of judicial selection in Tennessee were left to the legislature's prerogative.[75] While the constitution gives the voters a say, "[t]he method of electing the judges is left to the General Assembly."[76]

> Article VI of the Tennessee Constitution remains as it was drafted in 1870. It provides few limitations on the legislative authority to create and alter the judicial system. . . .
>
> The Constitution provides for the election of judges for eight-year terms. However, candidates screened for qualifications and endorsed by the governor may be placed on the ballot for voter approval or rejection, and this method (the Missouri Plan) may be developed by the legislature in such a manner as to constitute election within the meaning of the Constitution.[77]

c. The 1974 Partial Repeal

In 1971 the legislature's desire to secure a highly qualified, apolitical appellate bench led to the passage of the Tennessee Plan under which intermediate appellate and supreme court judges stood for retention elections. If the impetus behind the 1971 passage of the Tennessee Plan was government at its best, the 1974 repeal of the Plan for supreme court justices was politics at its worst. The circumstances which led to the repeal, omitted from Professor Fitzpatrick's essay, are a significant aspect of the legislature's historic control over the judiciary.

After the passage of the Tennessee Plan by the bipartisan Tennessee legislature with little or no opposition, the Plan became the spoils of a highly partisan battle between the Republican governor and the Democratic legislature.[78] Just as the 1853 amendment was not the result of a principled choice between judicial selection methods, neither did the 1974 repeal reflect a rejection of merit selection. In the end, the repeal of the Plan for supreme court

---

73. TENN. CODE ANN. § 17-4-115(b)(1) (1994 & Supp. 2007) (emphasis added).

74. *Id.* at (d)(1) (emphasis added).

75. *See* Laska, *supra* note 15, at 20.

76. *Id.*

77. Thomas R. Van Dervort, *The Changing Court System*, in TENNESSEE GOVERNMENT AND POLITICS: DEMOCRACY IN THE VOLUNTEER STATE 55, 57 (John R. Vile & Mark Byrnes eds., 1998).

78. *See* Parks, *supra* note 6, at 634; Carl A. Pierce, *The Tennessee Supreme Court and the Struggle for Independence, Accountability, and Modernization, 1974–1998*, *in* A HISTORY OF THE TENNESSEE SUPREME COURT 270, 271–73 (James W. Ely ed., 2002).

justices had little to do with the judiciary; rather, the Plan was a pawn to be given away in exchange for other political favors.[79]

When a justice on the Tennessee Supreme Court died in 1972, the Appellate Court Nominating Commission interviewed applicants and submitted three names to Governor Winfield Dunn.[80] Although the governor appointed his choice in July, he made the appointment effective September 1.[81] A lawsuit was filed by a supreme court aspirant challenging the governor's appointment and the constitutionality of the Tennessee Plan.[82] Ultimately, the court invalidated the appointment, but upheld, without equivocation, the constitutionality of the Tennessee Plan.[83] The process began anew, but the governor's first choice withdrew from consideration.[84]

Following a second appointment process, some Democrats became concerned about the likely replacements for other justices who might retire,[85] and the effect that new justices might have on the court's appointment of the state Attorney General,[86] the composition of the State Building Commission,[87]

---

79. *See* Parks, *supra* note 6, at 634 (the repeal came "amid charges of vote-swapping on other key legislative issues"); *id.* at 615 ("The partisan manner in which the . . . issue was resolved and the superficiality of the debate on the part of both sides have, however, tended to obfuscate rather than illuminate the significant and difficult questions posed by various methods of selecting judges.").

80. Robert Keele, *The Politics of Appellate Court Selection in Tennessee: 1961–1981*, *in* THE VOLUNTEER STATE: READINGS IN TENNESSEE POLITICS 231, 234–39 (Olshfski & Simpson eds., 1985).

81. The death of the justice created a vacancy which was to be filled by the governor in accordance with the merit selection appointment process. The governor was authorized to appoint his nominee to fill out the deceased justice's unexpired term. Rather than effectuate the appointment immediately, the governor appointed his nominee effective September 1, 1972, the beginning of the next term of office. Because the governor's authority to appoint extended only to the period of the unexpired term, his appointment for the subsequent term was invalid. State *ex rel.* Higgins v. Dunn, 496 S.W.2d 480, 491 (Tenn. 1973).

82. *Id.* Despite the appointment by Governor Dunn of Thomas F. Turley, Jr., one of the three nominees from the Appellate Court Nominating Commission, Robert L. Taylor, announced that he was running for the position, campaigned, and received write-in votes in forty-six counties. He declared himself elected and was issued a certificate of election by the Secretary of State. He then took the oath of office before a Chancellor. *Id.* at 482. Meanwhile, the governor issued a commission of appointment to Turley. *Id.* at 482–83.

83. *Id.* at 490–91.

84. Keele, *supra* note 80, at 236.

85. *Id.* at 236–37. Before the death of Justice Larry Creson, all members of the court were Democrats. *See id.* at 232–33. Justice Creson's ultimate replacement was Justice Fones who categorized himself as an Independent. *Id.* at 236.

86. *Id.* at 237. Tennessee is unique in its provision that the state Attorney General is appointed by the supreme court. TENN. CONST. art. VI, § 5 ("An Attorney General and Reporter for the State, shall be appointed by the Judges of the Supreme Court and shall hold his office for a term of eight years.").

87. Keele, *supra* note 80, at 237. The Attorney General served on the State Building Commission, "described at the time as 'one of the last sources of patronage for the state's

and the construction of a medical college in East Tennessee.[88] Two days after the supreme court had upheld the constitutionality of the Tennessee Plan, the legislature began the process of repealing it as it applied to supreme court justices.[89]

Governor Dunn vetoed the repealing legislation. His public statement was unadulterated logic:

> I am aware of no reasons for repealing the provisions of the 1971 [A]ct as it relates to members of the Supreme Court and allowing the [A]ct to remain in effect for other appellate judges. There is no basis for the establishment of a dual system to fill appellate court vacancies. If the modified Missouri Plan embodied in the 1971 [A]ct is desirable as the method for filling appellate court vacancies, then it should be retained. If it is not, then it should be repealed in its entirety. I cannot, however, sanction the establishment of a dual system.[90]

In seeking to repeal the Tennessee Plan's application to supreme court justices and to override the governor's veto, no member of the legislature ever suggested that the Tennessee Plan was unconstitutional. Even though the constitutional challenge was fresh, no one asserted a legal basis for repealing the Plan. Rather, they claimed that because justices were more "visible" than their "regional" appellate counterparts, the "electorate could be trusted to make an informed choice between competing candidates."[91] The governor's veto of the repealing statute demonstrated that the executive branch viewed the Plan as constitutional. All of these circumstances indicate that the legislative and executive branch concurred with the supreme court's decision upholding the Plan.

---

weakened Democrats.'" *Id.* Hence, the position (and the politics) of the State Attorney General had dual importance to the legislature. *Id.* ("'A switch from a Democrat to a Republican Attorney General would shift the partisan balance on that Commission and give the Republicans control of that body.'").

88. *Id.* at 239.

89. 1974 Tenn. Pub. Acts, ch. 433, at 4; *see* Keele, *supra* note 80, at 236–37.

90. S. 88, 1st Sess., at 1523 (Tenn. 1973) (Message from Governor Dunn to the Secretary of State on May 4, 1973). Governor Dunn's concern over the dual system was shared by others:

> The state's present dual system, whereby trial judges and judges of the highest court are elected, while intermediate appellate judges are appointed, is a historical anomaly which should be corrected to reflect public interest as it is presently perceived.
>
> . . . For the sake of consistency, one system, preferably the merit plan, should be used in selecting all judges.

Parks, *supra* note 6, at 635.

91. Keele, *supra* note 80, at 238.

d. The First Constitutional Challenge

Their concurrence was well founded. The supreme court's decision in *State ex rel. Higgins v. Dunn*[92] was based on venerable principles of law and the undisputed historical facts. Perhaps it is the irrefutable logic of the *Dunn* decision that leads to one of the more disturbing arguments in Professor Fitzpatrick's essay, an argument that must be refuted before turning to the merits of the opinion. Professor Fitzpatrick argues with regard to both *Dunn* and *State ex rel. Hooker v. Thompson*[93] that the decisions have diminished precedential value because they were authored by "special" and not "regular" justices of the Tennessee Supreme Court.[94] This argument is particularly troublesome in light of Professor Fitzpatrick's assumed fortification of the constitution.

In both *Dunn* and *Thompson*, the Tennessee constitution disqualified the "regular" justices from hearing the cases. In another constitutional provision that proscribes institutional interference with the judiciary, the Tennessee constitution provides that

> [n]o Judge of the Supreme or Inferior Courts shall preside on the trial of any cause in any event of which he may be interested .... In case all or any of the Judges of the Supreme Court shall thus be disqualified, ... the Court, or the Judges thereof, shall certify the same to the Governor of the State, and he shall forthwith commission the requisite number of men, of law knowledge, for the trial and determination thereof.[95]

Since both cases related to the method by which supreme court justices would retain their offices, all of the "regular" justices were "interested" in the cases and were therefore disqualified from hearing them.[96]

Once appointed by the governor, it logically follows that "a special judge has all the power and authority of the regular judge."[97] Otherwise the appointment process would be in vain. Since 1835 the Tennessee law has provided that "[t]he special judges so commissioned shall ... have the same power and authority in those causes as the regular judges of the court."[98] The

---

92. 496 S.W.2d 480 (Tenn. 1973).

93. No. 01S01-9605-CH-00106, 1996 WL 570090, at *2 (Tenn. Oct. 2, 1996).

94. Fitzpatrick, *supra* note 2, at 489–90.

95. TENN. CONST. art. VI, § 11.

96. Harrison v. Wisdom, 54 Tenn. 99, 111 (1872) ("It is of the last importance that the maxim that no man is to be a judge in own case, shall be held sacred, and it is not to be confined to a cause in which he is a party, but applies to one in which he has an interest. This will be a lesson to all inferior tribunals to take care, not only that in their decrees they are not influenced by their personal interest, but to avoid the appearance of laboring under such an influence." (quoting Dimes v. Proprietors Grand Junction Canal, 3 House of Lords Cases, 759)).

97. *See, e.g.*, Harris v. State, 100 Tenn. 287 (1898); Brewer v. State, 74 Tenn. 198 (1880); Henslie v. State, 50 Tenn. 202 (1871).

98. TENN. CODE ANN. § 17-2-103 (1994 & Supp. 2007).

Dunn and *Thompson* opinions—and any opinions rendered by a special supreme court—are entitled to the same weight as an opinion by the "regular" justices.[99]

The court in *Dunn* upheld the constitutionality of the Tennessee Plan without a struggle based on established principles of constitutional law. The court first recognized the inherent limited purpose of a constitution: to provide a broad outline of the organization and function of government.[100] Constitutions do not "provide the details for exercising governmental power .... [T]hey are not intended to establish all the law which, from time to time, may be necessary to meet changing conditions."[101] Article VI, section 3 is not self-executing. The executory details, which are not provided in the constitution, are left to the legislature. This legislative deferral is not only consistent with Tennessee tradition, it is also specifically addressed in the constitution.[102] The legislature assumed the duty and set forth the election details in the statutes.[103]

The *Dunn* court also applied traditional rules of construction to the terms used in the constitution, stating that

> [t]he Constitution of Tennessee does not define the words, "elect," "election," or "elected" and we have not found nor have we been referred to any provision of the Constitution or of a statute or to any decision of one of our appellate courts defining these words. ... [Since the Constitution in at least three instances refers to referenda and other methods of ratification as

---

99. Not only does the assertion that the "special" judges were not qualified to render a decision on a matter of constitutional importance ignore the law, it is also wholly uninformed. Among those who sat as members of the special appellate courts in these cases were Tennessee legal giants. *See, e.g.*, DeLaney v. Thompson, No. 01A01-9806-CH-00304, 1998 Tenn. App. LEXIS 486, at *23 (Tenn. Ct. App. July 16, 1998) ("The veteran judges making up this special court have served as judges at various times and places by every procedure known to the law: appointment, election, interchange, retention, litigant selection, bar election and special designation."). The special judges who sat on the *DeLaney* appellate panel—Judge William S. Russell, Judge Joe D. Duncan, and Judge Samuel L. Lewis—had more than seventy years of combined legal experience. The special justices who sat on the *DeLaney* supreme court included lawyers from all practice areas with nearly a century of combined legal experience. The special justices in *Thompson* included a former chief justice of the Tennessee Supreme Court. State *ex rel.* Hooker v. Thompson, No. 01S01-9605-CH-00106, 1996 WL 570090 (Tenn. Oct. 2, 1996).

100. State *ex rel.* Higgins v. Dunn, 496 S.W.2d 480, 487 (Tenn. 1973). *See generally* MCCLURE, *supra* note 9, at 25 ("The people, the fountain of all power, have delegated their sovereignty to their state governing agencies, the nature and organization of which are set forth in the constitutions ....").

101. *Dunn*, 496 S.W.2d at 487.

102. Article VII, section 4 provides that "[t]he election of all officers, and the filling of all vacancies not otherwise directed or provided by this Constitution, shall be made in such manner as the Legislature shall direct." TENN. CONST. art. VII, § 4. Article VI, section 3 provides that "[t]he Legislature shall have power to prescribe such rules as may be necessary to carry out the provisions of section two of this article." TENN. CONST. art. VI, § 3.

103. *Dunn*, 497 S.W.2d at 487–88.

> election], it cannot be said that [the 1971 statute] is unconstitutional because the elections therein provided for are limited to approval or disapproval.[104]

A few months after the supreme court upheld the constitutionality of the Tennessee Plan, the legislature, pursuant to its constitutional authority, overrode the governor's veto leaving Tennessee with a dual system for selecting appellate court judges. This incongruity would remain until 1994 when the legislature would enact a modified, incomparable plan for electing and evaluating all of Tennessee's appellate court judges.[105]

e. The 1977 Limited Constitutional Convention

While it is true that the citizens of Tennessee rejected a constitutional amendment that would have specified the details for electing appellate judges, it is disingenuous to suggest, as Professor Fitzpatrick does, that the 1977 vote somehow affects the constitutionality of the 1994 legislation. The essay's incomplete discussion of Tennessee's 1977 Limited Constitutional Convention creates a misimpression that the professor uses to buttress many of his arguments. The omitted details are discussed below.

The 1977 Limited Constitutional Convention was convened[106] to deal with a multitude of state problems, more than ever before undertaken in a meeting of its kind. While the issues were many, and varied, the primary impetus for the Convention was the state's dire fiscal situation compounded by a constitutional ceiling on interest rates.[107] "Although other groups had been seeking to change

---

104. *Id.* at 489. The court listed dozens of statutory provisions that used the word "elect" to describe various selection methods. *Id.* at 489 n.1.

105. *See infra* text accompanying notes 134–44.

106. After adjournment, it was determined that the entire 1977 Constitutional Convention was actually invalid because the governor had not signed the act calling for the convention. Crenshaw v. Blanton, 606 S.W.2d 285, 289 (Tenn. Ct. App. 1980) (holding that the constitution "does require the signature of the Governor on a measure submitting to the voters the question of calling a constitutional convention"). Crenshaw had filed a chancery action challenging the validity of the act providing for the convention. On appeal, the court of appeals found a constitutional deficiency, but was

> unwilling at this late date to invalidate the amendments to the Constitution which have been proposed by a convention called upon approval of the voters of the State who also have given final approval to the amendments. Judicial interference with the orderly framework of government as approved by the voters of the State is simply not justified by an omission which cannot be said to have interfered with the free exercise of the rights of the people of the State to change the form of their government.

Id. at 290.

107. Lewis L. Laska, *The 1977 Limited Constitutional Convention*, 61 TENN. L. REV. 485, 486–88 (1994). The primary promoters of the 1977 Constitutional Convention were the State Labor Council, the Tennessee Education Association, the Tennessee County Services Association (a lobbying group for county officials), the Tennessee Municipal League, and the Tennessee Congress of Parents and Teachers. *Id.* at 488 n.12. While none of the promoting

the constitution, lobbying by the financial industry (which included mortgage lenders and allies in the real estate industry) was the prime cause of the 1977 Limited Constitutional Convention."[108]

Although judicial reform was not a catalyst for the Convention, those who favored court reform supported the call.[109] The court reformers were not concerned about judicial selection methods. Rather, they were concerned about the overall inefficiency and dysfunction of the Tennessee court system. These concerns, documented in 1971 by the Institute of Judicial Administration,[110] grew out of Tennessee's antiquated and jumbled court system.[111] The Tennessee Bar Association and the Tennessee Law Revision Commission had sought a convention to deal with judiciary reform in the mid-1960s. In 1968, the legislature agreed,[112] but the people defeated the call for the convention.[113] When, in 1977, it became likely that a constitutional convention would be held, efforts to modernize the Tennessee court system began anew.

The 1977 call for convention included revisions to six of the eleven articles of the Tennessee constitution[114] For all of the articles, except one, the particular section sought to be revised was specified.[115] But the call relative to

---

groups or lobbyists promoted change in the state judiciary, state Supreme Court Justice Joe Henry is reported to have desired an opportunity to modernize the Tennessee court system. *Id.* at 494–95.

108. *Id.* at 488–89 (footnote omitted).

109. Pierce, *supra* note 78, at 297.

110. *See* JOHN M. SCHEB, II & STEPHEN J. RECHICHAR, THE POLITICS OF JUDICIAL MODERNIZATION: THE CASE OF THE TENNESSEE COURT SYSTEM 43–46 (The Univ. of Tenn., Bureau of Public Administration 1986). In 1971 the Tennessee Judicial Council created the Institute of Judicial Administration to study and recommend court reform measures in Tennessee. *Id.* at 46. The Institute documented "five serious shortcomings" of the Tennessee court system: "(1) Problems of multi-county districting associated with the dual law/equity system . . . ; (2) 'a lack of functional mobility among the judges' . . . and resulting case load inequities . . . ; (3) Judge shopping tendencies arising from 'an overdose of concurrent jurisdiction' . . . ;" (4) Problems of lack of uniformity in procedure; and "(5) An excess of judges at every level except the Supreme Court." *Id.* at 45–46. The study's only other reflection on the appellate courts was that the specialization in the appellate courts had produced a "high quality output with the benefits especially pronounced at the intermediate appellate level." *Id.* at 45.

111. *See* Frederic S. Le Clercq, *The Tennessee Court System*, 8 MEM. ST. U. L. REV. 185, 425 (1978); Van Dervort, *supra* note 77, at 56 (citing the hodgepodge court system and "judge shopping" which created caseload inequities as the "major problem" facing the Tennessee courts).

112. *See* 1968 Tenn. Pub. Acts 37.

113. *See* JOE C. CARR, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1969–1970, at 254–58 (1969).

114. 1976 Tenn. Pub. Acts ch. 848, § 1; *see* Governor Ray Blanton, *Proclamation by the Governor*, *in* THE LIMITED CONSTITUTIONAL CONVENTION OF 1977, STATE OF TENNESSEE, THE JOURNAL OF THE DEBATES OF THE LIMITED CONSTITUTIONAL CONVENTION OF 1977 (hereinafter *Proclamation by the Governor*).

115. 1976 Tenn. Pub. Acts ch. 848, § 1. The call listed article II, sections 8, 15, 18, and

the judicial article did not designate any particular section, but provided for consideration of the entire article.[116]

When the delegates had concluded the longest and most expensive convention in Tennessee's history, thirteen proposed amendments were submitted to the voters for approval. Most of the proposed amendments offered a single proposal to the voters.[117] But the amendment concerning the judicial department contained sixteen separate proposals,[118] consisting of more than 1,500 words. The amendments affected virtually every person serving in the justice system—judges, clerks, district attorneys, the state Attorney General, constables, and jurors.[119] Yet despite the number of separate proposals and the various constituencies affected, the voters were not allowed to vote separately on the provisions but were required to either accept or reject the amendment as a whole.

So complex were the changes to the judicial article that many of the delegates professed confusion over what was included in the final proposal.[120] In addition, the proposal omitted, perhaps by political design,[121] a constitutional provision that had protected judicial salaries from legislative tinkering during a judge's term of office.[122] This omission, coupled with the requirement of a

---

24; article III, sections 4 and 18; article IV, section 1; article VII, section 1 and 2; article XI, section 7, 11, 12, and 14. For article VI, the judiciary article, the call specified the entire article: "Article VI, consisting of Sections 1 through 15." *Id.*

116. *Id.*

117. *See Proclamation by the Governor*, *supra* note 114. For example, Proposal 1 required the voters to vote on whether the constitutional prohibition on interracial marriage should be repealed. *Id.* Proposal 3 called for the repeal of the constitutional homestead exemption. *Id.* Proposal 4 allowed a governor to serve two consecutive terms. *Id.* Proposition 7 allowed voters age eighteen and over to vote. *Id.* Proposal 10 deleted the constitutional maximum interest rate and allowed the legislature to set the maximum rate. *Id.*

118. Among the proposed changes to the judicial article were a complete restructuring and renaming of the court system; a combination of intermediate appellate courts; a reduction in the judicial term of office; the creation of a new "Superior Court"; the abolition of the Chancery Court; the creation of a state-wide General Sessions Court with "uniform" jurisdiction; the creation of a Court of Discipline and Removal; a change in the method of selection and the term of office of the State Attorney General; a reduction in the term of office of District Attorney Generals; the creation of a state-wide indigent defense system; and the elimination of clerks and masters. *See id.*

119. *See id.*

120. *See* Laska, *supra* note 107, at 549; Van Dervort, *supra* note 77, at 56.

121. The proposed constitutional prohibition on altering a judge's salary during the term of office was ultimately tied to a similar provision protecting the salaries of district attorneys and public defenders. This created concern and controversy, resulting in its deletion and ultimate omission from the proposal. Laska, *supra* note 107, at 550.

122. Van Dervort, *supra* note 77, at 56. Since the founding of the Republic the issue of removing the control of judges via reduction in salaries during terms of office had been prominent. DECLARATION OF INDEPENDENCE para. 11 (U.S. 1776) ("[The King] has made Judges dependent on his Will alone for the Tenure of their Offices, and the Amount and Payment of their Salaries."); *see* SUSAN B. CARBON & LARRY C. BERKSON, JUDICIAL RETENTION

unitary vote, assured that the proposed amendments to the judicial article would fail.

Those who had initially supported the inclusion of the article in the call for convention, including Chief Justice Joe Henry and the Tennessee Bar Association, vehemently opposed its passage.[123] Chief Justice Henry decried the interference with the independence of the judiciary: "It is incredible that in the last three quarters of the twentieth century a constitutional convention would make judges dependent upon the good will of the legislature for their compensation."[124] In the words of the chief justice, the amendment would assure a "devitalized, disorganized, demoralized, and subservient judiciary"[125] and should be rejected.

It was not only the vocal opposition to the amendment that led to its failure, but also its lack of support.

> The judicial article failed because there was no strong ally in support of it, and many discordant voices against it. The loudest was Chief Justice Joe Henry. . . . [H]is conclusion was pure Justice Henry: "I hope the people of Tennessee will consign the proposed judicial article to the oblivion it so richly deserves."

Despite their overwhelming support of judicial reform, the Tennessee Bar Association ultimately opposed the new judicial article. The Bar Association believed that the article adversely affected the traditional notions of checks and balances and separation of power of separate and equal branches of government. "In the end, the judicial article was abandoned by those whom it would have influenced the most: the supreme court (at least Justice Henry), the trial court judges, the court clerks, and even the nonlawyer general sessions judges."[126]

The legislature ultimately used its plenary powers to adopt many of the progressive court revisions contained in the rejected amendment to the judicial article.[127] The legislature reorganized the trial court system,[128] created a state-

---

ELECTIONS IN THE UNITED STATES 1 (1980).

123. Laska, *supra* note 107, at 570–71.

124. *Id.* at 551. Justice Henry was always a master of language, but his remarks to the Tennessee Municipal League in opposition to the amendment may be among his finest. In remembering a phrase used by Governor Gordon Browning, "Stand still, little pig, while I gut you," Justice Henry pronounced, "I won't be gutted!" SCHEB & RECHICHAR, *supra* note 110, at 58 (quoting Kirk Loggins, *Henry Launches Effort to Kill Judicial Article*, NASHVILLE TENNESSEEAN, Jan. 7, 1978, at 1).

125. SCHEB & RECHICHAR, *supra* note 110, at 60 (quoting Kirk Loggins, *Henry Launches Effort to Kill Judicial Article*, NASHVILLE TENNESSEEAN, Jan. 7, 1978, at n.45).

126. *Id.* at 570; *see* Van Dervort, *supra* note 77, at 55–57. According to Van Dervort, "[i]n the end the judicial article proposal failed because there was no strong lobby in support of it and many discordant voices, primarily those of the Chief Justice and the Tennessee Bar Association, against it." Van Dervort, *supra* note 77, at 57.

127. Van Dervort, *supra* note 77, at 58.

wide public defender system,[129] gave the supreme court extensive rulemaking powers,[130] and increased uniformity in the General Sessions Court.[131] And in 1994, the legislature revised the Tennessee Plan to assure the quality of the Tennessee appellate bench.

f. The 1994 Tennessee Plan

When the legislature revised the Tennessee Plan, it not only reinstated retention elections for supreme court justices, it also fashioned a merit election system that was unique to Tennessee. The legislature restated its clear and unambiguous purpose, first outlined in 1971: to secure a highly qualified apolitical appellate bench.[132] The Tennessee Plan was designed to assist the governor in the initial appointment and the citizens in the subsequent elections.[133] This time, the legislature added a feature to the Tennessee Plan that made the Plan uniquely able to "assist the electorate" in "elect[ing] the best qualified persons to the court."

The added dimension of the 1994 Tennessee Plan is a judicial performance evaluation program[134] by which court personnel, lawyers, and other judges evaluate the performance of Tennessee's judges.[135] In addition, the program includes self-evaluation and the opportunity for judges to discuss and reflect on their own strengths and weaknesses.[136]

The overriding purpose of the evaluation program is to "improve[e] the administration of justice in Tennessee . . . by instituting a program of continuous self-improvement . . . that empowers the judges, with the assistance of their peers, to enhance and to broaden their own judicial skills."[137] By assisting judges in identifying areas in which they need to boost their judicial skills, the program improves the overall quality of the Tennessee bench.[138]

---

128. TENN. CODE ANN. §§ 16-2-101 to -520 (1994 & Supp. 2007).

129. *Id.* §§ 8-14-201 to -212.

130. *Id.* § 16-3-401.

131. *Id.* §§ 16-3-501 to -504.

132. *Id.* §17-4-101.

133. *Id.* ("It is the declared purpose and intent of the general assembly . . . to assist the governor in finding and appointing the best qualified persons available for service on the appellate courts of Tennessee, and to assist the electorate of Tennessee to *elect* the best qualified persons to the courts . . . ." (emphasis added)).

134. At the time that Tennessee adopted its judicial evaluation program only nine other states in the country had similar programs providing for the evaluation of their judges. *See* Marla N. Greenstein, Dan Hall, and Jane Howell, *Improving the Judiciary through Performance Evaluations*, *in* THE IMPROVEMENT OF THE ADMINISTRATION OF JUSTICE (American Bar Association 7th ed. 2002); JUDICIAL PERFORMANCE EVALUATION HANDBOOK 3 (American Bar Association, 1996).

135. TENN. SUP. CT. R. 27, § 1.

136. *Id.* at § 1.04

137. *Id.* at § 1.03

138. *Id.* at § 1.02.

But for appellate judges, the purpose of evaluation is deeper than the mere desire for individual self-improvement. The program achieves the legislative purpose of "assist[ing] the electorate"[139] by providing information that "promote[s] informed retention decisions."[140] Each appellate judge standing for retention election is evaluated in order to inform the electorate about the judge's performance on the bench. This enables the voters to cast a more knowledgeable vote.[141]

By adopting the judicial evaluation program as part of the Tennessee Plan, the Tennessee legislature demonstrated a true commitment to assuring a quality appellate bench. For the last fourteen years, appellate judges in Tennessee have been appointed by the governor, evaluated by the Judicial Evaluation Commission, and elected by the voters.[142] The system has not only provided a unique model for other states; it has also produced a diverse and qualified appellate bench removed from partisan politics.

---

139. *See supra* note 133.

140. TENN. SUP. CT. R. 27, § 1.05 ("In addition to its primary purpose of self-improvement, the Judicial Performance and Evaluation Program must provide information that will enable the Judicial Evaluation Commission to perform objective evaluations and to issue fair and accurate reports concerning the appellate judges' performances.").

141. TENN. CODE ANN. § 17-4-201(a)(1) (1994 & Supp. 2007) provides that "[t]he purpose of the [judicial evaluation] program shall be to assist the public in evaluating the performance of incumbent appellate court judges." To this end, " [t]he judicial evaluation program shall require publication and disclosure of a final report." *Id.* § -201(c)(1). The report is publicly available and is published in six daily newspapers preceding the election. *Id.*

142. Professor Fitzpatrick complains that the Commission has recommended retention for "every single one" of the sixty-six judges that have been evaluated since 1994. Fitzpatrick, *supra* note 2, at 484. His critique implies that some of Tennessee's judges did not deserve either a positive evaluation or retention. An equally plausible explanation is that Tennessee's merit selection and evaluation system has produced good judges who do their jobs well and deserve to continue to do so. His criticism is also irrelevant—even before Tennessee moved to a merit selection system for its appellate judges, few appellate judicial races were contested and even fewer incumbents lost their seats. *See* Harry Phillips, *Our Supreme Court Justices*, 17 TENN. L. REV. 466, 468 (1942) ("Indeed, the caliber of Tennessee's appellate judges has been such that the State has seen few contests for the highest bench.").

g. The Second[143] Constitutional Challenge[144]

In 1996, a perennial litigant in Tennessee state and federal courts[145] filed suit to enable himself to run for a seat on the supreme court.[146] His attack on the Tennessee Plan in *State ex rel. Hooker v. Thompson* was based initially on the fact that the sitting justice[147] who was on the ballot for retention had not been evaluated by the Judicial Evaluation Commission.[148] Ultimately, a special

---

143. These two constitutional challenges and two others, one challenging the application of the system when a judge was not evaluated, *see infra* note 147, and the other dismissed by the federal court in March of this year, Johnson v. Bredesen, No. 3:07-0372, 2008 U.S. Dist. LEXIS 19738 (M.D. Tenn. Mar. 13, 2008), constitute the "several cases" that have caused the Tennessee Plan to be "mired in litigation." *See* Fitzpatrick, *supra* note 2, at 475.

144. I was a named defendant in the second suit challenging the constitutionality of the Tennessee Plan brought by Mr. John Jay Hooker. I am not unique; Mr. Hooker has sued every sitting Tennessee Supreme Court justice, the members of the Tennessee Judicial Selection Commission, at least three governors, and several State Attorney Generals, as well as at least two United States Senators, the mayor of Nashville, and the Federal Election Commission. *See* sources cited *infra* note 145.

145. With two exceptions, all of the litigation concerning the administration of the Tennessee Plan has been filed either by or on behalf of Mr. John Jay Hooker. In addition to these suits over the state judicial selection system, Mr. Hooker often challenges campaign finance systems in federal elections. *See, e.g.*, Hooker v. Fed. Election Comm'n, 21 F. App'x 402 (6th Cir. 2001); Hooker v. Thompson, 21 F. App'x 342 (6th Cir. 2001); Hooker v. Sasser, 893 F. Supp. 764 (M.D. Tenn. 1995); Hooker v. Alexander, No. M2003-01141-COA-R3-CV, 2005 Tenn. App. LEXIS 304 (Tenn. Ct. App. May 20, 2005).

146. At the time of this lawsuit, Mr. Hooker was not qualified to serve as a justice because "he failed to meet the requirement that a candidate for Supreme Court Justice must be an attorney licensed to practice law in Tennessee . . . ." State *ex rel.* Hooker v. Thompson, No. 01S01-9605-CH-00106, 1996 WL 570090, at *2 (Tenn. Oct. 2, 1996). Mr. Hooker's law license had been suspended for his failure to comply with continuing legal education requirements. *Id.* at *1 n.4. In addition, Mr. Hooker resided in the Middle Grand Division of the State and could not qualify for the seat because two sitting justices, Justice Drowota and Justice Birch, also resided in that Division. *Id.*; *see* TENN. CONST. art. VI, § 2 ("The Supreme Court shall consist of five Judges, of whom not more than two shall reside in any one of the grand divisions of the State.").

147. To state the obvious, that sitting justice was me.

148. A similar unsuccessful attack on the Plan was mounted by at attorney seeking to run for the Tennessee Court of Appeals, Middle Division, in 1998. Judge Henry Todd advised the Judicial Evaluation Commission that he did not intend to seek election at the end of his term. As a result the Commission did not perform an evaluation of Judge Todd. An aspirant for Judge Todd's seat sought and received injunctive relief against the application of the Tennessee Plan, claiming the Plan inapplicable since Judge Todd was not evaluated. The Davidson County Chancery Court decision granting relief was reversed by a special panel of the Tennessee Court of Appeals, which held the Tennessee Plan constitutional based upon rules of statutory and constitutional construction and the *Dunn* precedent. DeLaney v. Thompson, No. 01A01-9806-CH-00304, 1998 Tenn. App. LEXIS 486 (Tenn. Ct. App. July 16, 1998). The special court of appeals's decision was in turn reversed by a special supreme court which, based on equally long-standing principles, found it unnecessary to address the issue of the constitutionality of the

supreme court[149] assessed and upheld the constitutionality of the Tennessee Plan.[150]

The decision was not unexpected. Although the legislature had added an evaluation program to the Tennessee Plan, the remaining provisions were identical to those upheld by the court in 1973. While the 1973 precedent was a basis for the court's analysis,[151] it was not the sole foundation. The court also relied upon fundamental principles of statutory construction and constitutional interpretation essential to analyzing any constitutional challenge.

B. Fundamental Principles of Statutory Construction and Constitutional Law

The second cornerstone omitted from Professor Fitzpatrick's discussion is consideration of basic principles of statutory construction and constitutional interpretation. These principles, discussed below, are essential to evaluating the constitutionality of any legislative act. When properly utilized to analyze the Tennessee Plan, the principles lend further support to the conclusion that the Tennessee Plan is constitutional.

1. Presumption of Constitutionality

The most basic principle of statutory construction requires that courts indulge "every presumption" in favor of constitutional validity.[152] Statutes are presumed to be constitutional because it is within the province of the legislature to prescribe law by which society is governed.

The principle of presumed constitutionality requires that courts indulge every presumption in favor of upholding a legislative enactment. Every doubt as to the viability of a statute must be resolved in favor of constitutionality. So strong is the presumption that when two possible interpretations exist, the one that sustains constitutionality is imposed over the other.[153] Unless a plain and

---

Tennessee Plan. DeLaney v. Thompson, 982 S.W.2d 857, 858 (Tenn. 1998) ("It is the duty of all courts, including the Supreme Court, to pass on a constitutional question only when it is absolutely necessary for the determination of the case and of the rights of parties to the litigation.").

149. *See supra* text accompanying notes 93–98.

150. State *ex rel.* Hooker v. Thompson, No. 01S01-9605-CH-00106, 1996 WL 570090 (Tenn. Oct. 2, 1996).

151. *Id.* at *3 ("The issue of whether yes/no retention elections violate the Constitution of Tennessee has previously been decided by the Tennessee Supreme Court in the case of *State ex rel. Higgins v. Dunn*, and no compelling reason has been given to persuade this Court that it should disturb that ruling." (citation omitted)).

152. *See, e.g.*, Bank of State v. Cooper, 10 Tenn. (2 Yer.) 599, 608 (1831).

153. *See, e.g.*, Kirk v. State, 150 S.W. 83, 85 (Tenn. 1911); Cole Mfg. Co. v. Falls, 16 S.W. 1045, 1046 (Tenn. 1891).

unambiguous interpretation compels the conclusion that a statute violates the constitution, the statute must be upheld.[154]

2. Construction to Uphold Constitutionality

In addition to the presumption of constitutionality that adheres to all statutes, a court must construe a statute so as to preserve constitutionality.[155] If a statute lends itself to more than one construction, the construction that upholds constitutionality must be applied. A statute must not be declared unconstitutional if "it is possible to avoid doing so."[156] If doubt arises as to the meaning of the provision, a court must "harmonize [the conflicting] portions and favor the construction which will render every work operative rather than one which would make some words idle and meaningless."[157]

3. Legislative Objectives

When a statute's constitutionality is challenged, the court must look at the goals intended by the legislature and not the particular language used.

> In construing statutes, we look at the objects aimed at by the Legislature, and not to the particular verbiage, in which a statute, in some of its parts, may be expressed. If the real object aimed at is within legislative competency, and can be clearly seen from the whole statute taken together, the history of the prior legislation upon the same subject, the Court will not be turned aside by particular expressions, which, taken by themselves, might seem to indicate that the Legislature was assuming to transcend its constitutional power, but will give effect to the will of the Legislature thus discovered.[158]

C. The Constitutionality of the Tennessee Plan

1. The Principle of *Stare Decisis* in General

These fundamental rules of constitutional law and statutory construction viewed in light of Tennessee's constitutional history lead to the inescapable conclusion reached by the *Dunn* and *Thompson* courts that the Tennessee Plan does not violate the Tennessee constitution. These decisions are dismissed too summarily by Professor Fitzpatrick. His essay discounts the importance of

---

154. *See, e.g.*, Arrington v. Cotton, 60 Tenn. 316 (1872); Smith v. Normant, 13 Tenn. (5 Yer.) 271 (1833).

155. *See, e.g.*, Consolidated Enters., Inc. v. State, 263 S.W. 74, 75 (Tenn. 1924) (describing it as the "primary" rule); Turner v. Eslick, 240 S.W. 786, 789 (Tenn. 1921) (same).

156. Knoxville Power & Light Co. v. Thompson, 276 S.W. 1050, 1051 (Tenn. 1925).

157. Shelby County v. Hale, 292 S.W.2d 745, 748–49 (Tenn. 1956).

158. *Arrington*, 60 Tenn. at 319–320.

judicial precedent in two ways. In general, the essay disregards the principle of *stare decisis*. In particular, the essay erects illogical arguments to challenge the principle's application to the *Dunn* and *Thompson* decisions.

> The rule of stare decisis is peculiarly applicable in the construction of written constitutions. . . . "A cardinal rule in dealing with written instruments is that they are to receive an unvarying interpretation, and that their practical construction is to be uniform. A constitution is not to be made to mean one thing at one time, and another at some subsequent time, when the circumstances may have so changed as perhaps to make a different rule in the case seem desirable. A principal share of the benefit expected from written constitutions would be lost, if the rules they established were so flexible as to bend to circumstances or be modified by public opinion."[159]

The final arbiter of the Tennessee constitution has twice upheld the Tennessee Plan against constitutional challenges. Several United States District Courts and the Sixth Circuit Court of Appeals have relied on the supreme court's holdings in dismissing countless actions challenging the Plan.[160] The decisions upholding the Tennessee Plan have uniformly held that a retention election satisfies the constitutional requirement that the justices of the supreme court "be elected by the qualified voters of the State."[161] Thus, no challenge to the constitutionality of the Tennessee Plan has ever been successful.

2. The Principle of *Stare Decisis* applied to *Dunn* and *Thompson*

In addition to its general disregard for the importance of *stare decisis*, the essay floats specious arguments against the principle's application to the *Dunn* and *Thompson* decisions. The first taunt, addressed previously in this article, is that the decisions are not entitled to the effect of *stare decisis* because a "majority of regular justices" did not render the decisions. In addition to ignoring the constitutional provision requiring judicial disqualification,[162] the claim defies common sense. Advanced to its logical conclusion, Professor Fitzpatrick's point would create decisional chaos. Either "regular" judges would be forced to decide matters in which they had an interest, thereby

---

159. McCulley v. State (The Judges' Cases), 53 S.W. 134, 139–40 (Tenn. 1899); *see also* State *ex rel.* Pitts v. Nashville Baseball Club, 154 S.W. 1151, 1154–55 (Tenn. 1913).

160. Judges Higgins, Donald, and Campbell have all dismissed cases in which Mr. Hooker has claimed a property interest either in the right to run for justice or in the right to vote in a popular election of appellate judges. A threshold question in each case has been whether the Tennessee Plan violates state constitutional law. *See* Hooker v. Anderson, 12 F. App'x 323 (6th Cir. 2001); Hooker v. Thompson, 21 F. App'x 342 (6th Cir. 2001); Hooker v. Burson, No. 96-6030, 1997 U.S. App. LEXIS 2682 (6th Cir. Feb. 12, 1997); Johnson v. Bredesen, No. 3:07-0372, 2008 U.S. Dist. LEXIS 19738 (M.D. Tenn. Mar. 13, 2008).

161. TENN. CONST. art VI, § 3.

162. *Id.* § 11.

creating "good" precedent, or substitute judges would render a decision that was of no value.

The second jab is aimed only at the *Thompson* decision and claims that the decision has no precedential value because it was not published.[163] This argument relies upon a supreme court rule that specifies that certain intermediate appellate decisions will have "no precedential value."[164] But the unpublished *Thompson* decision does not fall in that category. Moreover, consistent with the essay's general disregard for *stare decisis*, the argument ignores the fact that *Thompson* relied on the precedent established twenty-five years earlier in *Dunn*.

3. The Essay's Four Remaining Arguments

This paper's earlier discussions of the legislature's historic involvement with the judiciary, Tennessee's constitutional history, and fundamental principles of constitutional law and statutory construction have exposed the fallacy of most of the arguments against the constitutionality of the Tennessee Plan. This section makes additional observations relative to Professor Fitzpatrick's four remaining arguments: that the legislature cannot give the governor the authority to appoint judges except when a midterm vacancy occurs; that retention elections are not "elections"; that retention races cannot be reconciled with democracy; and that the electorate's rejection of the 1977 constitutional amendment is evidence that the Tennessee Plan in unconstitutional.

a. The Legislature May Authorize the Governor to Fill all Appellate Court Vacancies

Professor Fitzpatrick claims that the Tennessee Plan is unconstitutional "to the extent [it] permits the governor to appoint a new judge to a position created when the previous judge served [a] full term . . . ."[165] If the legislature may empower the governor to fill end-of-term vacancies, Professor Fitzpatrick contends that the vacancy provision would nullify the election provision.[166] This argument fails for two reasons. First, it is contradicted by the plain

---

163. *See* Fitzpatrick, *supra* note 2, at 488–89, 489 n.143.

164. TENN. SUP. CT. R. 4(E)(1) ("If an application for permission to appeal is hereafter denied by this Court with a "Not for Citation" designation, the opinion of the intermediate appellate court has no precedential value."); *cf.* TENN. SUP. CT. R. 4(G)(1) ("An unpublished opinion shall be considered controlling authority between the parties to the case when relevant under the doctrines of the law of the case, res judicata, collateral estoppel, or in a criminal, post-conviction, or habeas corpus action involving the same defendant. Unless designated "Not For Citation," "DCRO" or "DNP" pursuant to subsection (F) of this Rule, unpublished opinions for all other purposes shall be considered persuasive authority.").

165. Fitzpatrick, *supra* note 2, at 492.

166. *Id.* at 491–92.

language of the constitution. Second, its legitimacy depends upon a forced, incorrect construction of the word "vacancy."

The constitution requires the legislature to determine the manner for filling all vacancies not otherwise provided for in the constitution.[167] This includes vacancies in the appellate courts. In circumscribing the legislature's power, the constitution has placed a limitation on the period of the appointment, providing that "[n]o appointment or election to fill a vacancy shall be made for a period extending beyond the unexpired term."[168] The legislature has abided by this constitutional mandate by providing that the term of an appointed judge expires on August 31 following the next biennial election.[169] The appointed judge either must be "elected by the qualified voters of the State," at that election or cease to serve;[170] otherwise the appointment would be in violation of the constitutional limitation imposed on the period of appointment. By virtue of these provisions, no appointed judge is able to avoid an election.

In fulfilling its constitutional mandate to determine the manner for filling vacancies, the legislature, by statute, has authorized the governor to fill all appellate court vacancies.[171] The statute plainly provides that "[w]hen a vacancy occurs in the office of an appellate court . . . by death, resignation, *or otherwise*, the governor shall fill the vacancy by [appointment.]"[172] The language makes it clear that all vacancies are to be filled by gubernatorial appointment. To circumvent this plain language used in the constitution and statute, Professor Fitzpatrick relies upon a forced and incorrect definition of the term "vacancy," surmising that the "constitution uses the word 'vacancies' to refer only to interim vacancies."[173] This strained construction is directly contradicted by more than a century of Tennessee law.

While neither the constitution nor the statute defines "vacancy," the courts have applied a consistent and unambiguous definition. The term is used in its ordinary sense, not in a limited or special one: "There is no technical or peculiar meaning to the word 'vacant' when applied to office. It means

---

167. TENN. CONST. art. VII, § 4.

168. *Id.* § 5.

169. TENN. CODE ANN. § 17-4-112(b) (1994 & Supp. 2007).

170. *Id.*

171. TENN. CODE ANN. § 17-4-112 provides that

(a) When a vacancy occurs in the office of an appellate court after September 1, 1994, by death, resignation or otherwise, the governor shall fill the vacancy by appointing one (1) of the three (3) persons nominated by the judicial selection commission, or the governor may require the commission to submit one (1) other panel of three (3) nominees. . . .

(b) The term of a judge appointed under this section shall expire on August 31 after the next regular August election occurring more than thirty (30) days after the vacancy occurs.

172. *Id.* § 17-4-112(a) (emphasis added).

173. Fitzpatrick, *supra* note 2, at 475 ("[I]t appears that the constitution uses the word 'vacancies' to refer only to interim vacancies—i.e., where the judges leave in the middle of their terms—rather than to positions that are vacant simply because judges choose not to run for reelection.").

unoccupied, without an incumbent, *regardless of whether it was ever filled, or when or how it subsequently became without an incumbent.*"[174]

Professor Fitzpatrick suggests that the issue of the constitutionality of gubernatorial appointments for end-of-term vacancies remains viable because "[n]one of the courts that have considered the constitutionality of the Tennessee Plan have addressed this point."[175] This argument disregards the plain unequivocal language of the appointment statute and recent precedent. The statute requires the governor to fill all vacancies created by "death, resignation or otherwise."[176] "Otherwise" means "in another way, or in other ways."[177] Thus, the governor must fill vacancies created by death, resignation, or created in any other way.

The only legitimate judicial interpretation of the statute is that the governor fills *all* appellate court vacancies, not just vacancies occurring midterm. This was the interpretation applied to the statute by the United States District Court for the Middle District of Tennessee in *Johnson v. Bredesen*. The District Court held that the statute "makes it abundantly clear" that it applies to vacancies created by appellate judges deciding not to pursue a new eight-year term. That vacancy "is to be filled by gubernatorial appointment followed by a retention election held at the next biennial August election . . . ."[178]

b. Retention Elections are Elections

Professor Fitzpatrick next argues that a retention election is unconstitutional because it cannot be reconciled with either traditional notions of democracy nor traditional definitions of election. In reality, his argument is that retention elections fail to satisfy his own definition of "election" and his concept of democracy.

The first argument—that a retention election does not fit the definition of election—fails because it turns on the assumption that the word "elect" means a popular election between candidates. The argument runs counter to the most

---

174. Richardson v. Young, 125 S.W. 664, 683 (Tenn. 1910); *accord* Conger v. Roy, 267 S.W. 122, 125 (Tenn. 1924); Ashcroft v. Goodman, 202 S.W. 939, 940 (Tenn. 1918); State *ex rel.* Gann v. Malone, 174 S.W. 257, 259 (Tenn. 1915); State *ex rel.* Witcher v. Bilbrey, 878 S.W.2d 567, 573–74 (Tenn. Ct. App. 1994).

175. Fitzpatrick, *supra* note 2, at 492.

176. TENN. CODE ANN. §17-4-112(a) (1994 & Supp. 2007).

177. 10 OXFORD ENGLISH DICTIONARY 984 (2d ed. 1989).

178. Johnson v. Bredesen, No. 3:07-0372, 2007 U.S. Dist. LEXIS 33897, at *16 n.5 (M.D. Tenn. May 8, 2007). In this case, plaintiffs, which included the Tennessee Center for Policy Research, challenged the constitutionality of the Tennessee Plan in United States District Court on the basis that it denied voters their Fourteenth Amendment property right to vote for a judge in a contested judicial election. *Id.* at *3. Mr. Johnson's suit was consolidated with Mr. Hooker's. Among the challenges was an attack on the authority of the governor to appoint a judge for an end-of-term vacancy. *Id.* at *14–15. The has been dismissed for failure to state a claim. Johnson v. Bredesen, No. 3:07-0372, 2008 U.S. Dist. LEXIS 19738 (M.D. Tenn. Mar. 13, 2008).

basic tenets of construction. Words must be given their natural and ordinary meaning. They must be construed in a common-sense fashion so as to not create inconsistencies within a document. To construe the word "elect" to refer specifically to popular elections would lead to internal conflict within the constitution. Rather than creating conflict by construction, courts are required to "'harmonize such portions and favor the construction which will render every word operative . . . .'"[179] By construing the word "elect" broadly to mean any kind of a selection process, the Tennessee courts have honored their obligation as interpreters of the law.

As a general proposition, neither "elect" nor "election" have a unilocular meaning. The word "elect" has many definitions and dozens of applications.[180] The Oxford English Dictionary defines "elect" to mean "[t]o choose (a person) by vote for appointment to an office or position of any kind."[181] Other definitions include "to choose" and "to select." While the word undoubtedly describes a selection process, it does not demarcate, nor mandate, the details of the process. Rather, it provides flexibility and a wide range of options.

None of Tennessee's constitutions have defined the term "elect," but all of them have used the word interchangeably to refer to numerous different selection processes. These include popular elections, legislative appointments, legislative balloting, retention elections, referenda, and ratifications, to name but a few.[182] While Professor Fitzpatrick criticizes the Tennessee Supreme Court for considering these various constitutional provisions[183] this interpretive mechanism used by the court in *Dunn* and *Thompson* is the very core of statutory construction.[184] By reference to other election procedures in the constitution, the court determined that retention elections satisfy the constitutional requirement.

In all of the various election processes provided for in the constitution, the details of the process have been left to legislative design. This is consistent with the recognition that the purpose of a constitution is to provide a general framework for government. It is neither appropriate nor desirable for a

---

179. State *ex rel.* Hooker v. Thompson, No. 01S01-9605-CH-00106, 1996 WL 570090 (Tenn. Oct. 2, 1996) (quoting Shelby County v. Hale, 292 S.W.2d 745, 749 (Tenn. 1956)).

180. *See generally* Erica Klarreich, *Election Selection*, 162 SCI. NEWS 280 (2002) (comparing plurality voting with other voting procedures used internationally, based upon principles of mathematics); Pippa Norris, *Choosing Electoral Systems: Proportionate, Majoritarian, and Mixed*, 18 INT'L POL. SCI. REV. 297, 299 (1997) (discussing four major categories of election types with at least twelve subcategories).

181. 5 OXFORD ENGLISH DICTIONARY 115 (2d ed. 1989). In the seminal early work on judicial retention elections in the United States, the authors likewise refer to retention elections as elections. CARBON & BERKSON, *supra* note 122, at 3.

182. *See supra* text accompanying notes 30–32, 49–66. Similarly, as the Tennessee Supreme Court has pointed out, the word is used in multiple ways in the Tennessee statutes. State *ex rel.* Higgins v. Dunn, 496 S.W.2d 480, 489 n.1 (Tenn. 1973) (listing thirteen separate statutory uses).

183. Fitzpatrick, *supra* note 2, at 492–94.

184. *See supra* text accompanying notes 152–58.

constitution to contain exhaustive details; doing so would limit the document's vitality over time.

Consistent with the underlying purpose of a constitution, the Tennessee judicial article provides generally for an electoral process but leaves the details to statute. The constitutional requirement that "the judges of the Supreme Court shall be elected by the qualified voters of the state"[185] is satisfied by any process by which the voters have a right to choose or select. In retention elections, voters choose whether a judge remains in office. By giving voters this choice, the constitutional requirement of an election is fulfilled.

Professor Fitzpatrick expresses concern that if the term "elect" is broadly construed consistent with the *Dunn* decision "then the legislature might permit governors to win second terms in uncontested retention referenda . . . ."[186] The sincere, albeit curt, response is "Yes, and your point is . . . ?" The simple truth is that the legislature could do so. It would not be unconstitutional, as a general proposition, for a state to have a retention election for governor or for any elected office. The fact that such a process might be unwise or unpopular does not mean that it would be unconstitutional. To the extent that the constitution does not mandate a particular electoral process for an office, it allows any process that involves some selection or choice.

c. Retention Elections Satisfy Democracy

Professor Fitzpatrick reasons that because retention elections were not customary when the constitution of 1870 was passed, they could not have been contemplated nor intended under its terms.[187] But he readily concedes that constitutions are intended to provide a general outline conducive to flexible interpretation, not a comprehensive description embracing every potential issue that might arise.[188] Moreover, he admits that the 1870 constitution used the word "elect" to refer to yes/no votes.[189] His criticism of the *Dunn* court for relying on two later amendments also providing for yes/no votes is unwarranted because not only did the 1870 constitution use the word "elect" to include yes/no votes, so did the two previous constitutions.[190] The fact that those who authored the document used the word even a single time to describe yes/no

185. TENN. CONST. art. VI, § 3. The very next sentence confirms legislative involvement in the details of the election. It provides that the "[l]egislature shall have power to prescribe such rules as may be necessary to carry out the provisions of section two of this article." The specific reference to section two does not limit the general power of the legislature to provide the details of the election process, but simply reiterates that the power is to be used to assure that no more than two judges reside in any of the state's three divisions. *Id.* § 2–3.

186. Fitzpatrick, *supra* note 2, at 493.

187. *Id.* at 494.

188. *Id.*

189. *Id* at 493–94.

190. *See supra* text accompanying notes 15, 30.

voting is sufficient to establish that it was understood and contemplated at the time.

Professor Fitzpatrick next suggests that retention elections might be valid if they "serve the democratic purposes of the 1870 constitution just as well as contested elections do." In essence, he constructs his own test for determining whether retention elections are constitutional. In order to do so, he continues to presuppose a rigid and forced construction of the word "elect" which cannot be justified.[191] The test that he creates is whether retention elections "facilitat[e] democractic accountability" as well as popular elections. The suggestion is that retention elections survive constitutional scrutiny only if they equal popular elections in facilitating accountability. But both the choice of this standard—"facilitating democratic accountability"—and the definition of accountability that is implicit in the essay's discussion are the author's alone.

Retention elections may be inconsistent with *some* ideas of democracy. But just as there is no one meaning of "elect," there is no one meaning of democracy. Without a doubt, the frontier Tennesseeans believed they were creating a democracy when they adopted the early constitutions. Yet both the 1796 and 1835 constitutions provided for the appointment of judges and the governor by the legislature. And while it is true that the 1870 constitution coincided with the development of Jacksonian democracy, the framers did not provide that judges would be popularly elected. Instead, they used the same word that they used to refer to yes/no votes on referenda, ratifications, and other approval processes.[192]

If the provisions of the 1870 constitution must accomplish "democratic accountability," and if, as Professor Fitzpatrick suggests, democratic accountability may be accomplished only by popular elections or their equivalent, then dozens of provisions of the Tennessee constitution and hundreds of Tennessee statutes are invalid. Surely, for example, the legislative election of the Speakers, Treasurer, and Comptroller[193] does not "serve[] the democratic purpose[] . . . as well as contested elections";[194] neither do the legislative appointments of interim members[195] or the Secretary of State.[196] Yet the constitution specifically provides for these selection methods.[197] Similarly,

---

191. *See supra* text accompanying notes 181–86.

192. *See* TENN. CONST. art. II, § 15; *id.* art. III, § 2.

193. *Id.* art. II, § 11; *id.* art. VII, § 3; *see* David Carleton, *The Governorship*, *in* TENNESSEE GOVERNMENT AND POLITICS: DEMOCRACY IN THE VOLUNTEER STATE 41, 47 (John R. Vile & Mark Byrnes eds., 1998).

194. Fitzpatrick, *supra* note 2, at 495.

195. TENN. CONST. art. II, § 15.

196. *Id.* art. III, § 17.

197. The constitution also provides that the legislature has the power to determine the method of selection for all officers not otherwise provided for. *Id.* art. VII, § 4. As one commentator explained, "[n]ow the legislature can call for an election or otherwise specify how an officer is to be selected." LEWIS L. LASKA, THE TENNESSEE STATE CONSTITUTION: A REFERENCE GUIDE 129 (1990).

the constitution provides for the gubernatorial appointment of judges (when regular judges are disqualified)[198] and temporary constitutional officers;[199] and for the appointment of the Attorney General[200] and the Clerks of the Court[201] by the supreme court. None of these appointment processes provides for democratic accountability in the way that a popular election does, but all are nonetheless constitutional.

Similarly, numerous statutes vest the power to appoint judges, sometimes permanently and sometimes temporarily, in either the executive or legislative branch. For example, the governor is empowered to fill judicial positions created by death, resignation, and removal[202] and to appoint special judges to hear cases when sitting judges are disqualified by sickness, incompetency, or disability.[203] The chief justice of the supreme court may appoint special judges;[204] county and municipal bodies appoint county and municipal judges;[205] sitting judges may appoint substitute judges,[206] and until 1997, with consent, the parties to a civil suit could appoint their own judge.[207]

Just as retention elections may be inconsistent with some ideas of democracy, they may also be inconsistent with some ideas of judicial accountability. Without expressly saying so, Professor Fitzpatrick implies that by judicial accountability he means the ability to influence judicial decisions. In other words, he links judicial accountability with majority public approval and finds it encouraging that "judges who run in referenda . . . report . . . that

198. TENN. CONST. art. VI, § 11.

199. *Id.* art. III, § 14.

200. *Id.* art. VI, § 5.

201. *Id.* § 13.

202. TENN. CODE ANN. § 17-1-301(a) (1998). The current version of the statute refers to a vacancy which occurs as a result of "death or other disqualifying event." TENN. CODE ANN. § 17-1-301(a) (1994 & Supp. 2007).

203. TENN. CODE ANN. § 17-2-102 (1994 & Supp. 2007) (incompetency); *id.* § 17-2-104 (illness); *id.* § 17-2-105 (incompetency, sickness, or disability of intermediate appellate judges); *id.* § 17-2-107 (incompetency, sickness, or disability of general sessions judges); *id.* § 17-2-115 (giving governor the power to appoint a judge in the event of incompetency); *id.* § 17-2-116 (giving governor the power of appointment in the event that a judge is certified as ill or disabled; providing that if the judge subsequently dies or retires, the successor shall continue to serve "until such time as the successor . . . is duly elected, qualified and installed in office in the manner provided by law . . . ."). The procedure set forth in section 17-2-116 has been at issue in all cases challenging the constitutionality of the Tennessee Plan.

204. *Id.* § 17-2-109(a)(1).

205. *Id.* § 17-1-303 (county judges); *id.* § 16-18-101 (municipal judges).

206. *Id.* § 17-2-118(a) ("If, for good cause, including, but not limited to, by reason of illness, physical incapacitation, vacation or absence from the city or judicial district on a matter related to the judge's judicial office, the judge of a state or county trial court of record is unable to hold court, such judge shall appoint a substitute judge to hold court, preside and adjudicate.").

207. TENN. CODE ANN. §17-2-108 (1996) (repealed 1997).

the prospect of running in the referenda influences their decisions on the bench."[208]

But accountability to majority rule and thus susceptibility to majority influence has never been the model for the American justice system. As Justice Felix Frankfurter explained, "Courts are not representative bodies. They are not designed to be a good reflex of a democratic society. . . . Their essential quality is detachment, founded on independence."[209] It was as important in 1870 that judges remain independent from undue political influence as it was that the people elect judges. It is disingenuous to assume, as Professor Fitzpatrick does, that the constitution intended one motivation to completely displace the other.

In the Tennessee Plan, the legislature has created a judicial selection method that satisfies the desire for public accountability while shielding judges from undue political influence. It is a unique system that responds to concerns about the absence of accountability by linking retention with satisfactory judicial performance.[210] By its passage, the legislature has evidenced its desire to provide for accountability but not at the expense of excellence. Moreover, accountability under the Tennessee Plan is based on criteria that signifies good judging,[211] rather than being at best a popularity contest and at worst a high dollar partisan political race.[212]

---

208. Fitzpatrick, *supra* note 2, at 497.

209. Dennis v. United States, 341 U.S. 494, 525 (1951) (Frankfurter, J., concurring).

210. TENN. SUP. CT. R. 27 (providing for the judicial performance and evaluation program).

211. Judicial performance is evaluated based on the following criteria:

(A) Integrity. In addition to other appropriate performance measures, the committee shall consider: (1) avoidance of impropriety and appearance of impropriety; (2) freedom from personal bias; (3) ability to decide issues based on the law and the facts without regard to the identity of the parties or counsel, or the popularity of the decision and without concern for or fear of criticism; (4) impartiality of actions; and (5) compliance with the Code of Judicial Conduct contained in TENN. S. CT. R. 10.

(B) Knowledge and Understanding of the Law. In addition to other appropriate performance measures, the committee shall consider: (1) understanding of substantive, procedural, and evidentiary law; (2) attentiveness to factual and legal issues before the court; and (3) proper application of judicial precedents and other appropriate sources of authority.

(C) Ability to Communicate. In addition to other appropriate performance measures, the committee shall consider: (1) clarity of bench rulings and other oral communications; (2) quality of written opinions with specific focus on clarity and logic, and the ability to explain clearly the facts of the case and the legal precedents at issue; and (3) sensitivity to the impact of demeanor and other nonverbal communications.

(D) Preparation and Attentiveness. In addition to other appropriate performance measures, the committee shall consider: (1) judicial temperament, including courtesy to all parties and participants; and (2) willingness to permit every person legally interested in a proceeding to be heard, unless precluded by law or rules of court.

(E) Service to the Profession and the Public. In addition to other appropriate performance measures, the committee shall consider: (1) efficient administration of

Although Professor Fitzpatrick acknowledges that the judicial performance evaluation system provides a measure of accountability, he argues that it does not fulfill "democratic accountability" because judges receive favorable evaluations and have routinely been retained. In other words, democracy fails unless judges are defeated. This cynical viewpoint ignores the more likely explanation for the positive evaluations and the high rate of retention among Tennessee's appellate judges—perhaps the judges are doing a good job. Those who have experience with the Tennessee judiciary have attributed the high retention rate to the "high caliber" of Tennessee's appellate judges.[213]

If, as Professor Fitzpatrick posits, democracy fails unless judges are defeated, then Tennessee's popular election period was a complete democratic failure. During that time, most Tennessee judges were appointed, not elected, to the bench, and few were ever opposed for their seats.[214] The tradition of appointment and non-opposition was so entrenched, that by 1947, the method of choosing state appellate judges would be described as an "approval" system:

> [N]early 60 percent of the regular judges who have served on our Supreme Court during the last one hundred years have been appointed by the Governor in the first instance. . . . Judges appointed to serve out unexpired terms are generally re-elected. Even when a judge first reaches the bench through the election route, he is not as a rule selected by the electorate. He is selected by the party leaders, and the party leaders are generally lawyers who have considerable information as to their selectee's qualifications for judicial office. The election by the people is only a formal approval of such selection

---

caseload; (2) attendance at and participation in judicial and continuing legal education programs; (3) participation in organizations which are devoted to improving the administration of justice; (4) efforts to ensure that the court is serving the public and the justice system to the best of its ability and in such a manner as to instill confidence in the court system; and (5) service in leadership positions and within the organizations of the judicial branch of government.

(F) Effectiveness in Working With Other Judges and Court Personnel. In addition to other appropriate performance measures, the committee shall consider: (1) exchanging ideas and opinions with other judges during the decision-making process; (2) commenting on the work of colleagues; (3) facilitating the performance of the administrative responsibilities of other judges; and (4) working effectively with court staff.

TENN. SUP. CT. R. 27 § 3.01.

212. The recent campaign for the position of Chief Justice of the Alabama Supreme Court cost $8.2 million. JAMES SAMPLE, LAUREN JONES, & RACHEL WEISS, THE NEW POLITICS OF JUDICIAL ELECTIONS 2006, at 5 (Jesse Rutledge ed.), *available at* http://www.justiceatstake.org/files/NewPoliticsofJudicialElections2006.pdf. The total spending in the race was $13.4 million. *Id.* Alabama is not an aberration; record totals were spent in Georgia, Kentucky, Oregon, and Washington in 2006 as well. *Id.* at 15.

213. *See supra* note 142.

214. *See id.*

> by the party leaders and that approval is generally obtained in Tennessee in an election in which there is no opposition.[215]

Part II of this Essay responds to the remainder of the claims related more generally to Professor Fitzpatrick's claim that the Tennessee Plan is not fulfilling the legislature's purpose.

d. Rejection of the 1977 Constitutional Amendment Did Not Render the Tennessee Plan Unconstitutional

Professor Fitzpatrick's last point, which he characterizes as "powerful, but not conclusive,"[216] is that the voter's rejection of the 1977 amendment favors the conclusion that the Tennessee Plan is unconstitutional. This is an indefensible and overly simplistic interpretation of the failed 1977 constitutional amendment. There is no legal basis for using the public's vote to evaluate the constitutionality of a legislative enactment;[217] nor is it proper to construe the vote as enjoining future legislative reform for the courts.

Even if the law attached legal significance to a failed public initiative, which it does not, it could not do so under the complex circumstances surrounding the 1977 Limited Tennessee Constitutional Convention. From the complex and intricate history of the Convention, described earlier in this paper,[218] Professor Fitzpatrick urges one conclusion: The people rejected the judiciary amendment because they wanted an elected judiciary. Under that logic, the 1979 statute creating the Court of the Judiciary would be unconstitutional, because the voters rejected the constitutional proposal to create the Court of Discipline and Removal.[219] Similarly, the 1989 statute providing for a state-wide public defender system would be unconstitutional,[220] because the voters rejected the constitutional proposal requiring that the General Assembly provide for the "adequate defense of indigents."[221] In

---

215. Parks, *supra* note 6, at 629 (quoting WILLIAM H. WICKER, *Constitutional Revision and the Courts*, *in* PROCEEDINGS OF THE SIXTH ANNUAL SOUTHERN INSTITUTE OF LOCAL GOVERNMENT 12, 14 (Bureau of Public Administration, University of Tennessee – Knoxville 1947)).

216. Fitzpatrick, *supra* note 2, at 498.

217. Professor Fitzpatrick cites Justice Souter's dissent in *Seminole Tribe of Florida v. Florida*, 517 U.S. 44 (1996), as authority for the proposition that "it is certainly not uncommon to use [rejected constitutional amendments] to interpret the meaning of a constitution." Fitzpatrick, *supra* note 2, at 498 n.211. The case deals with the whether suits by Indian tribes against states had been authorized by Congress consistent with the Eleventh Amendment. *Seminole Tribe*, 517 U.S. at 47. It's relevance on the point for which it is cited seems totally illusory.

218. *See supra* text accompanying notes 106–26.

219. *See* TENN. CODE ANN. § 17-5-101 to -314 (1994 & Supp. 2007); *see also supra* note 118.

220. TENN. CODE ANN. § 8-14-201 to -212.

221. This was the proposal set forth in section 12 of Proposal 13. *Proclamation by the Governor*, *supra* note 114; *see also supra* note 118.

addition, statutes providing for court redistricting[222] and supreme court rulemaking[223] would also violate the constitution. Thus, Professor Fitzpatrick's claim that the public's failure to ratify the judicial article represents a public mandate against merit selection finds no support in the circumstances or in the law.

II. AN ANALYSIS OF THE TENNESSEE PLAN'S FULFILLMENT OF ITS LEGISLATIVE PURPOSE

Just as the more complete story of Tennessee history has refined the discussion of the constitutionality of the Tennessee Plan, a more balanced account of merit selection will inform the discussion of the Tennessee Plan's success in fulfilling its legislative purpose.

Merit selection of judges originated from dissatisfaction with judicial elections, both partisan and nonpartisan. Roscoe Pound summarized this dissatisfaction in a famous 1906 speech to the American Bar Association entitled *The Causes of Popular Dissatisfaction with the Administration of Justice*: "Putting courts into politics, and compelling judges to become politicians, in many jurisdictions has almost destroyed the traditional respect for the bench."[224]

In 1914, Albert M. Kales of the American Judicature Society proposed an alternative selection process in a series of writings.[225] According to Kales, judges should be selected by the entity that is "most emphatically legal, conspicuous, subject directly to the electorate, and interested in and responsible for the due administration of justice."[226] The Kales Plan called for judges to be appointed by the chief justice, who would be popularly elected. Kales also proposed that a "judicial council" be given the authority to compile an "eligible list" of attorneys from which the chief justice would appoint judges.[227]

Under the Kales Plan, the tenure of judges appointed by the chief justice would be determined by voters in periodic noncompetitive elections.[228] Kales believed that such elections "present[ed] the essential features of a recall and at the same time [were] a fair substitute for the present periodic election" in that

---

222. *See* SCHEB & RECHICHAR, *supra* note 110, at 61–67; *see also supra* note 118.

223. TENN. CODE ANN. §§ 16-3-401 to -408.

224. Roscoe Pound, *The Causes of Popular Dissatisfaction with the Administration of Justice*, 46 J. AM. JUDICATURE SOC'Y 54, 66 (1962).

225. *See* ALBERT M. KALES, UNPOPULAR GOVERNMENT IN THE UNITED STATES (1914); *First Draft of an Act to Establish a Model Court for a Metropolitan District*, 4 AM. JUDICATURE SOC'Y BULL. (1914); *First Draft of a State-Wide Judicature Act*, 7 AM. JUDICATURE SOC'Y BULL. (1914).

226. *See First Draft of an Act to Establish a Model Court for a Metropolitan District*, *supra* note 225, at 36.

227. *See* KALES, *supra* note 225, at 250.

228. *See First Draft of an Act to Establish a Model Court for a Metropolitan District*, *supra* note 225, at 149–53.

they allowed the electorate to "retire unfit men" but relieved voters of the "largely impossible" task of choosing which lawyers should serve as judges.[229]

Founder of the American Judicature Society Herbert Harley offered a modified version of the Kales Plan in 1928, in which the governor would appoint judges from a list of names compiled through a bar plebiscite.[230] Participation of laypersons in the judicial nominating process was first suggested in 1931 by the Grand Jury Association in New York.[231]

In 1937, the American Bar Association adopted a resolution that combined the elements proposed by Kales and Harley, recommending the "filling of vacancies by appointment by the executive or other elective official or officials, but from a list named by another agency, composed in part of high judicial officers and in part of other citizens, selected for the purpose, who hold no other public office."[232] The American Bar Association resolution called for reappointment or retention elections after an initial term of office and periodically thereafter.[233]

Versions of this nominative-appointive-elective plan were considered in several states during the 1930s, but it was Missouri that first established what it termed the "Nonpartisan Court Plan" in 1940.[234] During the 1960s and 1970s, twenty-three jurisdictions adopted what had become known as the "Missouri Plan" or "merit selection."[235] Today, thirty-three states and the District of Columbia use merit selection to choose at least some of their judges.[236]

---

229. *See First Draft of a State-Wide Judicature Act*, *supra* note 225, at 164. While Professor Fitzpatrick maintains that the "architects of merit selection" proposed retention elections to provide "life tenure but without the appearance of life tenure," historians report that retention elections had two principal purposes: "to ensure that judges would be retained for lengthy terms of tenure once they had been chosen on the basis of professional merit," and "to accommodate the populists who insisted on a mechanism to hold judges publicly accountable." *See* CARBON & BERKSON, *supra* note 122, at 6.

230. Editorial, *The Eligible List of Judicial Candidates*, 11 J. AM. JUDICATURE SOC'Y 131 (1928).

231. *See* Glenn Winters, *The Merit Plan for Judicial Selection and Tenure—Its Historical Development*, *in* SELECTED READINGS: JUDICIAL SELECTION AND TENURE 36 (Glenn Winters ed., 1973).

232. John Perry Wood, *Basic Propositions Relating to Judicial Selection—Failure of Direct Primary—Appointment Through Dual Agency—Judge to "Run on Record"*, 23 A.B.A. J. 104–05 (1937).

233. *See id.*

234. Winters, *supra* note 231, at 36.

235. *See* AMERICAN JUDICATURE SOCIETY, JUDICIAL MERIT SELECTION: CURRENT STATUS (2008), *available at* http://www.judicialselection.us/uploads/documents/Judicial_Merit_Charts_0FC20225EC6C2.pdf (hereinafter CURRENT STATUS).

236. *See id.* Twenty-four states and the District of Columbia use merit selection to make initial appointments to some or all of their courts; nine states use merit selection to fill midterm vacancies only. *Id.* Eight states and the District of Columbia require legislative confirmation of gubernatorial appointments, and five states and the District of Columbia substitute a reappointment process for retention elections. *Id.*

When the Tennessee legislature created the Tennessee Plan in 1971 it announced four goals: selecting the best qualified judges, bringing more racial and gender diversity to the bench, insulating judges from political pressure and influence, and enhancing the prestige of and public respect for the courts.[237] In the sections that follow, this Essay examines the extent to which merit selection generally, and the Tennessee Plan specifically, accomplishes these objectives.

A. Selecting Highly Qualified Judges

Scholars have used a variety of approaches to address the question of whether merit selection systems produce better judges than do other selection methods, with mixed results.[238] Some studies have compared the educational backgrounds and professional experience of judges selected by appointment and election. The most comprehensive analysis of this kind reported that merit-selected and popularly-elected state high court judges did not differ significantly in the extent of their legal or judicial experience, but merit-selected judges were more likely than popularly-elected judges to have attended prestigious law schools.[239]

Other research has compared judges' performance once they attain their seats. A recent study examined the work product of state high court judges and concluded that, while elected judges were more productive than merit-selected judges, appointed judges' opinions were of higher quality.[240] Some analyses have assessed judicial performance through ratings or rankings by attorneys. Results of a survey of corporate attorneys indicated that three of the five states whose courts ranked highest on judges' competence were states in which

---

237. *See* TENN. CODE ANN. §§ 17-4-101, -102 (1994 & Supp. 2007).

238. Some studies comparing appointed and elected judges utilize inaccurate data for some judges, as they classify judges according to their formal selection method rather than the method through which they actually attained their seats. According to Holmes and Emrey, 52% of judges serving on high courts in elective states from 1964 to 2004 were initially appointed. Lisa M. Holmes and Jolly A. Emrey, *Court Diversification: Staffing the State Courts of Last Resort through Interim Appointments*, 27 JUST. SYS. J. 1, 1 (2006). Data available on the American Judicature Society's Judicial Selection in the States website indicates that 35% of judges currently serving on high courts in states with contestable elections were initially appointed to their seats. American Judicature Society, Methods of Judicial Selection, http://www.judicialselection.us/judicial_selection/methods/justices_of_the_supreme_court.cfm?state (last visited May 28, 2008).

239. *See* Henry R. Glick and Craig F. Emmert, *Selection Systems and Judicial Characteristics: The Recruitment of State Supreme Court Judges*, 70 JUDICATURE 228, 231–33 (1987).

240. *See* Stephen J. Choi, G. Mitu Gulati, & Eric A. Posner, *Professionals or Politicians: The Uncertain Empirical Case for an Elected Rather Than Appointed Judiciary* 1 (Univ. of Chi. Sch. of Law, John M. Olin Law & Econ. Working Paper No. 357, 2007), *available at* http://ssrn.com/abstract=1008989 (last visited May 20, 2008). Productivity was measured by the number of opinions judges wrote; opinion quality was measured by the number of citations to opinions by judges in other states. *Id.* at 2.

judges are appointed, while four of the five lowest ranking states on this criterion were elective states.[241] These findings are consistent with an early study of the effects of the Missouri Plan. While substantial proportions of both elected and merit-selected judges ranked in the highest quartile, fewer merit-selected than elected judges were ranked in the lowest quartile, suggesting that a merit plan "tend[ed] to eliminate the selection of very poor judges . . . ."[242]

A third approach to assessing whether appointive systems select "better" judges than elective systems is to compare the number of disciplinary incidents in which appointed and elected judges have been involved. Studies of this kind have uniformly found that elected judges were disciplined and removed from office with greater frequency than were appointed judges.[243]

It is not surprising that studies have found meaningful differences between judges chosen in appointive and elective systems. In a merit selection system, the emphasis is on qualifications and experience at the outset, and only the best qualified applicants are eligible for appointment. The Tennessee Plan is an example of how this process works in practice. Judicial vacancies are publicized when they occur, and applications are solicited from candidates who meet the constitutional and statutory requirements. Applicants are required to provide information about their professional background, judicial and administrative experience, education, and achievements. The judicial selection commission convenes a public meeting to receive comments on potential

---

241. *See* U.S. CHAMBER INSTITUTE FOR LEGAL REFORM, LAWSUIT CLIMATE 2008: RANKING THE STATES 14 (2008), *available at* http://www.instituteforlegalreform.com/states/lawsuitclimate2008/pdf/LawsuitClimateReport.pdf. The top ranking states were Delaware, Minnesota, Virginia, Nebraska, and Indiana; the lowest ranking states were Louisiana, Mississippi, West Virginia, Alabama, and Hawaii. *Id.* A follow-up analysis indicated that the average ranking of states with merit selection of judges (i.e., gubernatorial appointment from a nominating commission) was higher than states with any other selection method, while states with partisan judicial elections had the lowest average ranking. *See* Joshua C. Hall & Russell S. Sobel, *Is the "Missouri Plan" Good for Missouri? The Economics of Judicial Selection*, POLICY STUDY (Show-Me Institute, St. Louis, Mo.), May 21, 2008, *available at* http://showmeinstitute.org/docLib/20080515_smi_study_15.pdf.

242. RICHARD A. WATSON & RONDAL G. DOWNING, THE POLITICS OF THE BENCH AND THE BAR 283 (1969).

243. *See, e.g.*, Steven Zeidman, *To Elect or Not to Elect: A Case Study of Judicial Selection in New York City 1977–2002*, 37 U. MICH. J.L. REFORM 791, 808–10 (2004) (from 1977 to 2002, judges of New York City's Civil Court, who are elected, were substantially more likely to be disciplined than judges of the Criminal and Family Courts, who are appointed); CALIFORNIA COMMISSION ON JUDICIAL PERFORMANCE, SUMMARY OF DISCIPLINE STATISTICS 1990–1999, *available at* http://cjp.ca.gov/publicat.htm (disciplinary rates for elected judges from 1990 to 1999 were higher than those for judges who were initially appointed); The Florida Bar, Merit Selection and Retention, http://www.floridabar.org/DIVCOM/PI/BIPS2001.nsf/BIP+List?OpenForm (last visited May 28, 2008) (follow "Merit Selection and Retention" hyperlink) (since 1970, ten of the thirteen judges removed from the bench were elected rather than merit-selected, and 73% of the judges disciplined since 1998 initially reached the bench via election).

candidates, investigates and interviews applicants, and forwards the names of the three best qualified individuals to the governor.[244] There is no similar screening process for potential candidates in states with contestable elections, and political connections can take precedence over professional credentials.[245]

As has already been discussed, Tennessee has supplemented its selection and retention processes with a performance evaluation program designed both to promote judicial self-improvement and to enable voters to make more informed decisions in retention elections.[246] Under the Tennessee Plan, attorneys, other judges, and court personnel are asked to evaluate judges on several criteria, including integrity, knowledge and understanding of the law, ability to communicate, preparation and attentiveness, service to the profession, and effectiveness in working with other judges and court personnel.[247] The results of the evaluations of appellate judges are made public along with a recommendation for or against retention.[248] No similar, official performance evaluation programs exist in elective states.

The Tennessee Plan, both in theory and in practice, selects and retains highly qualified judges.

B. Bringing More Diversity to the Bench

Numerous studies have addressed whether particular selection methods are more likely to place diverse candidates on the bench, but the findings have been inconsistent.[249] While most of these studies consider only a state's formal selection method rather than how a judge actually reached the bench,[250] a recent analysis of state high courts over a forty-year period took into account the frequency of interim appointments in elective states and reported that gender

---

244. *See* TENN. CODE ANN. §§ 17-4-109 (1994 & Supp. 2007). When filling appellate vacancies, the governor may request a supplemental list of three names but is required to appoint a judge from the second list. *Id.* § 17-4-112(a).

245. In 2007, New York State created Independent Judicial Election Qualification Committees, a statewide network of screening panels to review the qualifications of trial court candidates. *See* New York State Unified Court System, Rules of the Chief Administrative Judge, http://www.nycourts.gov/rules/chiefadmin/150.shtml (last visited May 30, 2008). These committees are not comparable to nominating commissions, however, in that they simply rate candidates as qualified or not qualified, and candidates are not required to submit to screening in order to run for office.

246. *See supra* text accompanying notes 134–42.

247. *See* TENN. CODE ANN. §17-4-201; TENN. SUP. CT. R. 27.

248. *See* TENN. CODE ANN. § 17-4-201(a)(1), (c).

249. *Compare, e.g.*, Mark S. Hurwitz & Drew Noble Lanier, *Women and Minorities on State and Federal Appellate Benches, 1985 to 1999*, 85 JUDICATURE 84, 88–91 (2001) (women and minorities were no more likely to become state appellate judges under merit systems than non-merit systems), *with* M.L. HENRY, THE SUCCESS OF WOMEN AND MINORITIES IN ACHIEVING JUDICIAL OFFICE (1985) (women and minorities were more likely to attain judgeships through appointive systems than elective systems).

250. *See supra* note 238.

and racial diversification is more likely to occur through interim appointments than elections.[251] The demographics of state appellate courts in 2008 confirm these findings, with 65% of women judges and 76% of minority judges having been appointed rather than elected to their positions.[252]

A chief advantage of a merit selection system is that it is possible to structure the process so that opportunities for selecting a more diverse group of judges are enhanced.[253] The Tennessee Plan calls for consideration of the racial and gender population of the state in the appointment of members of the judicial selection commission,[254] and research has demonstrated that demographically diverse nominating commissions attract more diverse applicants and select more diverse nominees.[255]

According to data provided by the Administrative Office of the Courts, Tennessee's judicial selection commission has screened candidates for eighty-seven vacancies since 1994.[256] The commission has recommended 245 applicants to the governor to fill these vacancies, including sixty-three women and twenty-eight minorities. Of the governor's eighty appointees, twenty-two have been women and seven have been minorities. This contrasts markedly with the composition of Tennessee's benches before merit selection.[257] In the last decade alone, the number of women serving as appellate judges has tripled and the number of minorities serving on appellate benches has doubled.[258]

In Tennessee and nationwide, appointive systems have provided more diversity on appellate courts than have elective systems.

---

251. *See* Holmes & Emrey, *supra* note 238, at 7. A similar study found that women are significantly more likely to be selected to state high courts when initially appointed. *See* Kathleen A. Bratton & Rorie L. Spill, *Existing Diversity and Judicial Selection: The Role of the Appointment Method in Establishing Gender Diversity in State Supreme Courts*, 83 SOC. SCI. Q. 504, 504 (2002).

252. These figures include judges chosen through merit selection, gubernatorial appointment, or judicial appointment. Data on file with authors.

253. For a discussion of measures that may be used to promote diversity among nominating commission members and judicial appointees, see Leo M. Romero, *Enhancing Diversity in an Appointive System of Selecting Judges*, 34 FORDHAM URB. L.J. 485 (2007).

254. *See* TENN. CODE ANN. §17-4-102(b)(3), (d) (1994 & Supp. 2007).

255. *See* Kevin M. Esterling & Seth S. Andersen, *Diversity and the Judicial Merit Selection Process: A Statistical Report*, *in* RESEARCH ON JUDICIAL SELECTION 1999 (American Judicature Society ed., 2000).

256. Data provided by the Administrative Office of the Courts is on file with the authors.

257. *See* GENTRY CROWELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1988–1989, at 222–31 (1989); GENTRY CROWELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1989–1990, at 226–35 (1990); RILEY C. DARNELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1991–1994, at 248–58 (1994); RILEY C. DARNELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1995–1996, at 254–63 (1996); RILEY C. DARNELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1997–1998, at 250–60 (1998); RILEY C. DARNELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 1999–2000, at 264–73 (2000); RILEY C. DARNELL, TENN. SEC'Y OF STATE, TENNESSEE BLUE BOOK 2001–2004, at 288–97 (2004).

258. *See* sources cited *supra* note 257.

C. *Limiting Politics in Judicial Selection*

Regardless of which judicial selection method is used, it is impossible to entirely eliminate politics from the selection process.[259] In fact, in some appointive states, partisan politics is an explicit part of the process, with partisan balance required on judicial nominating commissions[260] or on the courts themselves.[261] But merit selection systems minimize the role of politics in judicial selection. Judicial aspirants in merit plan states are not required to raise money, seek party support, or campaign for office as are judicial candidates in elective states; and judicial campaigns in recent years have come to closely resemble campaigns for legislative and executive positions.

Judicial elections for the past decade have been characterized by unprecedented campaign fundraising and spending, increased special interest group involvement, and relaxed ethical standards for candidate speech. In the last four election cycles, candidates for state high courts have raised more than double the amount raised in the 1990s.[262] In a 2004 Illinois contest, candidates for a single district-based seat on the supreme court raised nearly $10 million, exceeding fundraising in eighteen of the thirty-four U.S. Senate races that year.[263] In 2006, candidates for the Alabama Supreme Court shattered previous records for judicial elections, raising a total of $13.4 million.[264]

At the same time, special interest groups have ramped up their efforts to influence the composition of state courts—making contributions to candidates, funding television advertising through independent expenditures, and pressuring candidates to discuss their political views. In the 2005–2006 election cycle, 44% of the contributions to state high court candidates came from business groups, and 21% came from trial attorneys.[265] These special interest groups also spent a total of more than $5 million on television ads in ten states with high court races in 2005–2006,[266] and in an April 2008 Wisconsin race, special interest groups spent approximately $4 million on a single supreme court race.[267]

---

259. *See, e.g.*, Melinda Gann Hall, *State Supreme Courts in American Democracy: Probing the Myths of Judicial Reform*, 95 AM. POL. SCI. REV. 315 (2001) (voter reactions to controversial policy issues and the extent of partisan composition in the state affected outcomes in all types of judicial elections—partisan, nonpartisan, and retention).

260. These states include Arizona, Connecticut, Delaware, Indiana, Nebraska, New York, South Dakota, Utah, Vermont. *See* CURRENT STATUS, *supra* note 235.

261. These states include Delaware and New Jersey. *See* American Judicature Society, Judicial Selection in the States, http://www.judicialselection.us (last visited May 20, 2008) (click on individual states shown on interactive map).

262. *See* SAMPLE ET AL., *supra* note 212, at 15.

263. *See* Robert Barnes, *Judicial Races Now Rife with Politics*, WASHINGTON POST, Oct. 28, 2007.

264. *See* SAMPLE ET AL., *supra* note 212, at 15.

265. *See id.* at 18.

266. *See id.* at 3.

267. *See* Emma Schwartz, *Elections for Judges are Getting Nastier*, U.S. NEWS & WORLD

Outside groups have also expanded their efforts to ascertain judicial candidates' views on controversial issues, distributing questionnaires regarding their positions on such subjects as abortion, the death penalty, and same-sex marriage, and publicizing their responses and failures to respond.[268] And in recent elections, candidates have been less constrained than in the past in responding to such questionnaires. According to a 2002 U.S. Supreme Court decision, candidates for state court seats are free to announce their views on legal and political issues—issues that may later come before them as judges.[269]

While no system of selecting judges can be completely insulated from politics, merit selection systems negate the importance of electoral campaigning, interest group activity, and candidate fundraising in the selection process.

D. Enhancing Public Confidence in the Courts

The increased politicization of judicial elections has not gone unnoticed by voters, and it seems to have taken a toll on the public's confidence in its courts. According to recent national surveys, between two-thirds and three-fourths of Americans believe that the need to raise money to conduct their campaigns influences judges' decisions.[270] More than four in five Americans are concerned that the U.S. Supreme Court's decision in *Republican Party of Minnesota v. White* will lead to special interest groups pressuring candidates to take positions on controversial issues,[271] and nine in ten fear that special interests are trying to use the courts to shape economic and social policy.[272] These concerns are reinforced by research that identifies correlations between campaign contributions and judicial decisions.[273]

REPORT, Apr. 4, 2008.

268. *See* Marcia Coyle, *Judicial Surveys Vex the Bench*, THE NATIONAL LAW JOURNAL, Sept. 8, 2006.

269. Republican Party of Minn. v. White, 536 U.S. 765 (2002).

270. *See* ANNENBERG PUBLIC POLICY CENTER, PUBLIC UNDERSTANDING OF AND SUPPORT FOR THE COURTS 3 (2007), *available at* http://www.annenbergpublicpolicycenter.org/Downloads/20071017_JudicialSurvey/Judicial_Findings_10-17-2007.pdf (69% of respondents believed that the need to raise money for elections affects judges' rulings to a moderate or great extent); JUSTICE AT STAKE CAMPAIGN, AMERICANS SPEAK OUT ON JUDICIAL ELECTIONS 1 (2004), *available at* http://www.justiceatstake.org/files/ZogbyPollFactSheet.pdf (71% of respondents believed that campaign contributions from interest groups have at least some influence on judges' decisions); JUSTICE AT STAKE CAMPAIGN, NATIONAL SURVEY OF AMERICAN VOTERS 7 (2001), *available at* http://www.justiceatstake.org/files/JASNationalSurvey Results.pdf (67% of respondents believed that individuals or groups who give money to judicial campaigns often receive favorable treatment).

271. *See* AMERICANS SPEAK OUT ON JUDICIAL ELECTIONS, *supra* note 270, at 1.

272. *See* NATIONAL SURVEY OF AMERICAN VOTERS, *supra* note 270, at 9.

273. *See, e.g.*, TEXANS FOR PUBLIC JUSTICE, PAY TO PLAY: HOW BIG MONEY BUYS ACCESS TO THE TEXAS SUPREME COURT (2001), *available at* http://www.tpj.org/docs/2001/04/reports/paytoplay/paytoplay.pdf (the Texas Supreme Court was four times more likely to accept a case

On the other hand, substantial majorities of voters nationwide and in individual states support merit selection and retention systems.[274] These systems significantly limit the involvement of parties, special interests, and money in the selection of judges, and in so doing, they preserve the public's confidence in its courts.

CONCLUSION

A few weeks before these Essays, wrestling with the constitutionality of the Tennessee Plan, were published, Tennessee's unique system for electing appellate court judges with its mutual accommodation of judicial independence and public accountability was dealt a likely fatal blow by the Tennessee General Assembly. Set to sunset in 2008, the Plan needed legislation to keep it alive. Because the legislation did not pass, the Plan is set to wind down completely in 2009, unless new legislation is passed. If the Tennessee legislature fails to revive the Tennessee Plan during the next calendar year, the Plan's demise will not be attributable to either author's rhetoric or logic, nor will it signify a considered rejection of merit selection. Rather, as has been true from the beginning, Tennessee's merit selection system will be yet another bargaining chip gambled away at the tables of the Tennessee General Assembly.[275]

---

for review if the petitioner had contributed to a justice's campaign); Madhavi M. McCall & Michael A. McCall, *Campaign Contributions, Judicial Decisions, and the Texas Supreme Court: Assessing the Appearance of Impropriety*, 90 JUDICATURE 214 (2007) (the likelihood of a justice voting in a party's favor was significantly higher if the party contributed to the justice's campaign); Vernon Valentine Palmer & John Levendis, *The Louisiana Supreme Court in Question: An Empirical Study of the Effect of Campaign Money on the Judicial Function*, 82 TUL. L. REV. 1291 (2008) (in nearly half of the cases heard by the court over a fourteen-year period, a litigant or attorney had contributed to at least one justice's campaign, and on average, justices voted in favor of contributors 65% of the time); Adam Liptak & Janet Roberts, *Campaign Cash Mirrors a High Court's Rulings*, NEW YORK TIMES, Oct. 1, 2006 (over a twelve-year period, justices of the Ohio Supreme Court routinely participated in cases involving campaign contributors and, on average, voted in favor of contributors 70% of the time).

274. *See* NATIONAL SURVEY OF AMERICAN VOTERS, *supra* note 270, at 12 (71% of voters nationwide supported a general merit selection and retention proposal); Memorandum from Patrick Lanne, Public Opinion Strategies, to Interested Parties (Dec. 11, 2007), *available at* http://www.justiceatstake.org/files/MissouriMemoAndOverallResults.pdf (71% of Missourians supported the state's current system of judicial merit selection and retention); Justice at Stake Campaign, Minnesota Statewide Survey January 2008, http://www.justiceatstake.org/files/MinnesotaJusticeatStakesurvey.pdf (last visited May 28, 2008) (74% of Minnesotans supported merit selection of judges with retention elections and performance evaluation).

275. [EDITOR'S NOTE: Professor Fitzpatrick has written a reply to this Essay. It is posted at http://papers.ssrn.com/abstract=1152413.]

ECONOMIC WARLORDS: HOW DE FACTO FEDERALISM INHIBITS CHINA'S COMPLIANCE WITH INTERNATIONAL TRADE LAW AND JEOPARDIZES GLOBAL ENVIRONMENTAL INITIATIVES

GREGORY H. FULLER

INTRODUCTION

In July 2007, Zheng Xiaoyu, former head of the Chinese State Food and Drug Administration (SFDA), was executed "for accepting bribes in exchange for approving substandard medicines" now linked to at least ten deaths.[1] In a gesture both symbolic and barbaric, the Chinese government attempted to assuage foreign fears of inadequate oversight while sending its citizens an undeniable message that corruption among state officials would no longer be tolerated.[2] The recent deluge of defective Chinese products has cast a long shadow across a country often dubbed the "World's Factory."[3] Yet, the relentless growth of the world's fourth—soon to be third—largest economy[4] is unlikely to subside anytime soon, and it is becoming increasingly clear that slowing the pace of China's economic juggernaut would be an enormous

1. Ariana Eunjung Cha, *China Executes Former Head of Food, Drug Safety*, WASH. POST, July 11, 2007, at D1.

2. Cao Wenzhuang, former director of the drug registration department at the SFDA, was also sentenced to death for accepting bribes from drug manufacturers. Orville Schell, Editorial, *China's Long Industrial Nightmare*, TAIPEI TIMES, July 13, 2007, at 8, *available at* http://www.taipeitimes.com/News/editorials/archives/2007/07/13/2003369376. Orville Schell, who directs the Asia Society's Center on U.S.-China Relations, remarked that "[b]oth verdicts were doubtless calculated to, as a famous Chinese proverb puts it, 'kill some chickens in order to scare the monkeys.'" *Id.*

3. *See, e.g.*, IRA KALISH, DELOITTE RESEARCH, THE WORLD'S FACTORY: CHINA ENTERS THE 21ST CENTURY 1 (2003), *available at* http://www.deloitte.com/dtt/cda/doc/content/DTT_DR_China21Century.pdf.

4. "The National Bureau of Statistics raised its estimate of China's 2006 growth rate from 10.7 percent to 11.1 percent. It nudged up its estimate of total output by 146.4 billion yuan ($18.8 billion) to 21.1 trillion yuan ($2.705 trillion)." Associated Press, *China to Be Third Largest Economy*, CHINA DAILY, July 11, 2007, http://www.chinadaily.com.cn/china/2007-07/11/content_5433153.htm. At least one report predicts that "Chinese statistics . . . are likely to show that the country is on track to leapfrog Germany as the third-biggest national economy this year, sooner than expected—yet another sign of just how quickly the global economic balance of power is shifting." Marcus Walker & Andrew Batson, *China's GDP Poised to Top Germany's as Power Shift Speeds Up*, WALL ST. J., July 16, 2007, at A2.

challenge.[5] From Beijing to Guangzhou, the opportunities afforded by Deng Xiaoping's revolutionary reforms continue to change the face of modern China. However, the enduring problems of governing the planet's oldest civilization[6] have proven somewhat impervious to centralized planning. As China scrambles to prepare for the 2008 Olympics, these enforcement obstacles are becoming ever more apparent.

Throughout Chinese history, local warlords engaged in epic battles to expand territory in efforts to increase their own political power and prestige.[7] Dynasties rose and fell on the whims of rebels who questioned the central government's authority and sought to establish their own independent kingdoms.[8] This Comment proposes that recent economic reforms and the rigorous competition of a global market have revived these ancient Chinese proclivities, transforming local government administrators into "economic warlords" who once again ignore the edicts of China's central government in furtherance of their own glory.

The execution of Zheng Xiaoyu exemplifies the extreme measures to which the central government is willing to resort in order to ensure adherence to domestic regulations and international agreements. The decentralization of economic decision-making advanced by Deng Xiaoping's reorganization of China's governmental structure has increased prosperity across China; however, it has also made regulating the world's most populous country more difficult than ever. Over time, these reforms have allowed a political and economic system of de facto federalism to evolve, resulting in the central government's inability to effectively enforce the legal agreements it reaches with foreign countries.[9] Until these structural enforcement problems are properly addressed, the widespread piracy, hazardous manufacturing techniques, and rampant pollution that characterize much of China's economic activity will continue, despite efforts of national lawmakers.

---

5. According to a report in July 2007, China's top economic advisor warned that "the economy is still in danger of overheating" due to unabated "investment[] in energy-intensive and polluting industries." *Top Planner: Overheating Trend Still Not Contained,* CHINA ECON. REV., July 6, 2007, http://www.chinaeconomicreview.com/dailybriefing/2007_07_06/Top_planner:_Overheating_trend_still_not_contained.html. The report described concerns of Ma Kai, chairman of the National Development and Reform Commission, that "industrial production, which grew 18.1% year-on-year in May, was still too high" and that the current fiscal system will make containing such trends difficult. *Id.*

6. JASPER BECKER, THE CHINESE 1 (Oxford Univ. Press 2002).

7. JOHN BRYAN STARR, UNDERSTANDING CHINA: A GUIDE TO CHINA'S ECONOMY, HISTORY, AND POLITICAL STRUCTURE 41 (rev. and updated ed. 2001).

8. *Id.* An examination of the ebb and flow of dynasties and emperors reveals that, in 500 of the 2000 years comprising Chinese history, local warlords vied for the chance to unite China. *Id.* As recently as the 1950s, the countryside was engulfed in full-fledged civil war. *Id.*

9. *See generally* ZHENG YONGNIAN, DE FACTO FEDERALISM IN CHINA: REFORMS AND DYNAMICS OF CENTRAL-LOCAL RELATIONS (2007) (analyzing long-term stability in China despite extensive decentralization).

The interplay between the legal, economic, and cultural attributes of the Chinese government has led to a political climate in which local regulation of intellectual property rights, trade activities, and environmental reforms has a truly global impact. Part I of this Comment outlines the cultural and economic underpinnings that have caused the dispersal of China's political power to various levels of provincial and local governments. Part II examines the international legal arrangements to which China has become a party from an administrative perspective, focusing on intellectual property law as an example. Part III describes how the diverse interests of China's large population, localism, and de facto federalism have made complying with these international accords especially arduous. Finally, Part IV considers the unmitigated environmental degradation fostered by lax enforcement in the context of global environmental initiatives.

I. DE FACTO CHINESE FEDERALISM

Much of China's difficulty in enforcing domestic and international laws stems from the diverse and fragmented nature of its society. From a purely functional standpoint, China's highly-centralized political structure ranks among the most successful and enduring civilizations in all of history.[10] Such a totalitarian system may in fact be the only effective way to govern a country of over 1.3 billion citizens, comprised of fifty-eight ethnicities and many cultural subgroups.[11] Indeed, even the concept of a unified Chinese language is limited to the written word; at least seven major "mutually unintelligible dialects" of Chinese exist.[12] Accordingly, the vast majority of Chinese citizens find it nearly impossible to converse with natives of large metropolitan areas, such as Shanghai or Guangzhou.[13] This linguistic phenomenon, coupled with the distinctive cuisines and persistent stereotypes associated with each region, has created strong local loyalties among the Chinese.[14] The prevalence of local

10. BECKER, *supra* note 6, at 1. For over 2,000 years, the raw power that the Chinese government exerts over its subjects has enabled the Middle Kingdom to survive. *Id.* Although this tight-fisted authoritarian system has exacted a tremendous cost in terms of human rights, the ten major Chinese dynasties and the contemporary communist regime have succeeded in controlling more territory and people than any other government in recorded history. *Id.*

11. *Id.*; STARR, *supra* note 7, at 9.

12. STARR, *supra* note 7, at 33.

13. This often leads fledgling students of the Chinese language to ask, "'Which [Chinese] language do you speak—Cantonese, Mandarin?'" ROSS TERRILL, THE NEW CHINESE EMPIRE 51 (2003) (alteration in original). In 1955, the central government instituted a Mandarin-only policy designed to promote national unity. Bret Wallach, Historical Developments, A Companion to UNDERSTANDING THE CULTURAL LANDSCAPE 3, http://ags.ou.edu/~bwallach/documents/Part2.pdf (last visited Apr. 22, 2008). In fact, the government takes this policy so seriously that billboards in Shanghai encourage locals to "speak Mandarin . . . be a modern person." *Id.*

14. STARR, *supra* note 7, at 32. "[N]ortherners in China are often thought of as reserved, formal, and aloof; southerners, by contrast, are seen as more outgoing, volatile, and

identities has caused preeminent Chinese scholar Zheng Yongnian to observe that "the concept of one China seems like a myth that papers over economic, political, and identity disparities, and tensions between coastal and inland China."[15]

In order to administer these diverse geographical areas, the central government has long relied on its provinces to govern local populations.[16] The modern People's Republic of China is composed of twenty-two provinces and "four 'directly administered cities,'—Beijing, Tianjin, Chongqing, and Shanghai."[17] The provincial governments are charged with managing more citizens than any comparable provincial government in the world; the largest provinces of Shandong, Guangdong, Sichuan, and Henan are so populous that they "can be easily ranked among the top ten countries in the world."[18] Within the provinces, territory is divided into over two thousand counties and twenty-six thousand townships managed by local governments.[19] It is this third tier of government that actually carries out central policies; consequently, the central government faces the considerable challenge of managing such a large number of administrations.

Traditionally, the central government has employed three avenues in managing the central-local relationship: bargaining, reciprocity, and coercion.[20] Professor Zheng Yongnian believes that these methods constitute informal institutions that have created a decentralized "*de facto* federal" system of norms and rules in the central-local relationships over the past three decades.[21] Within

---

spontaneous. . . . Residents of Guangzhou are often characterized as having a highly developed commercial sense; natives of the south-central provinces of Sichuan and Hunan, perhaps because of their peppery cuisine, are called hot tempered and impetuous." *Id.* at 33–34.

15. ZHENG YONGNIAN, *supra* note 9, at 26. Zheng notes Edward Friedman's conclusion that China has been divided by the modernization process: "[T]he North can be uniformly regarded as Leninist and be subject to procrastination, while the South is dynamic and modern." *Id.*

16. *See* STARR, *supra* note 7, at 34, 45–46.

17. *Id.*

18. ZHENG YONGNIAN, *supra* note 9, at xi. In 2005, Guangdong reported a population of over 110 million. Liang Qiwen, *Province Faces Population Pressure*, CHINA DAILY, Feb. 16, 2005, at 2, *available at* http://www.chinadaily.com.cn/english/doc/2005-02/16/content_416655.htm. According to the 2000 census, Shandong boasted 90.79 million citizens; Henan, 92.56 million; and Sichuan, 42.88 million. U.N. ECON. AND SOC. COMM'N FOR ASIA, STATUS OF POPULATION AND FAMILY PLANNING PROGRAMME IN CHINA BY PROVINCE, http://www.unescap.org/esid/psis/population/database/chinadata/intro.htm (last visited Mar. 2, 2008).

19. STARR, supra note 7, at 34. These counties are further subdivided into townships and towns. *Id.* "Each county has, on average, some two dozen townships . . . and twenty thousand town governments. There are just under 800,000 village governments, the lowest level of rural administration." *Id.*

20. ZHENG YONGNIAN, *supra* note 9, at 53–54.

21. *See id.* at 53. In the United States, federalism is seen as a means of protecting individual liberties and rights against the heavy hand of centralized government. In China, on

this system, bargaining is a bilateral process in which the central and provincial governments resolve conflicts by using their resources "to maximize their respective interests."[22] Further, reciprocity is the process in which the central government and provinces "achieve voluntary cooperation . . . through self-adjustment and deliberation."[23] This method is based upon obligation, "with each side behaving in a mutually acceptable way or with each side's behavior justifiable to the other side."[24] If these customs fail to produce the desired result, the central government may resort to coercion, [25] as typified by the execution of former SFDA director Zheng Xiaoyu.

In the past, in order to prevent local leaders from amassing a powerbase from which they might challenge the central government's authority, Chinese leaders subscribed to the "law of avoidance," which precluded administrators from serving in their home provinces.[26] The modern incarnation of this policy, which observers call the *nomenklatura* system, remains the primary means of managing provincial leadership.[27] The nomenklatura system comprises "'lists of leading positions, over which party units exercise the power to make appointments and dismissals.'"[28] By tightly monitoring the activities of provincial leaders, the central government was able to coerce intermediary provincial governments into implementing policies favored by Beijing.[29]

---

the other hand, unalienable rights of life, liberty, and the pursuit of happiness have traditionally been subverted to basic needs. In the food vs. freedom debate, many Chinese citizens are content with a government that provides its citizens with the so-called positive rights of food, shelter, education, healthcare, and employment. *See* Steve Chan, *Human Rights in China and the United States: Competing Visions and Discrepant Performances*, 24 HUM. RTS. Q. 1035, 1039–40 (2002) (noting the identification of these rights as "positive" in China).

22. ZHENG YONGNIAN, *supra* note 9, at 54

23. *Id.*

24. *Id.* Zheng argues, quite cogently, that because the provinces are a part of the overall government hierarchy, they are constrained by the legitimacy of the center's rule in areas such as national defense and foreign relations. *See* ZHENG YONGNIAN, *supra* note 9, at 31–32, 40. However, reciprocity is only one of three ways in which the center and provinces interact, and as such, if the local interests are significant enough, the province will resist central policies.

25. *See id.* at 53 (describing the practice of coercion).

26. STARR, supra note 7, at 46. While this policy effectively stymied local upstarts from anointing themselves warlords, there was also a tradeoff. *Id.* That is, "[w]ith no local ties and more often than not speaking a dialect different from that of the staff and clientele at his new post, the magistrate came to rely heavily on the local gentry (with whom he could communicate in Mandarin) for intelligence about the community." *Id.*

27. ZHENG YONGNIAN, *supra* note 9, at 57.

28. *Id.* (quoting THE CHINESE COMMUNIST PARTY'S NOMENKLATURA SYSTEM, at ix (John P. Burns ed., 1989)).

29. *Id.* at 93. The center was wise to resist forcing policies on lower governments by manipulating personnel. According to Professor Zheng, "[t]he frequent use of the system, e.g., dismissals and purges, however, can be seen as a sign not of central strength, but of weakness. Put simply, the use of these coercive measure means implies the failure of normal institutions between the center and the provinces." *Id.*

However, by the late 1970s, the inefficiency of centralized planning had become increasingly clear, as evidenced by widespread famines and stagnant economic growth.[30] As a result, the Chinese central government was forced to reevaluate its fiscal policies and implement a revenue-sharing arrangement between the various levels of authority.

A. Fiscal Decentralization

China's process of administrative decentralization can be separated into three distinct phases: pre-1979, 1980–1993, and post-1994.[31] Prior to 1979, the fiscal structure was best described as "'unified revenue collection and unified spending,'" where all levels of government were "eating from one big pot."[32] Under this arrangement, the central government created spending plans for the provinces, which collected local revenue generated within their jurisdictions and transferred the receipts to Beijing.[33] Beginning in 1980, however, these policies shifted toward a more decentralized scheme involving a fiscal contracting system between adjacent levels of government.[34] Nicknamed "eating from separate kitchens," this decentralization policy decreased the center's share of revenue and at the same time filled local treasuries.[35]

Although top leadership encouraged the decentralization policy, it was not applied uniformly; instead, the central government negotiated contracts with each province on an ad-hoc basis.[36] This process allowed richer provinces to garner preferential treatment from Beijing.[37] Fiscal contracts stipulated that "central fixed revenue," including direct taxes from state-owned enterprises (SOEs) and customs duties, would be remitted to the central government; any remaining income, dubbed "local revenue," would be split between the central and provincial governments according to pre-determined sharing schemes.[38] These arrangements allowed local and provincial governments to retain

---

30. *See* Vaclav Smil, *China's Great Famine: 40 Years Later*, 319 BRIT. MED. J. 1619, 1619 ("The origins of the famine can be traced to Mao Zedong's decision, supported by the leadership of China's communist party, to launch the Great Leap Forward.").

31. Hehui Jin, Yingyi Qian & Barry R. Weingast, *Regional Decentralization and Fiscal Incentives: Federalism, Chinese Style*, 89 J. PUB. ECON. 1719, 1723 (2005).

32. *Id.*

33. *Id.*

34. *Id.*

35. STARR, supra note 7, at 154.

36. *Id.* at 16. Almost immediately, provincial leaders used their newfound power to respond to the market conditions in their respective jurisdictions. *See id.* Soon after the provincial leaders began using their superior knowledge of local circumstances, however, the divergence of interests between the center and its provinces became clear. *Id.* The reduced level of oversight from the center enabled both provincial and local officials to abuse their positions of power, and dishonesty became embedded in local affairs. *See* BECKER, *supra* note 6, at 42.

37. ZHENG YONGNIAN, *supra* note 9, at 364.

38. Hehui Jin et al., *supra* note 31, at 1723.

significant amounts of surplus revenue, thereby creating an incentive to maximize local economic activities.[39] The increased discretion afforded local authorities reflected the central government's acknowledgement that provincial governments were better positioned to effectively utilize information pertaining to their respective jurisdictions.[40] As economics scholar Hehui Jin explains, "local governments have better access to local information, which allows them to provide public goods and services that better match local preferences than the national government."[41] As a result of the increased efficiency of local authority, more and more economic policy decisions were delegated to provincial governments, and surpluses became easier to hide from Beijing.

The trend of local fiscal autonomy was reversed in 1994 when the contracting system was replaced by a "separating tax system," which redefined local revenue and permitted the collection of "extra-budgetary revenue."[42] The new policy was in part a reaction to the remarkable success of decentralization in the 1980s. Faced with a declining share of revenue, the central government initiated reforms establishing a new central tax collection agency.[43] However, because the extra-budgetary revenues were not shared with the central government, [44] lower level governments were free to exert even greater influence over the economic conditions of their respective provinces. By acknowledging the right of a province to collect and retain certain classes of revenue, the central government tacitly institutionalized power in the hands of local leaders, further cementing de facto federalism.

B. Economic Liberalization

The decentralization of authority was not limited to the tax structure. In 1979, several areas including the prosperous Guangdong province were designated "special economic zones."[45] This status allowed local administrators to institute lower tax rates and exercise greater control over economic development.[46] Five years later, the central government identified

---

39. *Id.* at 1726.

40. *See id.* at 1724; BECKER, *supra* note 6, at 42, 60.

41. Hehui Jin et al., *supra* note 31, at 1720.

42. *Id.* at 1723–24.

43. STARR, *supra* note 7, at 154. "The reform set as a goal that 80 percent of all tax revenue would be collected by central-government agents and only the remaining 20 percent by local collectors, which, if successful, would reverse the situation in which most revenue was retained by the provinces." *Id.* However, by this time, the provinces had crossed the point of no return. Consequently, the percentage of revenues collected by the center reached only 50%, well below its original goal. *Id.* at 155.

44. Hehui Jin et al., *supra* note 31, at 1724.

45. Gabriella Montinola, Yingyi Qian & Barry R. Weingast, *Federalism, Chinese Style: The Political Basis for Economic Success in China*, 48 WORLD POL. 50, 62 (1995).

46. *Id.* at 62–63. These special economic zones, for example, "ha[d] the authority to approve foreign investment projects up to $30 million, while other regions' authority remained much lower." *Id.* at 63.

fourteen "coastal open cities," which enjoyed similar development authority.[47] The market-oriented shift in Beijing policy was reinforced by local governments' increasing control over many industrial SOEs, beginning as early as 1979.[48] By 1985, only 20% of industrial SOEs were controlled by the central government.[49] The locally controlled SOEs benefited from steady streams of foreign investments, which allowed local governments even greater influence over regional development without substantial oversight from Beijing.[50]

The model of decentralization implemented between Beijing and the provinces was emulated by many provincial governments with respect to their own local governments and enterprises. In the province of Zhejiang, for example, private enterprises were increasingly tolerated and SOE managers were afforded more discretion in organizational establishment, labor recruitment, and profit distribution.[51] Additionally, increased autonomy encouraged provincial and local governments to promote the expansion of the collective sector.[52] Former SOEs, as well as schools and hospitals, were transformed into more flexible collectives.[53] The market's invisible hand increasingly guided business decisions in the coastal open cities and special economic zones.

Local governments also enjoyed increased discretion in economic policy-making. Therefore, as Professor Zheng indicates, "local governments began to pursue even greater power from the central government, and continuous decentralization became inevitable since local governments were not satisfied with the existing power distribution between the center and the provinces."[54] In many cases, this decentralization placed local governments outside the reach of the central government's coercion altogether. Indeed, "even if the central government [were] capable of executing political control over the provincial leadership through the Party's *nomenklatura* system, it [did] not have the effective mechanisms to execute control over local government officials below the provincial level."[55]

Globalization has exacerbated the cycle of decentralization and hyper-competition. Professor Zheng notes:

> Globalization has affected China's central-local relations with the creation of two opposite forces, i.e., decentralization and centralization. On one hand, globalization has decentralized economic activities further to local

47. *Id.*
48. *Id.* at 61.
49. *Id.* at 61–62.
50. *Id.* at 62.
51. ZHENG YONGNIAN, *supra* note 9, at 198.
52. STARR, *supra* note 7, at 85.
53. *Id.*
54. ZHENG YONGNIAN, *supra* note 9, at 367.
55. *Id.* at 246 (emphasis added).

> governments and other local organizations, making it increasingly difficult for the center to access local economic resources. On the other hand, globalization requires the center to regulate the national economy in order to accommodate external economic forces resulting from globalization.[56]

This tension between local economic development and national regulation eventually led to the center's inability to regulate intellectual property, foreign trade, and the environment.

C. Consequences of De Facto Federalism

The greatest consequence of decentralization was increased competition between local governments in order to attract foreign investment, produce goods, and take their SOEs public.[57] This increase in regional competition reduced the impediment of central planning and greatly contributed to the privatization of the Chinese economy.[58] Commentators have termed the end result "market-preserving federalism."[59] This unique political and economical phenomenon is consistent with Professor Zheng's "de facto federalism" and embodies the following propositions:

> [F1] A *hierarchy* of governments with a *delineated scope of authority* . . . exists so that each government is autonomous within its own sphere of authority.
> [F2] The subnational governments have primary *authority over the economy* within their jurisdictions.
> [F3] The national government has the authority to police the *common market* and to ensure the mobility of goods and factors across subgovernment jurisdictions.
> [F4] Revenue sharing among governments is limited and borrowing by governments is constrained so that all governments face *hard budget constraints*.
> [F5] The allocation of authority and responsibility has an *institutionalized degree of durability* so that it cannot be altered by the national government either unilaterally or under the pressures from subnational governments.[60]

---

56. *Id.* at 367.

57. Shaomin Li, Shuhe Li & Weiying Zhang, *Cross-Regional Competition and Privatisation in China*, 9 MOCT-MOST: ECON. POL'Y TRANSITIONAL ECONS. 75, 76, 82–83 (1999).

58. *Id.* at 84.

59. *E.g.*, Montinola et al., *supra* note 45, at 55. This distinctly Chinese brand of federalism must be contrasted from its Western counterpart, which Zheng characterizes as "a form of government that differs from unitary forms of government, in terms of the distribution of power between central and sub-national governments; the separation of powers within the government; and the division of legislative powers between national and regional representatives." ZHENG YONGNIAN, *supra* note 9, at 32.

60. Montinola et al., *supra* note 45, at 55.

In short, the development of these conditions created an institutional framework that limited the extent to which the "political system[] could encroach upon its markets."[61] Beijing effectively traded a substantial degree of regulatory and policy-making power in exchange for rapid industrialization. The central government's reforms allowed the free market to integrate with China's economy, in the process opening a Pandora's Box of market forces.

The essence of a free market is competition, and China's version of federalism has induced ferocious rivalries between local economies. Zhao Jiancai, the enterprising mayor of Zhengzhou, is a good example of the economic warlords who disregard Beijing's decrees in quest of political fame.[62] Zhao plans to transform Zhengzhou from the dreary capital of Henan, one of China's poorest provinces, into the "'Chicago of the East'—a gateway between the booming coast and the vast interior—by more than tripling the city's size."[63] A $100 million waterfront arts complex, complete with dancing fountains and laser beams, illustrates the type of hyper-investment for which local governments vigorously compete.[64] In fact, despite the central government's repeated mandates to curb spending, top planners have resorted to using satellites to "spot bulldozers working on illegal construction projects in far-flung provinces."[65]

Indeed, much like the ancient warlords who ruthlessly expanded their powerbases in the celebrated Three Kingdoms era, "[e]conomic growth has become the path to career glory for city mayors."[66] The swords and spears once wielded by local strongmen have been replaced by new weapons of choice—financial subsidies and tax incentives—yet the goals are ultimately the same: economic power and political prestige. A Hong Kong economist recently compared "modern Chinese cities to corporations, and their mayors to chief executive officers, all competing with each other to expand their business empires."[67] In fact, much to the chagrin of top-level planners, "[t]he majority of China's 660 cities and 2,200 country-level towns have or are planning to

---

61. Barry R. Weingast, *The Economic Role of Political Institutions: Market-Preserving Federalism and Economic Development*, J.L. ECON. & ORG. 1, 3 (1995). Zheng argues that a "formal institutional perspective can hardly help us understand China's central-local relations properly simply because of the lack of a sound legal infrastructure . . . . China has never developed a system of rule of law." ZHENG YONGNIAN, *supra* note 9, at 36. Bargaining between the central and lower governments is inescapable, and central administrators are forced to negotiate for the enforcement of laws, contracts, and regulations. *Id.*

62. Andrew Browne, *Booming Municipalities Defy China's Effort to Cool Economy*, WALL ST. J., Sept. 15, 2006, at A1.

63. *Id.*

64. *Id.*

65. *Id.* at A6.

66. *Id.*

67. *Id.* Professor Zheng has made a similar comparison: "Since the governments at different levels could benefit greatly from local economic growth, they acted like entrepreneurs: acquiring raw materials and energy for enterprises; providing the financial resources enterprises needed; and creating markets for enterprise products." ZHENG YONGNIAN, *supra* note 9, at 166.

build wide roads and lavish squares" in furtherance of local administrators' legacies.[68] Chinese scholars refer to such localism as *chu-hou ching-chi*, an economic system in which "thirty huge 'fiefs' (provinces, municipalities, and autonomous regions), three hundred middle-sized fiefs (prefectures, cities), and over three thousand small fiefs (counties, cities) . . . operate, or try to operate, independently."[69] According to Chien-min Chao, this model is characterized by local government protectionism designed to benefit local industries by banning non-local goods.[70]

Localism may be defined as "an 'attachment to a locality, especially to the place in which one lives' and 'limitation of ideas, sympathies, and interests growing out of such attachment.'"[71] Fueled by geographic disparities and local identities, the economic warlords in charge of autonomous local governments have become the "biggest power wielders, with the right to initiate new industries and enterprises and to control local finance, banking, commerce, and trade," and have even begun working with the military to further their coinciding interests.[72] In one case, an escalating trade war between Guangdong and Hunan led to the deployment of troops to stop the outflow of rice to surrounding provinces.[73] These minor skirmishes reflect an enduring mistrust between provinces that led China scholar Nicholas Kristof to posit a scenario in which "economic cross-border competition between [the] Hunan and Hubei Provinces erupts into a civil war that an impotent central government cannot bring under control."[74]

The automobile manufacturing feud between Shanghai and the Hubei province presents another example of the clash between rival fiefdoms that can

---

68. Fu Jing, *Ministries: Outlandish Roads, Public Squares Have to Stop*, CHINA DAILY, Feb. 24, 2004, at 2, *available at* http://www1.china.org.cn/english/China/88293.htm. Many of these projects amount to little more than novelty tourist destinations. For instance, officials in the province of Chongqing recently opened the doors to what they insist is the world's largest bathroom. Associated Press, *China Public Restroom Has 1,000 Stalls*, July 6, 2007, http://www.foxnews.com/wires/2007Jul06/0,4670,ChinaLargestBathroom,00.html. This four-story "porcelain palace" boasts bizarre amenities including urinals shaped like crocodile mouths and busts resembling the Virgin Mary. *Id.*

69. Chien-min Chao, *T'iao-t'iao Versus K'uai-K'uai: A Perennial Dispute Between the Central and Local Governments in Mainland China*, *in* FORCES FOR CHANGE IN CONTEMPORARY CHINA 158, 161 (Bih-jaw Lin & James T. Myers eds., 1993) (citing Shen Li-jen & Tai Yüan-ch'en, *The Origin and Formation of the 'Fief Economy' in Our Country and Its Pitfalls*, 3 CHING-CHI YEN-CHIU [ECON. RES.] 12 (1990)).

70. *Id.* at 164.

71. ZHENG YONGNIAN, *supra* note 9, at 231 (quoting THE OXFORD ENGLISH DICTIONARY 1468 (1987)).

72. Chien-min Chao, *supra* note 69, at 163.

73. *Id.* at 165.

74. STARR, *supra* note 7, at 158. Indeed, many have speculated on the possibility of the People's Liberation Army being dispatched to quell regional conflicts. *See, e.g.*, *id.* at 161. Yet, even this ultimate form of coercion may be ineffective in taming local competition, because the army itself "suffers from regional divisions." *Id.*

result from local protectionism. In 1998, Shanghai passed local regulations to protect its "flagship auto enterprise," Shanghai Volkswagen, effectively raising the price of cars bought outside of Shanghai by $9,600.[75] In defiance of a central government directive banning restrictions on inter-provincial trade, Hubei retaliated by similarly taxing cars built in Shanghai.[76] A spokesman for Hubei explained that the province had taken these measures in response to its neighbors, who also break the central government's rules.[77] "To play by the market rules when everyone else cheats would mean losing out to the competition, threatening the livelihoods of workers and endangering already shaky social stability—something China's modern rulers fear most."[78] In fact, many local officials identify their wealthier neighbors, such as Shanghai and Guangdong, as "colonial centers" and consequently refuse to cooperate, due to deep-seated feelings of victimization.[79] These common conflicts reflect pervasive provincial ambitions to create self-sustaining local economies that are free from the commercial harassment of rivals and neighbors.[80]

The problem of local governments' over-investment partly stems from the seemingly endless supply of credit. The global capital flows propelling China's economy are too large and complex for Beijing to control; despite moves by the central bank to restrict lending, loan growth exploded by more than 70% in the first quarter of 2006.[81] As the gleaming skylines of its cities expand, the rest of the world will continue to turn to China as the largest marketplace for goods and services, as well as a base for manufacturing operations. Yet, dangerous products, piracy, and widespread environmental degradation have caused great concern in government halls and corporate headquarters around the globe, forcing companies to reconsider the consequences of doing business in China. The steady stream of toy recalls and tainted products has catapulted the problems stemming from Chinese federalism to the forefront of the global trade

---

75. Antoaneta Bezlova, *Cars in Collision on China's Free-Market Highway*, ASIA TIMES ONLINE, Jan. 15, 2000, http://www.atimes.com/china/BA15Ad01.html.

76. *Id.*

77. *Id.*

78. *Id.* Popular dissent is on the rise; the number of mass protests (those involving more than 100 protestors) rose from 8,700 in 1993 to 74,000 in 2004. SUSAN L. SHIRK, CHINA: FRAGILE SUPERPOWER 56 (2007).

79. ZHENG YONGNIAN, *supra* note 9, at 366. Professor Zheng notes that "[l]ocal governments compete[] with one another for local development and use[] all possible administrative methods to protect local industries." *Id.* The practice of thwarting central authority has become so widespread that "[t]he World Bank [has] warned that individual provinces ha[ve] a tendency to behave like independent countries, with an increase in external (overseas) trade and a relative decline in trade flows with each other." *Id.*

80. *See* Chien-min Chao, *supra* note 69, at 158–161, 164–65 (describing the perpetuation of localism in China through forces such as "economic protectionism" and "Balkanization" of the economy).

81. Brian Bremner, *Is China Growing Too Fast for Comfort?*, BUSINESSWEEK, July 18, 2006, http://www.businessweek.com/globalbiz/content/jul2006/gb20060718_426870.htm?chan=search.

debate and alerted millions of parents to the dangers of Chinese products.[82] Accordingly, before a company commits the millions of dollars necessary to set up shop in China, it must have some assurance that production techniques will be followed and that manufacturing technology will not be misappropriated.

II. CHINA'S INTERNATIONAL INTELLECTUAL PROPERTY OBLIGATIONS

China's intellectual property laws have come a long way in only twenty years. In the mid-1980s, there were virtually no significant laws against piracy or counterfeiting on the books in Beijing.[83] The modern patent law enacted in 1985 did not cover crucial industries such as chemicals and pharmaceuticals.[84] In response, many multi-national corporations began to clamor for meaningful protection of intellectual property rights.[85] This debate over intellectual property rights essentially pitted the developing and developed worlds against each other.[86] Negotiations stalled during the Uruguay Round of the General Agreement on Tariffs and Trade (GATT) talks, because developing countries such as India and Brazil resisted the incorporation of intellectual property protections into the umbrella of agreements governing international trade.[87] The developing nations argued that, if strong intellectual property restrictions were implemented, there "would be less of an opportunity to catch up to more advanced nations[,] and the gap between rich and poor would continue to expand."[88] Eventually, threats of trade sanctions convinced the developing nations that intellectual property rights were in fact trade issues appropriately included in the GATT agenda.[89]

This understanding led to the adoption of the Agreement on Trade-Related Aspects of International Property Rights (TRIPS) in 1994.[90] TRIPS "provided broader protections for intellectual property rights by granting most favored nation treatment for all signatories, establishing minimum terms of protection, imposing significant local enforcement and dispute settlement requirements, and authorizing trade sanctions against noncompliant nations."[91] Instead of

---

82. *See, e.g.*, David Schaper, *Thomas Tank Engine Toy Recall Angers Parents*, NPR: MORNING EDITION, June 22, 2007, http://www.npr.org/templates/story/story.php?storyId=11271805.

83. Joseph A. Massey, *The Emperor is Far Away: China's Enforcement of Intellectual Property Rights Protection, 1986–2006*, 7 CHI. J. INT'L L. 231, 231 (2006).

84. *Id.* at 231–32.

85. Robert C. Bird, *Defending Intellectual Property Rights in the BRIC Economies*, 43 AM. BUS. L.J. 317, 322 (2006).

86. *Id.* at 322–24.

87. *Id.* at 322–23.

88. *Id.* at 323–34 (citing Elizabeth Chien-Hale, *Asserting U.S. Intellectual Property Rights in China: Expansion of Extra-Territorial Jurisdiction?*, 44 J. COPYRIGHT SOC'Y U.S.A. 198, 226 (1997)).

89. *Id.* at 324.

90. *Id.*; TERRILL, *supra* note 13, at xv.

91. *Id.* at 324–25.

requiring signatories to implement the precise language of the agreement, TRIPS merely "provides standards and aims for the member states to integrate those standards into national legislation"; thus, member countries have wide discretion in developing judicial procedures and enforcement measures.[92] Nonetheless, China has to a large extent based its modern intellectual property laws on those of other Western nations.[93] In fact, critics assert that the "carbon copying of concepts and rules" is part of China's intellectual property (IP) enforcement problem, because administrators do not fully understand the rationale behind the rules.[94] Despite common misconceptions, however, these new standards were not China's first attempt at safeguarding IP rights.

A. Chinese Trademark Laws

The Chinese are responsible for some of the most influential inventions in human history, including gunpowder, movable type, and the compass.[95] Efforts to protect inventions are well documented, with evidence of the earliest trademarks being used as far back as the Northern Zhou Dynasty (556–580 A.D.).[96] Over time, as increasing numbers of merchants began to identify their crafts with logos, the use of trademarks became an important method of promoting the quality of Chinese products.[97] The first formal trademark law was enacted in 1904 by the Qing Dynasty (1644–1912 A.D.); however, the administration of the legal safeguards was handled mostly by foreigners, who substantially controlled China's trade during that period.[98] Successive governments modified and revised the underlying legal framework for many years until socialist ideology eventually dictated appropriation of intellectual property for the benefit of industry and commerce.[99] The Cultural Revolution was a time of widespread intellectual persecution that stymied innovation; as a result, the Chinese intellectual property rights regime suffered substantially.[100]

This trend was reversed by Deng Xiaoping in the late-1970s, when economic reforms injected the "concept of property rights into the socialist lexicon."[101] By opening China to the outside world and to foreign sources of capital, Deng Xiaoping set in motion a chain of events that would fill the

92. Brigitte Binkert, *Why the Current Global Intellectual Property Framework Under TRIPS Is Not Working*, 10 INTELL. PROP. L. BULL. 143, 144 (2006).

93. Jessica Jiong Zhou, *Trademark Law & Enforcement in China: A Transnational Perspective*, 20 WIS. INT'L L.J. 415, 416 (2002).

94. *E.g.*, *id.*

95. *Id.* at 417.

96. *Id.*

97. *Id.* at 417–18.

98. *Id.* at 418.

99. *See generally id.* at 418–21 (tracing China's transition into its modern phase of trademark law).

100. *Id.* at 420–21.

101. *Id.* at 421.

Chinese market with world class foreign brands that, at the time, few Chinese citizens could dream of owning.[102] These economic reforms ultimately led to the emergence of de facto federalism,[103] which placed the burden of protecting the intellectual property rights of foreigners on local governments. Meanwhile, the local governments' lax protection of intellectual property rights created endless opportunities for pirates to pilfer foreign technology and infringe upon registered trademarks.[104]

By 1993, China had signed several international agreements and promulgated extensive trademark regulations, but critics were still not satisfied with its enforcement initiatives.[105] New laws were passed, increasing administrative fines, refining the definition of infringements, and allowing the Administration of Industry and Commerce to order payment of damages.[106] Unfortunately, the economic warlords presiding over local governments ensured that these types of fines and damages were rarely levied upon violators.[107]

Eventually, the Seventh National People's Congress amended trademark laws, aiming to enhance "'the administration of trademarks' and encourage 'producers to guarantee the quality of their goods and maintain the reputation of their trademarks, with a view to protecting consumer interests.'"[108] The amendments also refined the definition of infringement "to include the sale of goods that one is 'fully aware' are counterfeits of a registered mark, the forgery or unauthorized manufacturing of representations of another's registered trademark, and the sale of trademark representations that were forged or manufactured without authorization."[109] A special Intellectual Property Rights Tribunal was created in Beijing to deal exclusively with infringement cases.[110] Satellite courts have been established in several other large markets; however, these courts still suffer from a lack of qualified legal administrators.[111] The Chinese legal system's "inquisitorial" nature aggravates the need for competent

---

102. *See* STARR, *supra* note 7, at 79.

103. *See supra* Part I.

104. *See* DANIEL C.K. CHOW, THE LEGAL SYSTEM OF THE PEOPLE'S REPUBLIC OF CHINA: IN A NUTSHELL 439 (2003) [hereinafter CHOW, THE LEGAL SYSTEM].

105. Zhou, *supra* note 93, at 427.

106. *Id.*

107. CHOW, THE LEGAL SYSTEM, *supra* note 104, at 440. "Under the reformist regime, the local government gained great autonomy in decision-making. . . . It now became a strong protector of the local community by protecting it from outside attack or criticism. Because the local community did not need to face pressure from above directly, its economy followed its own track." ZHENG YONGNIAN, *supra* note 9, at 211.

108. Zhou, *supra* note 93, at 427.

109. *Id.* (quoting Hamideh Ramjerdi & Anthony D'Amato, *The Intellectual Property Rights Laws of The People's Republic of China*, 21 N.C. J. INT'L L. & COM. REG. 169, 178–79 (1995)).

110. *Id.* at 431. This is the same court that observers have recently commended for prosecuting "egregious" infringers. *See, e.g., id.*

111. *Id.*

individuals who will vigorously prosecute violators of intellectual property rights without being swayed by local influence.[112]

Efforts to implement a viable intellectual property enforcement regime were greatly undercut in 1994 by the aforementioned third phase of Chinese economic reform, which created the separating tax system.[113] By allowing the local governments to retain tax money, the central government institutionalized incentives for local governments to support short-term projects that quickly realize taxable profits.[114] Trademark counterfeiting operations significantly contributed to local treasuries and therefore garnered local protection.[115] Local governments resolved the conflict between enforcing national intellectual property laws and protecting local trademark infringers by taking advantage of the Chinese political structure. Although the national government in Beijing creates policies and laws, provincial legislatures are responsible for promulgating "local laws and regulations to implement laws made by the national legislature."[116] Consequently, local lawmakers often fail to incorporate national mandates into local regulations; thus, it is difficult for law enforcers to locate applicable laws with which to prosecute offenders.[117]

By 2000, China had joined almost all the major international treaties and conventions, including the Paris and Madrid agreements, but the powerful market mechanisms driving China's rapid growth were also making effective enforcement of intellectual property rights extremely difficult.[118] In 2001, China again revised its trademark laws in order to gain admission to the World Trade Organization.[119] The amendments brought China into compliance with TRIPS and set out new guidelines for a first-to-file registration system.[120] Yet, the underlying problem persists: inadequate punishments fail to deter infringers from pursuing the lucrative counterfeiting trade.[121] In this climate of rapid industrialization, millions of Chinese entrepreneurs embrace the chance to create wealth by any means necessary; as Deng Xiaoping stated, "[T]o get rich is glorious."[122] Thus, economic interests have trumped legal considerations in China's approach to trademark enforcement over the last two decades. In the

---

112. *Id.* at 431–32.

113. *See supra* Part I.A.

114. Zhou, *supra* note 93, at 435.

115. *Id.*

116. *Id.* at 434.

117. *Id.* at 435. "Often times, a brief newspaper account is the only source for people to learn about important new laws. While the [National People's Congress] and the Supreme People's Court publish laws and judicial opinions in official gazettes, their provincial counterparts usually do not." *Id.* Thus, local laws remain unclear. *Id.*

118. *See* CHOW, THE LEGAL SYSTEM, *supra* note 104, at 439–40.

119. *See* Peter K. Yu, *From Pirates to Partners (Episode II): Protecting Intellectual Property in Post-WTO China*, 55 AM. U. L. REV. 901, 903–04 (2006).

120. CHOW, THE LEGAL SYSTEM, *supra* note 104, at 418–19.

121. Zhou, *supra* note 93, at 433–34.

122. Melinda Liu, *Mao to Now*, NEWSWEEK, Dec. 31, 2007, at 41, 45.

area of copyright, on the other hand, disregard for the creative works of authors and artists is more deeply rooted.

B. Chinese Copyright Laws

Despite evidence of early copyright laws dating back to the Tang Dynasty (618–906 A.D.), Chinese heritage and cultural characteristics exhibit a natural opposition to the concept of copyright.[123] For instance, "the yin-yang concepts of *Li* and *Fa* leave the Chinese people predisposed against" the idea of individual profit from literature or art.[124] Early Chinese law and Confucian principles, such as *Li*, dictated "that the individual should be submerged in the collective [and] that the individual-society relationship should be non-competitive."[125] The ancient Chinese believed that knowledge was a "cultural good," the value of which was maximized when distributed to society in general and not ascribed to the efforts of a single individual.[126] Similarly, Confucius taught that knowledge belongs to society and should be broadly disseminated, not individually owned.[127]

When the Communists seized power in 1949, they "sought to abolish all forms of private property" and severely limited intellectual property rights, as they "rewarded individual creativity and private initiative during a period in which the Party stressed collective endeavor and common ownership."[128] This philosophy has become ingrained in Chinese society, as evidenced by the billions of pirated DVDs peddled in markets across the country every year, many of which hit the streets the same week as their Hollywood premieres.[129] Microsoft suffers from rampant software piracy in China, where citizens commonly buy illegal copies of the Windows operating system for pennies on the dollar.[130] Some studies conclude that as much as 94% of all software used in China is pirated.[131] Piracy is distinguishable from its cousin, trademark counterfeiting, in that those pirating copyrighted material do not attempt to convince their clientele of their merchandise's authenticity.[132] Instead, the quality of the pirated content itself entices customers to purchase unauthorized copies of intellectual property. Chinese businesses experience tremendous

---

123. Graham J. Chynoweth, *Reality Bites: How the Biting Reality of Piracy in China Is Working to Strengthen Its Copyright Law*, DUKE L. & TECH. REV., Feb. 11, 2003, http://www.law.duke.edu/journals/dltr/articles/pdf/2003DLTR0003.pdf.

124. *Id.*

125. *Id.*

126. Binkert, *supra* note 92, at 146.

127. *Id.* at 147.

128. CHOW, THE LEGAL SYSTEM, *supra* note 104, at 411.

129. TED C. FISHMAN, CHINA INC.: HOW THE RISE OF THE NEXT SUPERPOWER CHALLENGES AMERICA AND THE WORLD 235 (2005).

130. *Id.* at 244–45.

131. CHOW, THE LEGAL SYSTEM, *supra* note 104, at 434.

132. *Id.* at 436.

savings by purchasing pirated software, and the fierce competition driving China's economy makes cost-cutting measures an indispensable weapon in the economic warlord's arsenal.

In the Chinese businessman's eyes, there is justification for the theft of foreign intellectual property, such as Microsoft's software. Many entrepreneurs contend that Microsoft is not really losing a source of revenue through Chinese piracy, because the majority of Chinese businesses cannot afford to purchase the software legally.[133] Thus, under this theory, because the license to operate Windows would not be sold in the first place, there is no ethical reason to prevent a shrewd businessman from obtaining the software at a steep discount. Piracy is also justified on moral grounds related to the theory of "reverse colonialism," which posits that a former colony should be allowed to benefit from the research and development of more advanced countries.[134] The memories of foreign occupation and exploitation are powerful motivators, and for the foreseeable future, it appears that this type of reasoning is likely to prevail.

China's copyright law, like its trademark law, was most recently revised in 2001 as part of the country's effort to conform to TRIPS.[135] Pursuant to the Berne Convention, copyright protection was extended to "works of literature, art, natural sciences, social sciences, engineering, and technology."[136] The copyright term was increased to the author's life plus fifty years, and foreign copyright holders are now afforded protection.[137] The revisions expand judicial remedies to permit preliminary injunctions, encourage the transfer of administrative cases to judicial venues for criminal prosecutions, and allow compensation of up to $60,000.[138] Thanks to these recent updates, China's national laws now adhere to the World Trade Organization and TRIPS requirements, and they parallel sophisticated Western IP statutes. China's political and economic realities, however, still pose significant barriers to the effective implementation of these new laws. As long as the commercial benefits of piracy and counterfeiting outweigh the tenacity of law enforcement officials, intellectual property theft will continue unabated.

III. WARNING: MADE IN CHINA

In China, counterfeiting foreign brands and selling knockoffs at cut-rate prices is big business. In 2004, counterfeit products accounted for more than 8% of the nation's gross domestic product, and observers estimate that as much as two-thirds of the world's bogus products are made in China.[139] Consumers

133. FISHMAN, *supra* note 129, at 246.
134. *Id.* at 252.
135. CHOW, THE LEGAL SYSTEM, *supra* note 104, at 428.
136. *Id.*
137. *Id.* at 429–30.
138. *Id.* at 431.
139. A.T. KEARNEY, THE COUNTERFEITING PARADOX (2005), http://www.atkearney.com/

can find counterfeit versions of virtually any product—from exploding phony Heineken bottles to exquisitely crafted fake Rolexes—in the stalls lining markets and bazaars across China.[140] In fact, pirated technology enables Chinese forgers to make copies of wristwatches sufficiently sophisticated to fool even Swiss watchmakers.[141] These counterfeit products, however, are not always of a quality similar to the originals—despite the recent uproar, defective products are nothing new on the Mainland.

Although some customers are pleased with their counterfeits, most are appalled when they learn of the risks and uncertainty accompanying a great spectrum of Chinese goods. For instance, scores of unsuspecting tourists and even some scientists have been duped by hustlers peddling fake fossils.[142] In short, anyone purchasing exotic goods, mundane items, or even an ice cold beverage while in China should exercise a large degree of *caveat emptor*. Yet, despite the well-publicized risks, customers travel miles—even coming from other countries—to purchase superb, albeit spurious, merchandise.[143] Unfortunately, frequent international news reports of unsafe Chinese goods demonstrate that these counterfeit and defective products are no longer confined to the Middle Kingdom.[144]

A. The Economics of Counterfeiting

In reality, counterfeit goods sold in places like Beijing's Silk Alley or Huai Hai Road in Shanghai form the cornerstone of many local economies. Counterfeiting creates jobs and supports other local businesses, such as warehouses, restaurants, and hotels.[145] The start-up costs of a counterfeiting operation are low, but its potential profits are enormous. United States businesses lose between $200 and $250 billion every year to counterfeiters, and the global counterfeiting trade is estimated to be worth half a trillion dollars

---

shared_res/pdf/Counterfeiting_Paradox.pdf.

140. FISHMAN, *supra* note 129, at 231. Phony bottled Heinekens and Budweisers are known for their unpredictable tendency to explode in the hands of unwitting victims. *Id.*

141. *Id.* at 239.

142. PaleoDirect.com, Fake Chinese Fossils, http://www.paleodirect.com/fakechinesefossils1.htm (last visited Feb. 3, 2008). While some of these fossils are described as magnificent specimens, others resemble fantastic creatures (such as the half-rat half-fish phony that recently fetched over three thousand dollars on eBay). *Id.*

143. FISHMAN, *supra* note 129, at 237–38.

144. *See, e.g.*, Mindy Fetterman, *Traditional Toy Sales Skirt China Recalls*, USA TODAY, Dec. 24, 2007, at B1 (noting the "series of high-profile recalls of Chinese-made toys" that occurred in 2007); Schaper, *supra* note 82 (discussing the impact of a Thomas the Tank Engine toy recall); Agence France-Presse, *Tree Harness, Helmet, Teethers: The Latest Made-in-China Recalls*, Dec. 18, 2007, http://afp.google.com/article/ALeqM5jVWx8B4noytTJFC183ptweCuaBvg.

145. *See* FISHMAN, *supra* note 129, at 238 (noting counterfeiting's importance to the local economy in the city of Yiwu, located in the Zhejiang Province).

annually.[146] The lucrative business has been compared to the drug trade, where the risk of violence and prolonged jail sentences induce profit margins of around 300%.[147] Counterfeiting and piracy, which present less risk of long-term incarceration, permit profit margins as high as 900%.[148] Thus, counterfeiting has become so rewarding that terrorist networks have recently begun conducting counterfeiting businesses to finance operations.[149]

Criminal organizations routinely use their economic clout to exert pressure on local officials to protect their illicit enterprises; as a result, "counterfeiting is heavily defended at local levels."[150] Moreover, counterfeiting activities have been so successful in Zhejiang and Guangdong that other towns seek to emulate their economic models.[151] By skipping the expensive steps of raising large amounts of capital, researching and developing products, and building brand awareness through costly advertising, counterfeiters are able to turn profits almost immediately.[152]

If counterfeit products are of a certain level of quality, they benefit local populations by allowing access to consumer goods—such as food, hygiene products, and medicine—that otherwise might be unaffordable to many citizens. However, great risks exist in the context of phony pharmaceuticals (such as those approved by Zheng Xiaoyu), which are estimated to account for 11% of global drug commerce per year.[153] In 2001, for example, "an estimated 192,000 people died in China because of counterfeit drugs, and as much as 50% of China's drug supply is counterfeit."[154] Given the illegal activity surrounding the counterfeiting trade, Beijing should be eager to crack down on the flow of fake goods, many of which are killing Chinese citizens. But despite the central government's best efforts, many anti-counterfeiting laws are simply ignored by the local governments charged with enforcement.[155] A similar failure of enforcement also hinders regulation of legitimate products, which can be tainted by the dangerous cost-cutting methods of unscrupulous manufacturers.[156]

---

146. Michael M. DuBose, *Criminal Enforcement of Intellectual Property Laws in the Twenty-First Century*, 29 COLUM. J.L. & ARTS 481, 483 (2006).

147. *Id.* at 483–84.

148. *Id.*

149. *Id.* at 485–86.

150. FISHMAN, *supra* note 129, at 238.

151. *Id.*

152. A.T. KEARNEY, *supra* note 139.

153. DuBose, *supra* note 146, at 481.

154. *Id.* (citing Amy Reeves, *Clamping Down on Counterfeit Drugs*, INVESTOR'S BUS. DAILY, Oct. 20, 2003, at A9).

155. *See* Fareed Zakaria, *The Rise of a Fierce Yet Fragile Superpower*, NEWSWEEK, Jan. 7, 2008, at 38–39 ("On almost every issue . . . the central government issues edicts that are ignored by the provinces.").

156. *See* Pete Engardio & Dexter Roberts, *The China Price*, BUSINESSWEEK, Dec. 5, 2004, at 102, 104 (discussing the effect of the "China Price" on United States manufacturing).

B. Defective Products and the Hidden Costs of the "China Price"

The single greatest factor propelling China's breakneck economic growth is referred to by supply chain managers worldwide as the "China Price," a term denoting the huge price gap between domestic production and production in China.[157] Enabled by currency manipulations and the astounding cheapness of Chinese labor, the China Price of manufacturing is typically 30–50% lower than average costs in the United States.[158] These tremendous savings fuel the perennial exodus of American factory jobs across the Pacific.[159] Not surprisingly, the greatest beneficiary of the China Price is Wal-Mart, the world's largest retailer.[160] Beloved by budget-conscious consumers and pilloried by fair trade critics, Wal-Mart has earned a reputation for ruthless procurement policies by forcing its suppliers to constantly roll back prices.[161] This policy has exacerbated the migration of manufacturing jobs, as 70% of Wal-Mart's products are now made in China.[162]

The enormous pressure on manufacturers to meet this pricing scheme invariably leads to production shortcuts, as evidenced by massive recalls of products such as toothpaste, tires, and toys. After the recall of over eighteen million toys due to excessive levels of lead paint and hazardous magnets, global toy giants are reeling from dangerous Chinese manufacturing practices.[163] Much of the problem arises from China's production infrastructure—"[o]n average, it takes China 17 separate parties to produce a product that would take [the United States] three."[164] Furthermore, the Consumer Products Safety Commission (the American agency charged with monitoring imports) relies primarily on the industry's self-regulation.[165] In fact, "[o]nly 15 trained inspectors regularly monitor goods at U.S. ports, where hundreds of millions of toys—about three quarters from China—come in yearly. Fewer than 100 inspectors have to cover the rest of the country, scouring store shelves for safety problems among the 15,000 products regulated by CPSC."[166]

---

157. *See id.*

158. *Id.* at 104.

159. *Id.* This exodus cost over 2.7 million U.S. workers their livelihoods between 2000 and 2004. *Id.*

160. Jiang Jingjing, *Wal-Mart's China Inventory to Hit US$18b This Year*, CHINA DAILY (BUSINESS WEEKLY), Nov. 29, 2004, http://www.chinadaily.com.cn/english/doc/2004-11/29/content_395728.htm.

161. Fareed Zakaria, *Does the Future Belong to China?*, NEWSWEEK, May 9, 2005, at 26.

162. Jiang Jingjing, *supra* note 160.

163. *See* Shu-Ching Jean Chen, *Trapped in the Chinese Toy Closet*, FORBES, Aug. 21, 2007, http://www.forbes.com/markets/2007/08/21/china-toy-industry-markets-equity-cx_jc_0821markets1.html (noting Mattel's decision to "begin to look elsewhere" for manufacturing opportunities after its massive recall).

164. Jeremy Haft, *The China Syndrome*, WALL ST. J., July 16, 2007, at A12.

165. Michael Weisskopf, *Who Regulates America's Toymakers?*, TIME, Aug. 18, 2007, http://www.time.com/time/business/article/0,8599,1654132,00.html.

166. *Id.*

Given China's culture of corruption and the poor regulation on both sides of the Pacific, it is little wonder that Chinese products have fallen under heightened scrutiny. As such, the traditionally contentious trade disputes involving quotas and dumping are increasingly taking a backseat to safety concerns.[167] Yet, before any significant progress can be made, the structural enforcement factors contributing to de facto Chinese federalism's sweeping impact must be comprehensively addressed from the bottom up.

C. Enforcement Impediments

The attitudes of local governments assigned the massive job of administering China's intellectual property and consumer safety laws are best captured by an old Chinese adage: "The mountains are high and the emperor is far away."[168] Unlike in the 1980s, when the Chinese government itself was involved in pirating United States software, Beijing has promulgated a number of laws to curb the practice of counterfeiting in recent years.[169] After extensive negotiating, the first major revision in China's rather toothless intellectual property regime was announced in 1992.[170] A bilateral agreement with the United States compelled China to enter the Berne and Geneva Conventions.[171] Yet, even as the ink was drying on the agreement, Chinese resolve was called into question when a senior provincial leader of Guangdong, a well-known counterfeiting hotbed, claimed that the agreement was irrelevant in his province.[172]

This reaction was typical of the economic warlords, who were reluctant to shut down operations that amassed large amounts of local revenue. In the early 1990s, the economic reforms of Deng Xiaoping and Zhao Ziyang were just beginning to disperse the power of regulating China's economy from Beijing to the provincial capitals.[173] Flush with newfound power, local leaders were unwilling to divert the extra money and manpower required to enforce an agreement made by the "emperor" many miles away in Beijing.[174] A similar tension between northern and southern China is well documented throughout Chinese history, from the days when the kingdoms of Wei and Wu battled for supremacy of the Mainland until as recently as 1917, when Sun Yat-sen challenged the North's legitimacy by forming a rival government in Guangzhou.[175] The economic and legislative freedoms bestowed upon the

---

167. Andrew Batson & Lauren Etter, *Safety Becomes a Hot Trade Issue*, WALL ST. J., July 16, 2007, at A4.

168. Massey, *supra* note 83, at 231.

169. *Id.* at 232.

170. *Id.* at 235.

171. *Id.*

172. *Id.*

173. ZHENG YONGNIAN, *supra* note 9, at 291–95.

174. *See* STARR, supra note 7, at 197.

175. ZHENG YONGNIAN, *supra* note 9, at 236. For a classic Chinese historical novel

south as a special economic zone further emboldened its leaders to ignore Beijing's fiats.[176]

Local governments' refusals to adequately enforce the national intellectual property regime are a direct reflection of de facto federalism's evolution. The burdensome restrictions of central government regulations serve to reduce revenue flowing into the local governments' treasuries.[177] This conflict of interests creates a strong disincentive for local governments to enforce any constraint on economic activity.[178] Many "businesses that sell counterfeit and infringing goods . . . negotiate a fixed amount of taxes to be paid to the local government."[179] In addition, the jobs created through counterfeiting and the legitimate businesses that they support have integrated illegal activity into the local economies to an extent that many local residents "are ready to use any means necessary to protect" counterfeiting, as it is in their economic interests.[180] Professor Daniel C.K. Chow has noted that a "crackdown on counterfeiting would result in shutdown of the local economy with all of the attendant costs of unemployment, dislocation, social turmoil, and chaos."[181]

China's failure to enforce its agreement with the United States led to the adoption of further bilateral treaties that created committees to monitor and prevent counterfeiting in the provinces.[182] Much of the distribution of counterfeit products takes place in large open-air markets, which "are established and regulated by the local Administration of Industry and Commerce (AIC), a branch of the local government responsible for promoting, regulating, and policing commercial activity" and enforcing anti-counterfeiting laws.[183] As Professor Daniel C.K. Chow describes, "AICs are faced with a conflict of interest[,] as they are charged with policing and enforcing the very markets in which AICs and the local government have a substantial investment and financial interest."[184] Systemic corruption is another factor preventing the enforcement of intellectual property rights. Journalist Nicholas Zamiska recently observed that China's so-called crackdown on corruption is a one way fight; Chinese officials limit their prosecution of bribery to the officials who

---

discussing the rise of the Wei, Wu, and Shu kingdoms after the fall of the Han Dynasty, see generally GUANZHONG LUO, ROMANCE OF THE THREE KINGDOMS (C.H. Brewitt-Taylor trans., Charles E. Tuttle Co., 1959) (circa 1330).

176. ZHENG YONGNIAN, *supra* note 9, at 258.

177. Hehui Jin et al., *supra* note 31, at 1726.

178. *Id.*

179. FISHMAN, *supra* note 129, at 238.

180. *Intellectual Property Protection as Economic Policy: Will China Ever Enforce Its IP Laws?: Roundtable Before the Cong.-Exec. Commission on China*, 109th Cong. 29 (2005) (statement of Daniel C.K. Chow, Professor, Ohio State University College of Law) [hereinafter *Intellectual Property Protection*].

181. *Id.*

182. Massey, *supra* note 83, at 235–36.

183. *Intellectual Property Protection*, *supra* note 180, at 29.

184. *Id.*

receive it, for the most part ignoring the private entities making the bribe.[185] The direct involvement of many local governments in the distribution of illegal counterfeit goods exacerbates the problem. For instance, in the Zhejiang province, local officials have invested millions of dollars in markets now famous for their selection of dubious goods.[186] This blatant disregard for the intellectual property regime and the culture of corruption[187] in which many Chinese government officials thrive hinders the protection of intellectual property rights.

Furthermore, competition between provinces fosters an environment in which each provincial government ignores central government mandates that would cause it economic harm out of fear that other provinces will not enforce the restriction in order to gain a competitive advantage. Additionally, the sheer size of China's market forces hyper-competition between provinces and encourages officials to disregard any law that will constrain local economies. Thus, the decentralization of administrative and economic authority has institutionalized de facto federalism and, in effect, tied the hands of policymakers in Beijing.

D. Judicial Issues

The protection of intellectual property rights is also stifled by a weak legal infrastructure that fails to deter criminal activity. Although newspapers frequently feature stories of government raids on "massive counterfeiting operations," the seemingly endless supply of counterfeit goods persists.[188] While current laws make it "relatively easy to obtain an administrative action in the form of a raid," the fines levied on violators are often insignificant and fail to deter counterfeiters.[189] For example, the average fine for counterfeiting in 2000 was a mere $794, and administrators sent only one in five hundred cases to judicial authorities for criminal prosecution.[190] Often, merchandise is simply confiscated, and offenders remain free to return to clandestine warehouses and reload their specious stock.[191] With such symbolic punishments, it is

185. Nicholas Zamiska, *China Targets Bribe Takers, but What About Givers?*, WALL ST. J., July 9, 2007, at A6. Despite repeatedly emphasizing its commitment to ending corruption, Chinese officials barred a newspaper article calling for prosecution of bribe offerors: "[A] reporter at China Business News, a Shanghai-based newspaper that published a similar article detailing the bribes, was suspended from work for a month, along with the paper's assistant chief editor . . . ." *Id.*

186. A.T. KEARNEY, *supra* note 139.

187. *See generally* Guilhem Fabre, *State, Corruption, and Criminalisation in China*, 53 INT'L SOC. SCI. J. 459 (2001) (examining the limited scope of functionalist and culturalist interpretations of systemic corruption throughout a decentralized China).

188. FISHMAN, *supra* note 129, at 235.

189. *Intellectual Property Protection*, *supra* note 180, at 29.

190. *Id.* at 29–30.

191. *See id.* (noting that after seizures of property, "whatever sanctions are meted out do not create detterence").

unsurprising that local governments have spurned Beijing's demands to shut down the lucrative businesses that raise tax revenue and support the local citizenry. The highly publicized raids that occur are more likely the result of "turf wars among the government fiefdoms that are themselves knee-deep in counterfeiting" than any genuine crackdown.[192]

One recent court ruling, however, has given anti-counterfeit crusaders some hope. In December 2005, Judge Shao Minyan of the Beijing Second Intermediate Court ordered the owner of a local market to pay $24,800 in damages to brand owners of counterfeit merchandise sold by the market's vendors.[193] The plaintiffs' attorneys are pushing for a "two-strike rule" that would subject all future leases to a provision requiring a 30-day suspension for a vendor's first counterfeiting offense and a termination of the lease upon a second violation.[194] This particular crackdown has been dubbed an attempt to protect Beijing's reputation in anticipation of the 2008 Olympics.[195] Because local officials outside Beijing are subject to less scrutiny by the central government, it is unlikely that similar judgments will spread to distant areas, such as Zhejiang and Guangdong.

The general lack of judicial response can partly be attributed to the difficulty of proving counterfeiting.[196] In addition, the factories producing both counterfeit and defective goods are often dispersed in the countryside and difficult to locate.[197] Trained intellectual property prosecutors are scarce; as a result, the maximum sentences are rarely exercised.[198] Moreover, despite the recent decision in Beijing, courts seldom award compensation for damages, instead granting injunctions against future infringement.[199] In sum, these factors make enforcement a colossal challenge at a time when China's place in the global economy is of supreme importance. Indeed, the structural problems plaguing the enforcement of intellectual property and consumer safety laws pervade the Chinese legal system, and nowhere is this administrative conundrum more pronounced than in the area of environmental protection.

IV. ENVIRONMENTAL DEGRADATION

In a world increasingly concerned with climate change and carbon-footprints, China continues to flirt with impending ecological catastrophe. In June 2007, a report released by the Netherlands Environmental Assessment Agency announced that China had surpassed the United States as the world's

192. FISHMAN, *supra* note 129, at 236.

193. Dexter Roberts, *A Bigger Stick Against Chinese Fakes*, BUSINESSWEEK, Jan. 10, 2006, http://www.businessweek.com/bwdaily/dnflash/jan2006/nf20060110_3904_db039.htm.

194. *Id.*

195. *See id.*

196. A.T. KEARNEY, *supra* note 139.

197. *Id.*

198. *Id.*

199. Zhou, *supra* note 93, at 433.

largest emitter of carbon dioxide.[200] However, this ominous changing of the guard in no way reveals the sobering extent of China's environmental nightmare. China is home to nine of the world's ten most polluted cities.[201] In fact, at least 90% of Chinese cities fail the central government's relatively lenient clean-air standards, and the city of Benxi is so polluted that it occasionally becomes completely enveloped in smog, disappearing from view on satellite maps.[202] Some studies indicate that more than 25% of California's pollution may be traced across the Pacific to China's factories.[203] Scientists studying this phenomenon report that plumes of aerosols and pollutants stretch up to three hundred miles wide and six miles deep; once aloft, these toxic rivers can circle the globe in as little as three weeks.[204] Air pollution is so pervasive that the World Bank recently suggested China's air quality could soon be responsible for as many as 750,000 deaths annually.[205] As if noxious air were not enough for officials to combat, China's water woes present an even greater challenge.

The Yangtze River, China's longest, suffers from "cancerous" pollution, prompting experts to predict that its death may occur in less than five years.[206] One third of the species populating the Yellow River are now extinct.[207] This devastation is undoubtedly caused by 21,000 chemical plants located on shores of China's rivers and coastline, 11,000 of which line the Yellow and Yangtze rivers.[208] In 2005, the Yellow River was polluted with "4.35 billion [tons] of contaminated effluents, 88 million more than in 2004."[209] Much of the

---

200. Associated Press, *China Crucial in Climate Talks*, USA TODAY, June 30, 2007, http://www.usatoday.com/weather/climate/2007-06-30-china-climate_N.htm.

201. TERRILL, *supra* note 13, at 309.

202. STARR, *supra* note 7, at 178. On average, Beijing's air quality rates sixteen times worse than that of New York City. *Id.*

203. Srini Sitaraman, *Regulating the Belching Dragon: Rule of Law, Politics of Enforcement, and Pollution Prevention in Post-Mao Industrial China*, 18 COLO. J. INT'L ENVTL. L. & POL'Y 267, 271 (2007). Acid rain buffets the Chinese countryside and damages buildings in Japan. *Id.* at 270–71. In fact, "of the 555 cities monitored in 2002, 279 cities (50.3%) have registered occurrences of acid rain." Wang Canfa, *Chinese Environmental Law Enforcement: Current Deficiencies and Suggested Reforms*, 8 VT. J. ENVTL. L. 159, 165 (2007).

204. Robert Lee Hotz, *Huge Dust Plumes From China Cause Changes in Climate*, WALL ST. J., July 20, 2007, at B1.

205. David Barboza, *China Reportedly Urged Omitting Pollution-Death Estimates*, N.Y. TIMES, July 5, 2007, at A3.

206. Reuters, *Yangtze River "Cancerous" With Pollution*, CHINA DAILY, May 30, 2006, http://www.chinadaily.com.cn/china/2006-05/30/content_604228.htm.

207. *Fish Dying in Yellow River*, ASIANEWS, Jan. 18, 2007, http://www.asianews.it/index.php?l=en&art=8264&dos=103&size=A [hereinafter *Fish Dying in Yellow River*]. In October 2006, a half-mile stretch of the Yellow River ran red with chemical discharge. Audra Ang, *Chinese River Mysteriously Turns Red*, WASHINGTON POST, Oct. 24, 2006, http://www.washingtonpost.com/wp-dyn/content/article/2006/10/24/AR2006102400987.html.

208. *Fish Dying in Yellow River*, *supra* note 207.

209. *Id.*

country's surface water is unfit for industrial use, let alone drinking—as many as 320 million of China's 1.3 billion citizens lack access to potable water.[210] Water scarcity is reaching pandemic levels, as 440 of China's 669 major cities face "moderate to severe" shortages.[211] Despite official acknowledgement of widespread degradation and constitutional provisions requiring the government to protect China's environment, [212] the effective regulation of pollution has proven nearly impossible.

A. China's Environmental Laws

Much like China's enactment of intellectual property laws, the adoption of environmental regulations began after Chairman Mao's death and was achieved primarily by importing comprehensive Western standards.[213] Under Mao's rule, the Chinese government neglected environmental issues, since the official ideology "promoted the view that environmental problems [were] particular to capitalistic economic systems because capitalism is based on the logic of profiteering, exploitation, and inequality."[214] Currently, China's environmental laws include legislation addressing environmental protection, natural resource protection and conservation, and special litigation for pollution prevention;[215] the country has also ratified many international conventions and treaties.[216]

The central government began to warm to the idea of environmental protection in 1971, when top leaders attended the United Nations Conference on Environment and Development in Stockholm, Sweden.[217] During the Conference, Chinese diplomats were reluctant to admit their environmental problems and instead blamed developed countries for global pollution.[218] Eventually, China began to address these concerns by holding its own National Conference on Environmental Protection and entrusting the creation of new policies to the State Council and Environmental Protection Leadership Group.[219] These initial steps were largely symbolic; meaningful measures were

---

210. Susan Jakes, *China's Water Woes*, TIME, Oct. 2, 2006, http://www.time.com/time/asia/2006/environment/water.html.

211. Martin Lagod, *We're Running Out of Water*, S.F. CHRON., July 8, 2007, at E5.

212. XIAN FA art. 26 (1982) (P.R.C.), *available at* http://english.people.com.cn/constitution/constitution.html. "The state protects and improves the living environment and the ecological environment, and prevents and controls pollution and other public hazards." *Id.*

213. *See* Sitaraman, *supra* note 203, at 281–82, 298.

214. *Id.* at 287. "Moaists argued that communist societies, based on the principles of equality, fairness, and distributive justice, are inherently friendly towards the natural environment; hence, they believed that China was unlikely to encounter environmental difficulties." *Id.*

215. *See* Wang Canfa, *supra* note 203, at 162–63 (discussing civil and criminal litigation based on environmental damage).

216. *Id.* at 163.

217. Sitaraman, *supra* note 203, at 287–88.

218. *Id.* at 288.

219. *Id.* at 289.

not seriously pursued until the rise of Deng Xiaoping in 1978.[220] Following formal adoption of the Environmental Protection Law (which served as the foundation for coordination of economic development, social progress, and environmental protection) in 1979, China began to enter into major multilateral agreements.[221]

Between 1979 and 2004, the State Council "generated fifty-nine major environmental laws, more than thirty of which specifically focus[ed] on issues such as wildlife preservation, conservation of endangered species, air and water quality, and the prevention of soil erosion, desertification, [and] atmospheric and marine pollution."[222] Yet, the sheer number of rules and overlap of enforcement responsibilities have led Professor Wang Canfa to conclude that many of these laws are both frivolous and unrealistic.[223] Indeed, many of these laws may fairly be characterized as legislation for legislation's sake; the rapid pace of development and deadlines for implementation have caused officials to focus more on creating coalitions between agencies than carefully crafting practical, enforceable regulations.[224] The end result was a comprehensive substantive environmental regime that lacked effective procedural mechanisms.[225]

Thus, the superficial nature of Chinese environmental legislation in the late twentieth century was applied to a setting dominated by administrative decentralization and economic development. Consequently, enforcement again took a backseat to the policies that were entrenching de facto federalism into the Chinese political system. Like its intellectual property laws, China's environmental initiatives were largely ignored at the provincial and local levels.

B. Lax Enforcement

The same implementation issues that plague the Chinese intellectual property rights regime pervade local administration of central government environmental policies. Much of the problem flows from the inefficacious organizational structure of the agencies charged with executing environmental reforms. Although the State Environmental Protection Agency (SEPA) is officially responsible for promulgating laws, conducting impact assessments, and generating specific regulations, the local Environmental Protection Boards (EPBs) are charged with enforcement.[226] In reality, the "EPBs are only

220. *Id.*

221. *Id.* at 293–95.

222. *Id.* at 295. In addition, "China has joined 48 international conventions on environmental protection to date, including the United Nations Framework Convention on Climate Change" and many others. Wang Canfa, *supra* note 203, at 163.

223. Wang Canfa, *supra* note 203, at 169.

224. *Id.* at 170.

225. *Id.*

226. Sitaraman, *supra* note 203, at 309.

nominally under the supervision of SEPA."[227] Unlike its American counterpart, which boasts 17,000 full-time employees, SEPA staffs a scant 300 personnel.[228] Furthermore, the EPBs are funded by the same local governments that turn a blind eye to counterfeiting.[229] Professor Srini Sitaraman clearly articulates the problem:

> This dual supervision has produced a dilemma of governance because local governments control the budget, personnel decisions, promotions, and the allocation of resources such as automobiles, housing, and office space. Local EPB officials are heavily dependent on provincial, city, and county governments because career advancement and budgetary decisions are made by the local governments.[230]

Thus, the advancement of local interests—cultivated by decentralization and de facto federalism—depends entirely on the tax revenue generated by local polluters. Any local government that chooses to elevate environmental consciousness over economic growth suffers doubly, as "construction is prevented or delayed on factories that might bring it income while poisoning the local air and water, and the local government not only loses the tax revenue the factories would have generated but also must pay for the EPB officials who stand in the way."[231]

In familiar fashion, the vast majority of violators actually prosecuted for pollution activity receive administrative penalties.[232] Complying with central government mandates makes little economic sense from the local enterprise's perspective, as nearly 80% of fines levied against violators are returned to them in the form of pollution abatement subsidies.[233] Despite the addition in 1997 of criminal sentences as sanctions for environmental protection violations, fewer than twenty cases out of 387—less than 5%—have been prosecuted.[234]

China's abysmal record for environmental protection has led to an even more disturbing trend: the importation and smuggling of toxic waste.[235] Once

---

227. *Id.*

228. *Id.* at 334. This problem is exacerbated by the fact that, despite its name, SEPA is not considered to be a full-fledged State ministry. *See id.* at 307–09, 334. Bureaucratic prestige is an indispensable tool in Communist China, where the influence associated with the Party is highly coveted.

229. *Id.* at 309–10.

230. *Id.* Accordingly, national policies encounter local counter-measures designed to exploit loopholes in national legislation in furtherance of local goals. *Id.*

231. STARR, *supra* note 7, at 188. This conflict of interests easily explains why the central government regulations permitting large fines for polluters are rarely assessed by local officials, who operate a substantial portion of the very factories targeted by environmental laws. *See* BECKER, *supra* note 6, at 85.

232. Wang Canfa, *supra* note 203, at 168.

233. Sitaraman, *supra* note 203, at 312–13.

234. Wang Canfa, *supra* note 203, at 168.

235. *See generally* Sitaraman, *supra* note 203, at 327–32 (discussing China's involvement

again, the thriving Guangdong Province provides a somber example of de facto federalism's steady erosion of Beijing's control. As a peripheral province with a long history of commercial innovation, Guangdong was an ideal testing ground for Deng Xiaoping's experimental reforms.[236] Because of its vicinity to Hong Kong, the province had been exposed to Western influence; further, its coastal location caused it to become an important trading center.[237] The central government's influence in Guangdong was often resisted by natives, whose local identity was "reinforced by the province's unique culture, manifested in its physiography, climate, language, folklore and products."[238]

This localism ultimately led Sun Yat-sen and his followers to overthrow the Qing dynasty (1644–1912 A.D.) and establish the Republic of China with its capital in Guangzhou.[239] Together, Sun Yat-sen and his successor Chiang Kai-shek led a crusade against powerful warlords and succeeded in reuniting China in 1928.[240] After the Chinese Communist Party took power in 1949, Guangdong's political influence subsided.[241] However, following Deng Xiaoping's rise to power, Guangdong once again found itself at the center of Chinese politics, as liberal policies were initiated with the hope of transforming Guangzhou into a "second 'Hong Kong.'"[242] Out of these reforms eventually evolved de facto Chinese federalism; as such, of all the provinces, Guangdong remains the most resistant to Beijing.[243]

In the case of toxic trade, the Guangzhou town Giuyu "has emerged as ground zero for the global dumping of hazardous wastes."[244] It is estimated that Britain alone exports as much as 200,000 tons of rubbish and 500,000 tons of paper and cardboard to Chinese cities like Giuyu;[245] furthermore, America sends 50–80% of its hazardous electronic waste to such towns.[246] In this highly unregulated trade, waste smugglers cut costs by employing young children and the elderly at illegal dumpsites, paying them approximately one dollar a day.[247]

---

in "toxic trade" and its compliance with the Basel Convention).

236. *See generally* ZHENG YONGNIAN, *supra* note 9, at 232–33 (discussing Guangdong's unique geographic and historical attributes).

237. *Id.* at 233. "As early as the Han dynasty [202 B.C.-220 A.D.], trade routes to as far as Malacca were opened, and trade became an important component of local economic activities." *Id.* Eventually, Guangdong became the most commercialized province on the mainland. *Id.* at 234.

238. *Id.*

239. *Id.* at 236.

240. BRIAN CROZIER, THE MAN WHO LOST CHINA 121 (1976); *see also* ZHENG YONGNIAN, *supra* note 9, at 236.

241. ZHENG YONGNIAN, *supra* note 9, at 236–37.

242. *Id.* at 239.

243. *See id.* at 232 (noting that the central government "initiated a campaign" specifically intended to curb localism in that province).

244. Sitaraman, *supra* note 203, at 327.

245. *Id.* at 328.

246. *Id.* at 327–28.

247. *Id.* at 329.

As a result of this appalling activity, water supplies have become contaminated to such a degree that water for consumption is now "transported in plastic containers from towns more than thirty miles away."[248]

Chinese decentralization has resulted in an administrative climate in which conflicting interests, needless complexity, and widespread inefficiencies abound. With China now leading the world in carbon emissions, any serious global reduction initiative must confront and co-opt local Chinese administrators who are more concerned with economic development than any meaningful environmental protection.

CONCLUSION

Nearly three decades after Deng Xiaoping's revolutionary reforms, the consequences of economic decentralization still reverberate across China. A roaring economy has catapulted China to the forefront of the international intellectual property and environmental debate, compelling Beijing's policymakers to accede to the major multilateral intellectual property and environmental agreements. Sophisticated Western legal frameworks have been adopted, numerous committees have been created to enforce the multitude of new laws, and government media campaigns have repeatedly highlighted the enforcement measures in action. Yet, the supply of counterfeit goods and pirated optical discs never seems to diminish, and recalls of Chinese products continue. The transfer of administrative power from the central to provincial governments has had a trickle-down effect, empowering local leaders to shirk their critical enforcement duties in favor of enhancing their own political reputations through economic growth. These economic warlords shrewdly manage their cities and provinces with a desire to maximize commercial output and further their own legacies, ignoring Beijing's attempts to slow the economy or enforce international agreements. These local leaders are both a product of Deng Xiaoping's vision of a modern China and a reflection of ancient Chinese localism.

De facto federalism has institutionalized the free-market in China, intensifying local rivalries in the process. Beijing's 2008 Olympics offer hope that the central government will renew its efforts to enforce its obligations to other countries. Ultimately, local administrators must realize that, for the long-term good of the country, China must respect the intellectual property and environmental rights of foreigners and natives alike. In the coming decades, China may find itself the leading nation in the areas of semi-conductors, automobiles, and biotechnology. If this happens, it will be in the interest of all Chinese governments—central, provincial, and local—to protect the proprietary rights of Chinese industry. Unless local managers begin to effectively enforce international accords, China will suffer from a decrease in investment. China

248. *Id.* at 328.

also risks facing a legal environment in which inventors and entrepreneurs are reluctant to introduce their ideas into the world's largest market.

The fiscal and administrative decentralization that created de facto federalism in China continues to spur economic growth while undermining international legal agreements. Whether these enforcement issues will be resolved before China steps into the global spotlight in 2008 remains unclear; however, if central planners cannot exert more control over their far-flung provinces, charismatic local leaders will certainly continue to flout their governmental superiors and serve the economic interests cultivated by de facto federalism. Without comprehensive reform, Chinese provinces will become increasingly distinct as local identities and diverging economic interests slowly drive regions apart. If followed to its logical conclusion, de facto federalism may one day cause the People's Republic of China to resemble a loose confederation of quasi-independent states—much like the European Union—where local cultures and language barriers prevent a truly unified system of government. In fact, the recent turmoil in Tibet belies the central government's fragile grip on western China and reveals the simmering resentment which many ethnic minorities harbor toward Beijing.[249] Although the Communist Party would undoubtedly deny any such balkanization, the rest of the world should elevate economic reality over hollow national legislation when shaping its China policy. While far from certain, this scenario is strikingly similar to China's long dynastic history. In the coming years, taming these economic warlords may prove to be the "emperor's" greatest challenge of all.

249. *See* Loretta Chao & James T. Areddy, *Beijing Controls Tibet Visit but Not Ethnic Tension*, WALL ST. J., March 31, 2008, at A7.

PROPERTY—ADMINISTRATION OF WILLS—COMMON LAW ADEMPTION BY EXTINCTION AND THE APPLICABILITY OF TENNESSEE CODE ANNOTATED SECTION 32-3-111

Stewart v. Sewell, 215 S.W.3d 815 (Tenn. 2007).

I. INTRODUCTION

In 1994, the testator Clara Stewart executed her last will and testament.[1] She made a specific bequest of real property that included her house and approximately seven acres to her stepson, the plaintiff.[2] She named her two natural children as the remainder beneficiaries under her will.[3] In the same year, she also executed a durable power of attorney in which she named her two natural children as her attorneys-in-fact.[4] In 1997, the testator's natural children, acting under their power of attorney, obtained appraisals of two separate portions of the real property—an undeveloped six-acre tract and a tract of approximately one acre that included the house.[5] After determining the appraised value of the undeveloped six acres at approximately $110,000, the children sold that portion for $80,000 to members of their family and close friends.[6] Approximately one year later, the testator died and her natural children inherited the remaining proceeds from the sale of the property as the remainder beneficiaries under the will.[7] The plaintiff inherited the one-acre tract that had not been sold.[8]

The plaintiff sued the testator's children and the purchasers of the six-acre tract, alleging that the children had "fraudulently conveyed their mother's

---

1. Stewart v. Sewell, 215 S.W.3d 815, 818 (Tenn. 2007).

2. *Id.* The real property, known as the "Tim's Ford Lake property," was inherited by the testator from her late husband upon his death in 1981. *Id.*

3. *Id.*

4. *Id.* The durable power of attorney provided that the rights of the attorneys-in-fact included the power "to buy and sell both real and personal property on [the testator's] behalf to the full extent as if [she] transacted the sale or purchase in person. This shall specifically include the right and power to execute deeds and other instruments conveying personal and real property." *Id.*

5. *Id.* The children testified that they considered selling the six-acre tract to assist in funding the testator's nursing home care. *Id.* at 819. After being found in a coma in December 1996, the testator was placed in a nursing home in January 1997. *Id.* at 818.

6. *Id.* at 819. Before the sale, the children contacted the plaintiff to determine if he wanted to buy the property, but he declined because he believed he was entitled to receive the land through bequest. *Id.*

7. *Id.*

8. *Id.* at 817.

property to keep [the plaintiff] from inheriting said property."[9] The trial court dismissed the plaintiff's complaint without making specific findings of fact.[10] On appeal, the Tennessee Court of Appeals reversed the trial court and awarded the plaintiff judgment against the testator's children.[11] The court found that the testator's children breached their duties as attorneys-in-fact and that the rule of ademption by extinction did not apply because of the passage of section 32-3-111 of the Tennessee Code Annotated.[12] The court of appeals, however, affirmed the trial court's dismissal of the claims against the purchasers of the property.[13] Upon review by the Tennessee Supreme Court, *held*, reversed.[14] The attorneys-in-fact did not breach their fiduciary duties, and the specific bequest of real property was adeemed by extinction when the property no longer existed in the estate at the time of the testator's death.[15] Tennessee Code Annotated section 32-3-111 did not apply retroactively.[16] *Stewart v. Sewell*, 215 S.W.3d 815 (Tenn. 2007).

II. ISSUES OF LAW

Ademption by extinction occurs when specifically bequeathed property no longer exists in the estate at the time of the testator's death because the property has been given away, sold, destroyed, or has been so materially altered that it cannot be substituted for the specific bequest.[17] Tennessee has traditionally followed the *in specie* theory (or identity theory) of ademption by extinction, whereby courts only evaluate whether the bequest is specific and if so, whether the bequeathed property exists in the estate at the time of the testator's death.[18] In 2004 the Tennessee General Assembly enacted Tennessee Code Annotated section 32-3-111, which provides several exceptions to the common law *in specie* theory of ademption by extinction. One such exception applies when a conservator or agent acting within the authority of a durable power of attorney

---

9. *Id.* at 820.

10. *Id.* at 821.

11. *Id.*

12. *Id.* The court of appeals found that because the attorneys-in-fact had "acted in contravention" of their fiduciary duties, the plaintiff was entitled to enforce a constructive trust on the net proceeds of the sale, with the plaintiff named as the beneficiary. Stewart v. Sewell, No. M2003-01031-COA-R3-CV, 2005 Tenn. App. LEXIS 222, at *59 (Tenn. Ct. App. Apr. 14, 2005).

13. *Stewart*, 2005 Tenn. App. LEXIS 222, at *63–64.

14. *Stewart*, 215 S.W.3d at 828.

15. *Id.* at 825.

16. *Id.* at 828.

17. *See* 80 AM. JUR. 2D *Wills* § 1467 (2002); *see also* JACK W. ROBINSON, SR. & JEFF MOBLEY, PRITCHARD ON THE LAW OF WILLS AND ADMINISTRATION OF ESTATES § 486, at 709–10 (5th ed. 1994 & Supp. 2006) ("The term 'ademption' . . . . also is used with reference to the loss of the legacy, destruction of the subject matter, or transfer or termination of testator's interest before his death. This is sometimes called 'ademption by extinction.'").

18. *In re* Estate of Hume, 984 S.W.2d 602, 605 (Tenn. 1999).

for an incompetent testator sells or otherwise disposes of property that is the subject of a specific bequest.[19] In *Stewart v. Sewell*, the Tennessee Supreme Court addressed the applicability of the *in specie* theory of ademption by extinction and whether Tennessee Code Annotated section 32-3-111 would apply retroactively.[20]

III. DEVELOPMENT OF THE LAW OF ADEMPTION BY EXTINCTION

A. Tennessee Common Law Principles of Ademption by Extinction

Ademption is "the extinction, alienation, withdrawal, or satisfaction of the legacy by some act of the testator by which an intention to revoke is indicated; the doing of some act with regard to the subject-matter which interferes with the operation of the will."[21] The two primary forms of ademption are ademption by satisfaction and ademption by extinction.[22] Ademption by satisfaction occurs during the lifetime of the testator when the testator provides the bequeathed property, or a valid substitute, to the beneficiary so as to show that the bequest is revoked or satisfied.[23] In contrast, ademption by extinction is determined at the time of the testator's death when the specific bequest cannot be fulfilled because the bequeathed property no longer exists in the estate.[24] The Tennessee Supreme Court has adopted the *in specie* theory of ademption by extinction.[25] The *in specie* theory of ademption by extinction does not look to the testator's intent.[26] Rather, the *in specie* theory focuses on whether the bequest is specific, and if it is, whether the bequeathed property is still held by the estate.[27]

In *American Trust & Banking Co. v. Balfour*, a foundational Tennessee case involving the law of ademption by extinction, the Tennessee Supreme Court ruled that the bequest of life insurance proceeds was a specific bequest and that the termination of the life insurance policies adeemed the bequest.[28] The testator executed his will and bequeathed his life insurance proceeds to his daughter to pay for her education, provided that the testator died before his

---

19. TENN. CODE ANN. § 32-3-111(b) (2004).

20. *Stewart*, 215 S.W.3d at 817.

21. Am. Trust & Banking Co. v. Balfour, 198 S.W. 70, 71 (Tenn. 1917); *see also* 80 AM. JUR. 2D *Wills* § 1458 (2002) ("'Ademption' is generally defined as the extinction, alienation, withdrawal, or satisfaction of the legacy by some act of the testator by which an intention to revoke is indicated.").

22. ROBINSON & MOBLEY, *supra* note 17, at 709–710.

23. *Hume*, 984 S.W.2d at 604.

24. 97 C.J.S. *Wills* § 1749 (2001).

25. *Hume*, 984 S.W.2d at 605.

26. *Id.*

27. *Id.*

28. Am. Trust & Banking Co. v. Balfour, 198 S.W. 70, 71 (Tenn. 1917).

daughter finished her education.[29] Approximately two years after executing his will, the testator terminated his life insurance policies, collected the cash surrender values, and eventually invested the proceeds into real estate mortgage notes.[30] The testator still owned the mortgage notes at the time of his death.[31] The executor of the will filed a complaint for a construction of the will to determine how to treat the bequest of the life insurance proceeds.[32] The Tennessee Supreme Court held that the bequest of life insurance proceeds was specific because life insurance policies "may not be acquired by purchase at will in open market by the testator or his personal representative to respond to and satisfy the bequest."[33] After finding that the bequest was specific in nature, the court then concluded that the bequest was adeemed when it was converted into mortgage notes. The court reasoned that "a specific legacy is adeemed when there has been a material alteration or change in the subject-matter, and that the property into which it was converted in such change cannot be substituted as or for the specific bequest."[34]

One year later, the Tennessee Supreme Court restated the law of ademption by extinction when it decided *Ford v. Cottrell*.[35] In *Ford*, the testator executed a will in which she bequeathed the rents from a house she had owned to her sister for the remainder of the sister's life.[36] The will also provided that upon the death of the testator's sister, the house was to be sold and the proceeds from the sale were to be given to a specific orphans' home.[37] The house mentioned in the will was sold by the testator prior to her death, but the sales proceeds were identifiable through monthly note payments, paid by the purchasers.[38] The Tennessee Supreme Court first determined that the bequests of rents to the sister and sale proceeds to the orphans' home were specific bequests.[39] After deciding that the bequests were specific, the Tennessee Supreme Court held that the specific bequests were adeemed because "the sale of [the] property by the testator during her lifetime . . . was an ademption to the legacies provided for in . . . the will . . . ."[40]

Further establishing the law of ademption by extinction in Tennessee, *Wiggins v. Cheatham* demonstrated the importance of the court's determination

---

29. *Id.* at 70. If the testator died after his daughter finished her education, the life insurance proceeds were to be placed in a trust fund for her benefit. *Id.* The remainder of the estate was to be divided, half to the daughter in trust and half to the testator's wife. *Id.*

30. *Id.* at 70–71.

31. *Id.* at 71.

32. *Id.* at 70.

33. *Id.* at 71.

34. *Id.*

35. 207 S.W. 734, 736 (Tenn. 1918).

36. *Id.* at 735.

37. *Id.*

38. *Id.*

39. *Id.* at 737.

40. *Id.*

of whether the property that is subject to a specific bequest still exists in the estate.[41] In *Wiggins*, the testator's will stated, "I give my entire whisky business now conducted at 1221–1223 Market [S]treet, Chattanooga, Tenn., in equal parts to [legatees]."[42] Less than a year after the will was executed, the testator was forced to close his whisky business in Chattanooga and move the whisky to Kentucky because of adverse legislation.[43] The testator died approximately one year after moving the whisky.[44] The executor sold all of the whisky in storage, and the funds were held subject to construction of the will.[45] Upon the eventual review by the Tennessee Supreme Court, the court stated:

> We do not think that the fact that the testator was not actually engaged in business at the time of his death can be said to destroy the specific character of the property intended to be bequeathed in his will. That part of the specific property bequeathed, and which remained unsold, was in existence at the time of the testator's death, and was subject to identification.[46]

Because the court ruled that the subject of the specific bequest, the whisky business, was still identifiable upon the death of the testator, the bequest was not adeemed; thus, the legatees were entitled to the proceeds of the sale of the whisky.[47]

In a more modern ademption by extinction case, *In re Estate of Hume*, the Tennessee Supreme Court reinforced the applicability of the common law ademption by extinction doctrine when it ruled that the identifiable proceeds from a foreclosure sale of a house could not be substituted for the specific bequest of the house.[48] In that case, the testator executed a will in which he bequeathed a house he owned in Atlanta to the plaintiff, with the remainder of his estate going to a college.[49] Shortly before his death, the testator stopped

---

41. *See* Wiggins v. Cheatham, 225 S.W. 1040, 1042 (Tenn. 1920).

42. *Id.* at 1040.

43. *Id.* While the testator kept the whisky in warehouses in Louisville, Kentucky, he took out a loan from a Chattanooga bank to make significant tax payments related to the whisky, and he sold a portion of the whisky to assist in repaying one of his loans from the bank. *Id.* The testator also commented to several friends that he was seeking a new location to re-open the whisky business. *Id.*

44. *Id.*

45. *Id.* at 1040–41. The legatees of the whisky business filed a complaint to construe the testator's will to determine if the proceeds from the sale of the whisky should be devised to the specific legatees or to the general estate. *Id.* at 1041. The defendants argued that the removal of the whisky from Chattanooga and the halting of the whisky business adeemed the bequest to the plaintiffs because the property was so altered or changed that the property could not be identified. *Id.*

46. *Id.* at 1042.

47. *Id.* at 1041–42.

48. *In re* Estate of Hume, 984 S.W.2d 602, 606 (Tenn. 1999).

49. *Id.* at 603.

making mortgage payments, and the bank sold the house at foreclosure.[50] The trial court and the Tennessee Court of Appeals both ruled that the sales proceeds should be paid to the plaintiff because there was no evidence that the testator had actual knowledge of the foreclosure proceedings and because the testator clearly intended that the house should be given to the plaintiff.[51] Referencing its earlier holdings in *American Trust & Banking Co. v. Balfour*, *Ford v. Cottrell*, and *Wiggins v. Cheatham*, the Tennessee Supreme Court overturned the lower courts, stating that "it only matters that the subject of the specific bequest no longer exists because of 'the doing of some act;' it is irrelevant who or what initiates 'the doing.'"[52] Quoting a decision from another jurisdiction, the Tennessee Supreme Court reaffirmed the common law doctrine of ademption by extinction by putting forth the *in specie* test, which focuses on only two questions: "'(1) whether the gift is a specific legacy and, if it is, (2) whether it is found in the estate at the time of the testator's death.'"[53] Because the bequest was specific and the property no longer was owned by the estate, the court held that the bequest to the plaintiff was adeemed and she therefore did not receive the foreclosure sale proceeds.[54]

B. Uniform Probate Code Section 2-606, Tennessee Code Annotated Section 32-3-111, and Their Effect on the Common Law Principles of Ademption by Extinction

Because of the sometimes harsh results of the common law *in specie* theory of ademption by extinction, several jurisdictions have modified the doctrine either through court decisions or by statute.[55] Prior to the Tennessee Supreme Court's consideration of *Stewart v. Sewell*, approximately two-thirds of the case holdings outside of Tennessee made an exception to the *in specie* theory for situations in which testators became incompetent after making specific bequests.[56] These cases generally have ruled that no ademption occurs when

---

50. *Id.* The testator died soon after the foreclosure sale and the proceeds from the sale were paid to the estate's executrix, who understood the bequest of the house to be adeemed and believed the proceeds should be paid to the residual beneficiary. *Id.* The plaintiff filed an exception to the executrix's final accounting of the estate in the probate court, claiming a right to the surplus proceeds from the foreclosure sale. *Id.*

51. *Id.* at 603–04.

52. *Id.* at 604 (quoting Am. Trust & Banking Co. v. Balfour, 198 S.W. 70, 71 (Tenn. 1917)).

53. *Id.* at 605 (quoting McGee v. McGee, 413 A.2d 72 (R.I. 1980)).

54. *Id.*

55. *See* UNIF. PROBATE CODE § 2-606 cmt. (amended 2006), 8 U.L.A. 177 (Supp. 2007) ("Recently some courts have begun to break away from the 'identity' theory and adopt, instead the so-called 'intent' theory.").

56. *See* Jeffrey F. Ghent, Annotation, *Ademption or Revocation of Specific Devise or Bequest by Guardian, Committee, Conservator, or Trustee of Mentally or Physically Incompetent Testator*, 84 A.L.R. 4th 462, 467–68 (1991). The other third of the holdings were closely divided between partial ademption and total ademption. *Id.* at 468.

the guardian or conservator of an incompetent testator sells the property that is the subject matter of the testator's specific bequest and retains the proceeds of the sale, in whole or in part, at the time of the testator's death.[57] The comments to Uniform Probate Code section 2-606 encourage states to adopt these exceptions, stating that "[t]he application of the 'identity' theory of ademption has resulted in harsh results in a number of cases, where it was reasonable [sic] clear that the testator did not intend to revoke the devise."[58]

The Tennessee General Assembly adopted legislation, which took effect on June 8, 2004, that largely mirrors Uniform Probate Code section 2-606.[59] Tennessee Code Annotated section 32-3-111 follows the modern trend by modifying the strict *in specie* theory with several exceptions.[60] The statute provides:

> A specific legatee or devisee has a right to the specifically gifted or devised property in the testator's estate at death or if the property has been disposed of and a contrary intention is not manifest during the testator's lifetime: (1) Any balance of the purchase price, together with any security interest, owing from a purchaser to the testator at death by reason of sale of the property; (2) Any amount of a condemnation award for the taking of the property unpaid at death; (3) Any proceeds unpaid at death on fire or casualty insurance on, or other recovery for injury to, the property; and (4) Property owned by the testator at death and acquired as a result of foreclosure, or obtained in lieu of foreclosure, of the security interest for a specifically devised obligation.[61]

The legislature also adopted Uniform Probate Code section 2-606(b), which establishes an exception for an ademption that occurs as a result of the acts of a conservator or guardian of an incompetent testator.[62] Addressing the potential challenges related to a conservator managing the estate of an incompetent testator, the statute provides:

> If specifically devised or bequeathed property is sold or mortgaged by a conservator or by an agent acting within the authority of a durable power of attorney for an incapacitated principal, or if a condemnation award, insurance proceeds, or recovery for injury to the property are paid to a conservator or to an agent acting with the authority of a durable power of attorney for an incapacitated principal, the specific devisee has the right to a general

---

57. *Id.* at 468. If the guardian uses the proceeds of the sale for the testator's maintenance and support, however, no ademption occurs. *Id.*

58. UNIF. PROBATE CODE § 2-606 cmt. (amended 2006), 8 U.L.A. 177 (Supp. 2007) (listing as notable examples *McGee v. McGee*, 413 A.2d 72 (R.I. 1980), and *Estate of Dungan*, 73 A.2d 776 (Del. Ch. 1950)).

59. *See* Stewart v. Sewell, 215 S.W.3d 815, 825–26 (Tenn. 2007).

60. *See* TENN. CODE ANN. § 32-3-111 (2004).

61. *Id.* § 32-3-111(a).

62. *Id.* § 32-3-111(b); UNIF. PROBATE CODE § 2-606(b) (amended 2006), 8 U.L.A. 177 (Supp. 2007).

> pecuniary devise equal to the net sale price, the amount of the unpaid loan, the condemnation award, the insurance proceeds, or the recovery.[63]

Essentially, the Tennessee General Assembly declared that upon the enactment of section 32-3-111(b), the *in specie* theory of ademption by extinction does not apply when a conservator or an agent, acting under a durable power of attorney for an incompetent person, sells or otherwise disposes of the property that is subject of a specific bequest.[64]

IV. STATUTE DATE OF ENACTMENT AND THE EFFECT ON THE COMMON LAW

When newly enacted statutes change the law related to probate matters, the courts must determine when the statutes become effective in order to apply the appropriate law to the construction of a will.[65] The Tennessee Supreme Court has stated, "It may generally be said that all substantive rights in the estate, be they vested or inchoate, are controlled by the law existing at the time of death."[66] In *Marler v. Claunch*, the testator died seven days before the Tennessee General Assembly enacted legislation that changed a statute which governed the provision of support allowances to the decedent's family in the administration of wills.[67] The testator's widow dissented to the final accounting of the will, arguing that the amended statute applied and that she was entitled to receive more from the estate than if the statute did not apply.[68] The Tennessee Supreme Court concluded that the time of death establishes the substantive rights of the parties under the will but that newly enacted statutes could change the procedural process controlling how wills are administered.[69] The court provided, "It is at the time of death that the entire substance of the right is fixed, though the procedural mechanics for its realization are subject to change by subsequent legislative directive."[70]

In *Fell v. Rambo*,[71] the testator executed his will in 1959, devising his entire estate to his wife in a life tenancy with unlimited powers of disposition.[72] The will left a remainder interest to the couple's twelve nieces and nephews.[73]

---

63. TENN. CODE ANN. § 32-3-111(b).

64. *See id.*

65. *See* Marler v. Claunch, 430 S.W.2d 452, 453 (Tenn. 1968).

66. *Id.* at 454.

67. *Id.* at 453. The testator died on April 23, 1967; his will was probated on May 1, 1967; the testator's widow filed her dissent from the will on May 4, 1967. *Id.* Also on May 1, 1967, the Tennessee General Assembly enacted legislation that changed the statute governing support allowances to family members. *Id.*

68. *See id.* at 454.

69. *Id.*

70. *Id.*

71. 36 S.W.3d 837 (Tenn. Ct. App. 2000).

72. *Id.* at 840.

73. *Id.*

The testator died in 1963, and his widow continued living on their farm.[74] In 1986, the widow executed a general power-of-attorney to her niece, the defendant.[75] In 1991, the widow executed her own will, making a few bequests and dividing the remainder of her estate between the defendant and one other nephew.[76] With the defendant as her attorney-in-fact, the widow sold the farm in 1993; she died in 1994.[77] Upon the widow's death, some of her nieces and nephews who were excluded from her will asserted that they were entitled to the farm sale proceeds based upon the deceased husband's earlier will.[78] The central issue in the litigation that followed was whether the pre- or post-1981 version of Tennessee Code Annotated section 66-1-106 applied to the deceased husband's 1959 will.[79] Following the Tennessee Supreme Court's holding in *Marler v. Claunch*, the court of appeals ruled that the law in effect at the time of the testator's death governed the substantive rights of the parties.[80] Because the plaintiffs' remainder interest was based on the deceased husband's 1959 will and his 1963 death, the pre-1981 version of the statute applied and the plaintiffs did not have a right to the proceeds of the sale.[81]

In *Nutt v. Champion International Corp.*, a case involving workers' compensation, the Tennessee Supreme Court addressed the retroactivity of statutes by stating that "[s]tatutes are presumed to operate prospectively unless the legislature clearly indicates otherwise."[82] The *Nutt* case involved a 1996 amendment to Tennessee Code Annotated section 50-6-114(b) which allowed employers to offset long-term disability payments against workers' compensation awards for permanent total disability.[83] The statute became effective after the date upon which the plaintiff suffered his injury.[84] Referencing earlier judicial decisions and the Tennessee Constitution's prohibition on retrospective laws, the court reiterated the general rule that statutes affecting legal rights cannot apply retroactively.[85] However, the court recognized an exception for procedural or remedial statutes: "Statutes deemed

---

74. *Id.* The widow was forced to leave the farm in 1991 due to poor health. *Id.*

75. *Id.*

76. *Id.* Except for the defendant and one other nephew, the widow excluded all other nieces and nephews who were included in her deceased husband's earlier will. *Id.*

77. *Id.* at 841.

78. *Id.*

79. *Id.* at 844. If the pre-1981 version of section 66-1-106 applied, the plaintiffs would lose their remainder interest and rights to the sale proceeds when the widow sold the farm. *Id.* On the contrary, if the amended post-1981 version of section 66-1-106 applied, the plaintiffs would have a statutory right to any proceeds that were not used to pay the life tenant's debts incurred during her lifetime. *Id.*

80. *Id.* at 845 (citing Marler v. Claunch, 430 S.W.2d 452, 454 (Tenn. 1968)).

81. *Id.* at 846.

82. Nutt v. Champion Int'l Corp., 980 S.W.2d 365, 368 (Tenn. 1998) (citations omitted).

83. *Id.* at 367.

84. *See id.* at 367.

85. *Id.* at 368. The Tennessee Constitution provides that "no retrospective law, or law impairing the obligations of contracts, shall be made." TENN. CONST. art. I, § 20.

remedial or procedural apply retrospectively to causes of action arising before such acts became law and to suits pending when the legislation took effect."[86] The court explained that a procedural or remedial statute is one that "does not affect the vested rights or liabilities of the parties"[87] and that "addresses the mode or proceeding by which a legal right is enforced."[88] Because the injury pre-dated the statute's effective date and because the legislature had not included language to make application of the statute retroactive, the court held that the employer was not entitled to an offset.[89]

V. *STEWART V. SEWELL*: CLARIFYING THE COMMON LAW OF ADEMPTION BY EXTINCTION AND THE APPLICABILITY OF TENNESSEE CODE ANNOTATED SECTION 32-3-111

In *Stewart v. Sewell*, the Tennessee Supreme Court addressed "the applicability of the rule of ademption by extinction and of Tennessee Code Annotated section 32-3-111 concerning the sale of specifically devised property."[90] Justice Clark wrote the court's unanimous opinion.[91] The Tennessee Supreme Court reinstated the trial court's dismissal of the plaintiff's complaint because the rule of ademption by extinction applied to the specific bequest and the bequest was adeemed when the property was sold.[92] Further, the court stated that the court of appeals "erred in applying retroactively Tennessee Code Annotated section 32-3-111 and in imposing a constructive trust in order to avoid that result."[93]

A. The Rule of Ademption by Extinction

In her discussion of ademption, Justice Clark began by setting forth the Tennessee common law rule of ademption by extinction and affirming its current applicability.[94] Relying upon the court's earlier holdings in *American Trust & Banking Co. v. Balfour* and *In re Estate of Hume*, Justice Clark

86. *Nutt*, 980 S.W.2d at 368.

87. *Id.* (citations omitted); *see also* Kuykendall v. Wheeler, 890 S.W.2d 785, 787 (Tenn. 1994) ("Whether a statute applies retroactively depends on whether its character is 'substantive' or 'procedural.' If 'substantive,' it is not applied retroactively because to do so would 'disturb a vested right or contractual obligation.'" (quoting Saylors v. Riggsbee, 544 S.W.2d 609, 610 (Tenn. 1976))).

88. *Nutt*, 980 S.W.2d at 368.

89. *Id.*

90. Stewart v. Sewell, 215 S.W.3d 815, 817 (Tenn. 2007).

91. *Id.*

92. *Id.*

93. *Id.*; *see also supra* note 12 (explaining the court of appeals's imposition of a constructive trust).

94. *See id.* at 824.

reiterated "Tennessee's longstanding rule that a devise of specific property is extinguished upon 'the doing of some act with regard to the subject-matter [of the devise] which interferes with the operation of the will.'"[95] Specifically, Justice Clark emphasized that in Tennessee, the common law rule of ademption by extinction "prevails *without regard to the intention of the testator or the hardship of the case* . . . ."[96] Justice Clark endorsed the *in specie* theory of ademption by extinction and its premise that intent is irrelevant, citing its advantages of "ease of application, stability, uniformity, and predictability."[97]

After establishing the enduring applicability of the common law rule of ademption by extinction, the unanimous court concluded that the testator's specific bequest of real property was adeemed upon the sale of the six-acre tract.[98] Justice Clark stated that the sale "was clearly 'the doing of some act with regard to the subject-matter which interfere[d] with the operation of the will.'"[99] Though she criticized the court of appeals's findings of fact,[100] Justice Clark elected not to address the lower court's conclusion that an exception to the common law rule of ademption by extinction is required when there is a breach of a fiduciary duty by an attorney-in-fact.[101] Rather, the court found that the attorneys-in-fact in this case acted in accordance with their duties and did not act in an unfaithful or disloyal manner.[102] The court concluded by holding that the plaintiff was not entitled to the proceeds of the sale of the six-acre tract because the specific bequest was adeemed.[103] Further, because the attorneys-in-fact did not violate their fiduciary duties and the property was adeemed, the lower court's finding of a constructive trust on the proceeds from the sale was inappropriate.[104]

---

95. *Id.* at 824 (quoting *In re* Estate of Hume, 984 S.W.2d 602, 604 (Tenn. 1999)).

96. *Id.* (quoting *Hume*, 984 S.W.2d at 604).

97. *Id.* at 825 (citing *Hume*, 984 S.W.2d at 605).

98. *Id.*

99. *Id.* (quoting Am. Trust & Banking v. Balfour, 198 S.W. 70, 71 (Tenn. 1917)).

100. *Id.* at 826–27. Justice Clark emphasized that there was no evidence in the record to contradict the assertions of the attorneys-in-fact that they had placed the proceeds of the sale into a credit union account bearing the testator's name along with their own, so that the testator would have access to the funds if anything happened to them. *Id.* at 827. She also noted that no evidence in the record that the attorneys-in-fact made any improper use of the funds. *Id.*

101. *Id.* at 825; *see supra* note 12 (describing the court of appeals's holding).

102. *Id.* The Supreme Court detailed how the natural children, acting under their power of attorney, deposited the real estate sales proceeds into a credit union account. *Id.* at 820. The natural children then issued checks totaling approximately $40,000 from the credit union account for the testator's nursing home care. *Id.* Checks totaling approximately $2,200 were also issued from the same credit union account for the testator's pharmacy bills. *Id.*

103. *Id.* at 825.

104. *Id.* at 827 ("[N]o improper use of the proceeds occurred. And because the original bequest was partially adeemed by the sale of the Undeveloped Tract, [the plaintiff] had no interest in the proceeds."); *see supra* notes 12, 93 and accompanying text.

B. The Applicability of Tennessee Code Annotated Section 32-3-111

Justice Clark next turned a critical eye to the court of appeals's application of Tennessee Code Annotated section 32-3-111.[105] The court of appeals found that the sale of the property did not adeem the specific bequest under the language of the statute which provides that "[i]f specifically devised or bequeathed property is sold or mortgaged by a conservator or by . . . a durable power of attorney for an incapacitated principal, . . . the specific devisee has the right to a general pecuniary devise equal to the net sale price. . . ."[106]

First citing the Tennessee Constitution's ban on retrospective laws and then referencing both *Nutt v. Champion International Corp.* and *Fell v. Rambo*, Justice Clark reiterated that "the law in effect when the testator dies controls all substantive rights in the estate, whether vested or inchoate."[107] Justice Clark noted that the testator died in 1998 and that Tennessee Code Annotated section 32-3-111 became effective on June 8, 2004.[108] The Tennessee General Assembly did not include any language in the statute concerning retrospective application.[109] Justice Clark emphasized that section 32-3-111 could not be read as simply a remedial or procedural statute because its application would significantly affect the vested rights of the testator's children to the funds remaining in her bank account.[110] In summary, the court concluded that section 32-3-111 was not applicable and thus could not affect the vested rights of either the plaintiff or the defendants.[111] Ultimately, the common law rule of ademption by extinction, without exceptions, prevailed.

C. The Effects of Stewart v. Sewell

The Tennessee Supreme Court rendered a sound decision in *Stewart v. Sewell* regarding the retroactivity and application of statutes by refusing to apply Tennessee Code Annotated section 32-3-111 to the facts of this case. The court's decision provided consistency in the area of probate law by not applying the statute; however, it also made evident the problems with the *in specie* doctrine of ademption by extinction. The facts of *Stewart v. Sewell* demonstrate why the *in specie* doctrine needed to be modified and why section 32-3-111 improves the law of ademption in Tennessee. Further, the case affirms the importance of including detailed instructions in powers of attorney to clarify the intent of the parties regarding specifically bequeathed property.

---

105. *Stewart*, 215 S.W.3d at 825–26.

106. *Id.* (quoting TENN. CODE ANN. § 32-3-111(b) (Supp. 2004)).

107. *Id.* at 826 (quoting Fell v. Rambo, 36 S.W.3d 837, 845 (Tenn. Ct. App. 2000)).

108. *Id.* (citations omitted).

109. *Id.*

110. *Id.*; *see supra* notes 86–88 and accompanying text (giving the Tennessee Supreme Court's definition of a procedural or remedial statute and explaining the rule that such a statute may apply retroactively to the facts of a particular case).

111. *Stewart*, 215 S.W.3d at 826.

Justice Clark's opinion promotes consistency in Tennessee probate law by refusing to apply retroactively a statute that did not expressly include retroactive language. The Tennessee Supreme Court has consistently ruled that the law in effect at the time of the testator's death controls the beneficiaries' substantive rights in the estate.[112] Section 32-3-111 was enacted more than six years after the testator in *Stewart v. Sewell* died[113] and more than seven years after the property that was the subject of the specific bequest was sold.[114] Retroactively applying a statute more than six years after the death of a testator would have caused confusion and apprehension for even the most thoughtful estate planners. If the Supreme Court had chosen to apply the statute to the facts of this case, it would have established a precedent whereby future courts could modify wills, long after the death of the testator, that were drafted under the law when the testator was alive. Instead, the court avoided significant long-term confusion by reaffirming earlier holdings that the law in effect at the time of the testator's death controls the substantive rights of the estate and that statutes affecting vested rights only have retrospective application if the legislature so specifies. By upholding the long-recognized rule, the Tennessee Supreme Court chose to maintain an environment where lawyers can draft wills, and courts can interpret them, with a sound understanding of the law and its applicability.

While the court appropriately denied retroactive application of the statute, *Stewart v. Sewell* provides an example of the harsh results that sometimes occur with the *in specie* doctrine of ademption by extinction. The apparent intent of the testator was not dispositive in this case.[115] The testator made the specific bequest of the property while she was competent and aware of her actions. The property was sold and the bequest was adeemed while the testator was unable to participate in the decision because of incapacity. The plaintiff, who expected to receive the property through the bequest, lost the property when it was sold by the testator's guardians. *Stewart v. Sewell* demonstrates that while the *in specie* doctrine provides for clear decisions and bright lines, it also often leads to defeated intent and failed bequests.

The passage of Tennessee Code Annotated section 32-3-111 improves the law of ademption in Tennessee by superseding the common law *in specie* doctrine of ademption by extinction. Section 32-3-111 finally does what the Tennessee Supreme Court had refused to do through its past decisions: The statute protects the testator's intent and prevents a testator's incapacity from defeating that intent when property is sold by conservators or guardians. Specifically, section 32-3-111(b) preserves the right of the specific devisee to "a general pecuniary devise equal to the net sale price, the amount of the

112. *See* Marler v. Claunch, 430 S.W.2d 452, 454 (Tenn. 1968); *see also supra* Part IV (discussing the general rule in Tennessee).

113. *Stewart*, 215 S.W.3d at 826.

114. *Id.* at 819.

115. *See supra* notes 95–97 and accompanying text (explaining the rule that the testator's intent is irrelevant when applying the doctrine of ademption by extinction).

unpaid loan, the condemnation award, the insurance proceeds, or the recovery."[116] If the testator in *Stewart v. Sewell* had died after the enactment of section 32-3-111(b), the statute would have protected her original intent by allowing the plaintiff to receive the proceeds from the sale of the specifically bequeathed property because it was sold by "an agent acting with the authority of a durable power of attorney for an incapacitated principal."[117] Unfortunately for the testator and the plaintiff, they were caught by the intent-defeating *in specie* doctrine, and the bequest was adeemed as a result.

Another important lesson from *Stewart v. Sewell* is that estate planners should encourage testators either to avoid specific bequests or to include detailed instructions about how they want specifically bequeathed property handled by conservators or guardians in their durable powers of attorney.[118] If the testator in this case had provided specific instructions in her power of attorney to prohibit the sale of the bequeathed property, the ultimate issue never would have arisen. The parties could have agreed how to handle the property before the testator became incapacitated. Because the parties did not address this possibility in the durable power of attorney, they were forced to turn to the courts for a resolution.

VI. CONCLUSION

The *Stewart v. Sewell* holding provides a solid foundation for the application of statutes to probate matters in Tennessee. Instead of looking at a sensitive factual scenario and working to avoid harsh results by incorrectly applying a statute retroactively, the Tennessee Supreme Court chose legal consistency over case-by-case confusion. The court's decision also provides a detailed example of why Tennessee Code Annotated Section 32-3-111 is a welcome change to the common law of Tennessee. By prohibiting ademption when the property is sold by a guardian or conservator acting under a durable power of attorney, the statute will protect the intent of future testators and allow parties to better plan for death and incapacity.

JOSHUA A. MULLEN*

116. TENN. CODE ANN. § 32-3-111(b) (2004).

117. *Id.*

118. *See* Ghent, *supra* note 56, at 471.

* The author wishes to thank Professor Amy Morris Hess for her constructive comments and critique of earlier versions of this note.

EVERYTHING IS PATENTABLE

MICHAEL RISCH*

INTRODUCTION

The currently confused and inconsistent jurisprudence of patentable subject matter[1] can be clarified by implementing a single rule: any invention that satisfies the Patent Act's requirements of category, utility, novelty, nonobviousness, and specification is patentable.[2] In other words, if a discovery otherwise meets the requirements of patentability, then the discovery will be properly patentable without need to consider non-statutory subject matter restrictions such as the bars against mathematical algorithms, products of nature, or natural phenomena.[3]

This Article's proposal is based on both positive and normative analysis. Positively, a historic review of United States Supreme Court opinions provides evidence that general patentability criteria—and not subject matter—were key to the Court's primary subject matter precedents. In each case reviewed, the Court's analysis concerned the underlying patentability of the particular claim at issue—

---

* Michael Risch. Associate Professor of Law and Project Director – Entrepreneurship, Innovation and Law Program, West Virginia University College of Law; Of Counsel, Russo & Hale LLP, Palo Alto, California. The author thanks Miriam Bitton, Kevin Collins, John Duffy, Mark Lemley, David Olson, Caprice Roberts, Joshua Sarnoff, David Schwartz, John Taylor, R. Polk Wagner, participants of the 2007 IP Scholars Conference, the 2007 WIPIP Conference, and faculty workshops at Cardozo Law School and West Virginia University College of Law. Valuable research assistance was provided by Mary Long and Gabriele Wohl.

1. *See generally* 35 U.S.C. §§ 101, 102, 103, 112 (2000). Section 101 lists the subject matter for which patents may be obtained. Patentable subject matter is "any new and useful process, machine, manufacture, or composition of matter...." *Id.* § 101. Sections 102 and 103 require an invention's novelty and nonobviousness. Novelty means that an invention is new, that it has not been patented in the United States or a foreign country, or that no one has applied for a patent for that invention. *Id.* § 102. An invention is obvious if the subject matter is "such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains." *Id.* § 103. Section 112 requires specificity in a patent application. A patent application must contain "a written description of the invention, and of the manner and process of making and using it...as to enable any person skilled in the art...to make and use the same..." *Id.*

2. *See id.* §§ 101–103, 112.

3. One case suggested such an approach nearly fifty years ago. Merck & Co. v. Olin Mathieson Chem. Corp., 253 F.2d 156, 162 (4th Cir. 1958) ("[W]here the requirements of the Act are met, patents upon products of nature are granted and their validity sustained."). The approach suggested by the *Merck* court has not been widely accepted doctrinally. *See, e.g.*, Parker v. Flook, 437 U.S. 584, 593 (1978) ("First, respondent incorrectly assumes that if a process application implements a principle in some specific fashion, it automatically falls within the patentable subject matter of § 101 and the substantive patentability of the particular process can then be determined by the conditions of §§ 102 and 103.").

problems such as obviousness or insufficient disclosure—even if its opinions nominally recited broad subject matter limitations. Further, other patentability criteria could easily meet the underlying policy concerns of the Court. As a result, current patentable subject matter jurisprudence is based not on actual issues the Court historically decided, but instead on sweeping dicta that outlined unsubstantiated concerns about broad patent claims.[4]

Normatively, if courts always reach the right result for the wrong reasons, then little need be done to change the status quo. However, due to the lack of clear and rigorous precedential support for limiting patentable subject matter, current patentable subject matter jurisprudence is inconsistent and, if extended to logical conclusions, would bar patentability of almost any invention or discovery, which certainly would present a suboptimal outcome.[5] For example, in *Parker v. Flook*, the Supreme Court stated, "[W]e must proceed cautiously when we are asked to extend patent rights into areas wholly unforeseen by Congress."[6] However, the subject matter areas of some of the most important breakthroughs in history could not have been foreseen by Congress when patent laws were originally drafted. Many patents throughout history, from the telegraph to the airplane to the transistor, would be invalid under the *Flook* approach to unforeseeable technology areas.[7]

Only two years after *Flook*, in *Diamond v. Chakrabarty* the Court considered whether live bacteria used to clean oil spills could be patented.[8] The

---

4. *See, e.g., In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *58 (Fed. Cir. Oct. 30, 2008) (Rader, J., dissenting) ("Much of the court's difficulty lies in its reliance on dicta taken out of context from numerous Supreme Court opinions dealing with the technology of the past. In other words, as innovators seek the path to the next tech[no]-revolution, this court ties our patent system to dicta from an industrial age decades removed from the bleeding edge.").

5. *See* Kristin Osenga, *Ants, Elephants Guns, and Statutory Subject Matter*, 39 ARIZ. ST. L.J. 1087, 1093–1103 (2007) (describing inconsistent subject matter decisions).

6. *Flook*, 437 U.S. at 596. Interestingly, the Court cites *Deepsouth Packing Co. v. Laitram Corp.*, 406 U.S. 518, 531 (1972), to support its position. Congress enacted 35 U.S.C. § 271(f) in direct response to the narrow view of the patent law taken by the Court in *Deepsouth Packing Co.* *See* Microsoft Corp. v. AT&T Corp., 127 S. Ct. 1746, 1751–52 (2007); Pub. L. No. 98-622, Title I, § 101(a), 98 Stat. 3383 (Nov. 8, 1984). This "correction" implies that Congress intended the act to be construed broadly, and that the Court in *Deepsouth* and *Flook* need not have interpreted the statute so narrowly.

7. *See* Thomas D. Kiley, *Common Sense and the Uncommon Bacterium – is "Life" Patentable?*, 60 J. PAT. OFF. SOC'Y 468, 474 (1978) (arguing that the Court's approach in *Flook* would mean that new technologies could not be patented until Congress decided otherwise after the fact); Expanded Metal Co. v. Bradford, 214 U.S. 366, 384–86 (1909) (determining that the definition of "process" should be construed broadly, not narrowly).

8. Diamond v. Chakrabarty, 447 U.S. 303, 305 (1980). Prior attempts to clean oil used a combination of bacteria, whereas Chakrabarty genetically engineered a single bacterium that could "eat" multiple chemicals. *Id.* at 305 n.2. The *Chakrabarty* Court expressly compared the new bacteria with the combination in *Funk Bros. Seed Co. v. Kalo Inoculant Co.*, 333 U.S. 127 (1948), discussed in Part III.A. *Id.* at 310.

Court ruled that a living organism could very well be patentable if it was novel.[9] Importantly, the Court made clear that simply because a technology was unforeseen at the time a statute was enacted was no reason to exclude that technology from patentability.[10] This approach makes intuitive sense, given that the primary justification for patent law is to encourage new technologies.[11] However, *Chakrabarty*'s ruling is directly contrary to the policy set forth in *Flook*, which has never been expressly overruled.[12]

Nearly thirty years after *Flook*, court rulings have not borne any further clarity. In a 2006 dissent from dismissal in *Laboratory Corp. of America Holdings v. Metabolite Laboratories, Inc.*, three justices admitted that "the category of non-patentable 'phenomena of nature,' like the categories of 'mental processes,' and 'abstract intellectual concepts,' is not easy to define."[13] Similarly, the Federal Circuit's recent en banc opinion in *In re Bilski* admitted the difficulty.[14]

Despite the difficulty of providing clear definitions, these cases are trending toward more subject matter rejections. Further, scholars continue to advocate the "gatekeeping" role of courts in barring patents of a particular subject matter.[15]

---

9. *See id.* at 309–10 ("His claim is not to a hitherto unknown natural phenomenon, but to a nonnaturally occurring manufacture or composition of matter—a product of human ingenuity 'having a distinctive name, character [and] use.'" (quoting Hartranft v. Wiegmann, 121 U.S. 609, 615 (1887))).

10. *Chakrabarty*, 447 U.S. at 315.

11. *See id.* at 308 (stating that the Patent Act "embodied [Thomas] Jefferson's philosophy that 'ingenuity should receive a liberal encouragement'" (quoting Letter from Thomas Jefferson to Oliver Evans (May 2, 1807), *in* V THE WRITINGS OF THOMAS JEFFERSON 76 (H.A. Washington, ed., 1859).

12. *Id.* at 316 ("This is especially true in the field of patent law. A rule that unanticipated inventions are without protection would conflict with the core concept of the patent law that anticipation undermines patentability."); *see also* U.S. CONST. art I, § 8, cl. 8 ("To promote the Progress of . . . useful Arts . . . .").

13. Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 126 S. Ct. 2921, 2926 (2006) ("After all, many a patentable invention rests upon its inventor's knowledge of natural phenomena; many 'process' patents seek to make abstract intellectual concepts workably concrete; and all conscious human action involves a mental process.") (Breyer, J., dissenting). Courts invalidating claims based on subject matter have recognized this much as well. Parker v. Flook, 437 U.S. 584, 589 (1978) ("The line between a patentable 'process' and an unpatentable 'principle' is not always clear.").

14. *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *5 (Fed. Cir. Oct. 30, 2008) ("Unfortunately, this inquiry is hardly straightforward. How does one determine whether a given claim would pre-empt all uses of a fundamental principle? Analogizing to the facts of *Diehr* or *Benson* is of limited usefulness because the more challenging process claims of the twenty-first century are seldom so clearly limited in scope as the highly specific, plainly corporeal industrial manufacturing process of *Diehr*; nor are they typically as broadly claimed or purely abstract and mathematical as the algorithm of *Benson*.").

15. *See, e.g.*, Eileen M. Kane, *Splitting the Gene: DNA Patents and the Genetic Code*, 71 TENN. L. REV. 707, 725 (2004); David Olson, *Patentable Subject Matter: The Problem of the Absent Gatekeeper* 3–4 (Stanford Law School Center for Internet and Society 2006), *available at*

As a result, historical reliance on unexamined dicta may now lead to the wrong results, one of this Article's principal normative concerns.

The virtue of the proposed rule is that it provides a doctrinal standard for determining patentability that is more consistent and more rigorous than supposed "bright line" subject matter rules[16]—or at least as rigorous as the remainder of the statute will allow. The goals of this proposal are utilitarian: to increase the benefits of the patent system and to decrease the costs.

The proposal is agnostic about whether too many patents will result, in part because it is simply too hard to identify, let alone measure the effect of subject matter rules on innovation. Instead, the Article focuses on where historical court decisions focused: rejecting patents that do not pass muster.

Thus, this Article assumes that maximum social value is obtained by the issuance of only those patents that are justified under the statute.[17] As part of the analysis the Article examines constitutional limitation and statutory interpretation arguments, finding that the proposal is at best mandated and at worst not foreclosed by precedent.

Under this normative statutory metric, rigorous application of the Patent Act's patentability criteria ensures optimal patent issuance and scope.[18] On the one hand, extra-statutory, unprincipled subject matter bars do not reduce the number of bad patents, but might cause harm in other ways. In fact, even if without rigorous application of other patentability criteria, unclear subject matter rules create costs without adding much benefit. Some of the costs associated with the status quo are unsettled expectations, over and under-allowance of bad or

---

http://ssrn.com/abstract=933167. *But see* Rick Nydegger, *B2B, B2C and Other "Business Methods": To Be or Not To Be Patent Eligible?*, 9 U. BALT. INTELL. PROP. L.J. 199, 216 (2001) ("A statutory section that is as deeply founded on policy considerations as § 101 is ill suited to serve as a 'gatekeeper' to the grant of patent protection. That role is best left to considerations of the merits (e.g., novelty and nonobviousness under §§ 102 and 103) of a particular invention in the given technological field.").

16. Julie E. Cohen, *Reverse Engineering and the Rise of Electronic Vigilantism: Intellectual Property Implications of "Lock-Out" Programs*, 68 S. CAL. L. REV. 1091, 1168–69 (1995) ("In particular, the requirements . . . that a claimed invention be novel and nonobvious . . . may be used to accomplish what the statutory subject matter inquiry cannot achieve: a rule that permits analytic dissection of claims into statutory and non-statutory elements for purposes of identifying which computer program-related inventions are patentable."); Osenga, *supra* note 5, at 1091–92 ("The PTO and some commentators are using § 101 rejections as a means to avoid tackling other policy or practical issues that should be handled through other means. The rejections thus serve as proxies for inquiries that should be made more appropriately under other requirements of patentability, such as utility, novelty, nonobviousness, adequate written description, and enablement.").

17. *See, e.g.*, Dan L. Burk & Mark A. Lemley, *Policy Levers in Patent Law*, 89 VA. L. REV. 1575, 1669 (2003) [hereinafter Burke & Lemley, *Policy Levers*] ("[W]e think the solution is for the courts to get their decisions right, rather than for them to wash their hands of involvement in the calibration of policy."). This Article argues that courts cannot "get their decisions right" if those decisions are based on generalized subject matter rules.

18. *See* 35 U.S.C. §§ 100–103, 112.

good patents respectively, reduction in innovation caused by uncertainty, unnecessary examination costs, and increased litigation costs.

On the other hand, those who favor non-expansion and even contraction in patent protection might fear that the proposed rule will expand the types of discoveries considered patentable, leading to patents covering inventions that should be in the public domain.[19] Reliance on judicially created subject matter rules to answer this concern is misplaced for several reasons. First, judges lack the empirical information to make subject matter policy. Second, opinions must focus on a single case rather than entire industries (or multiple industries), which leads to unintended effects of any given rule. Third, the judiciary should not be responsible for legislating patent eligibility beyond the categories defined by Congress, especially where such pronouncements are admittedly divorced from the statute. For those who argue that application of the current patentability criteria would yield too many patents, overhaul of the statute rather than fidelity to it is warranted.

However, statutory overhaul may not be necessary to achieve the goal of reducing patenting in controversial areas. While this Article asserts no position on the number and types of patents that should issue, it does demonstrate that abandoning subject matter restrictions in favor of rigorous application of patentability requirements will not necessarily lead to more patents in controversial areas.[20] In fact, the proposal may reduce the number of discoveries that are currently considered patentable in a manner consistent with the goals of the Patent Act.[21]

Furthermore, the proposed rule does not foreclose congressional restriction—in a narrowly tailored and consistently applicable manner—of patentable subject matter based on actual evidence of harm caused by particular types of patents.[22] Of course, broad congressional action that suffers from the same problems as judicial opinions may not be desirable, but the judiciary should not limit the subject matter of *all* patents based on *any single case* at bar, and it certainly should not do so without concrete evidence of the supposed harm that an entire class of patents might allegedly cause.

Part I briefly describes the five prerequisites for obtaining a patent, including the requirements of patentable subject matter. Part II is descriptive; it examines

---

19 For example, by allowing tax methods patents, though even more subtle expansion is possible.

20 *Bilski*, 2008 WL 4757110, at *62 (Rader, J. dissenting) ("[R]eading section 101 as it is written will not permit a flurry of frivolous and useless inventions. Even beyond the exclusion for abstractness, the final clause of section 101-'subject to the conditions and requirements of this title'-ensures that a claimed invention must still satisfy the 'conditions and requirements' set forth in the remainder title 35. Id. These statutory conditions and requirements better serve the function of screening out unpatentable inventions . . . . ").

21 "Rigorous" here means strict fidelity to the Patent Act and its requirements of category, utility, novelty, nonobviousness, and description. 35 U.S.C. §§ 101, 102, 103, and 112. This is more fully described in Part II.D.

22 *See, e.g.*, 42 U.S.C. § 2181(a) (2000) (banning patents on certain nuclear technologies).

several patentable subject matter judicial decisions and reconciles each "subject matter" outcome with the five prerequisites of patentability. In each case, a patentability decision could have been—and often was—reached without determinative consideration of the patent's subject matter.

Parts III and IV are normative. Part III discusses how the proposed rule should apply to current or controversial technologies such as DNA and business method patents. Part IV discusses and responds to potential criticism of the proposed rule.

I. Obtaining a Patent

In order to obtain a patent, the inventor(s) must file an application that meets several criteria.[23] First, the invention must be described in sufficient detail so that one with ordinary skill in the subject matter of the patent (the "art") can both make and use the invention.[24] The invention described must meet the requirements for patentability: it must be useful,[25] it must be novel,[26] and it must not be obvious to one with ordinary skill in the subject matter of the patent.[27]

Additionally, inventions are patentable only if they are of an approved subject matter.[28] Subject matter standards emanate from two sources: legislation and case law.[29] The Patent Act describes broad subject categories, allowing patents for any "process, machine, manufacture, or composition of matter, or any new and useful improvement thereof."[30]

Case law, however, is more restrictive than the Patent Act. Since the early 1800s, courts have stated that patents incorporating products of nature,[31] natural phenomena,[32] mental steps,[33] and mathematical algorithms[34] should not issue.[35]

---

23. *See* 35 U.S.C. § 111.

24. *Id.* § 112. The specific invention must be "claimed" so that others know when they might be infringing. *Id.*

25. *Id.* §§ 101, 112.

26. *Id.* § 102(a). Novelty generally means that no one has "anticipated" the invention by previously inventing, describing, or using it.

27. *Id.* § 103(a).

28. *Id.* § 101.

29. *See id.* § 101; Diamond v. Chakrabarty, 447 U.S. 303, 309 (1980) (identifying subject matter that courts treat as unpatentable).

30. 35 U.S.C. § 101. There is often an important distinction between processes and the other possible subjects of inventions. Processes are a series of steps used to accomplish some end result or product, while machines and compositions are specific physical objects. Corning v. Burden, 56 U.S. 252, 267 (1854) ("A new process is usually the result of discovery; a machine, of invention."). As a result, processes may often be more abstract with less specific elements. On the other hand, sometimes a process for creating a composition may be novel, even though the composition is not. The product/process difference becomes a pivot on which many cases relating to patentable subject matter turn.

31. *See Funk Bros.*, 333 U.S. at 130.

32. *See generally* O'Reilly v. Morse, 56 U.S. 62, 116 (1854) (stating that "the discovery of a principle in natural philosophy or physical science" is not patentable).

However, during the last thirty years courts have removed many subject matter limitations; for example, the United States Patent and Trademark Office (PTO) has issued and courts have affirmed patents covering segments of DNA,[36] business methods,[37] and computer programs incorporating mathematical algorithms.[38] As a result, courts invalidate few patents on subject matter grounds, though the historical bars have never been overruled.[39]

This recent expansion of patentable subject matter, for example into tax strategies,[40] has caused consternation among scholars.[41] A primary concern is the notion that inventors might remove from the public domain not just particular inventions, but the broad types or categories of inventions that would create greater social value in the public domain.[42] This concern over patentable subject matter is somewhat misplaced. As discussed in Part II, patentable subject matter uncertainties in Supreme Court jurisprudence stem from a failure of the particular invention to qualify for a patent on other grounds. Further, as discussed in Parts

---

33. *See In re* Musgrave, 431 F.2d 882, 893 (C.C.P.A. 1970). *See generally In re* Comiskey, 499 F.3d 1365 (Fed. Cir. 2007) (stating that mental processes by themselves are not patentable).

34. *See* Gottschalk v. Benson, 409 U.S. 63, 67 (1972). *See generally Flook*, 437 U.S. at 591 (stating that a mathematical algorithm, without more, is not patentable).

35. Many of these cases predate the 1952 Patent Act, though judicial exclusions have never been tied to particular statutory language.

36. *See, e.g.*, Regents of the Univ. of Cal. v. Eli Lilly & Co., 119 F.3d 1559, 1566–69, 1571–74 (Fed. Cir. 1997) (invalidating one of the plaintiff's patents covering cDNA on written description grounds, not on patentable subject matter grounds, and finding a patent on "DNA Transfer vector and transformed microorganism" valid, but not infringed. U.S. Patent No. 4,431,740 (filed June 8, 1982)); U.S. Patent No. 5,589,579 (filed July 19, 1994) (presenting the identification, cloning, and characterization of tumor-associated carbonic anhydrase and its cDNA sequence).

37. *See generally* State St. Bank & Trust Co. v. Signature Fin. Group, Inc., 149 F. 3d 1368 (1998).

38. *See generally* Diamond v. Diehr, 450 U.S. 175 (1981). The algorithm in the abstract is still barred.

39. Dan L. Burk & Mark A. Lemley, *Inherency*, 47 WM. & MARY L. REV. 371, 406–07 (2005) [hereinafter Burk & Lemley, *Inherency*]; Eileen M. Kane, *Patent Ineligibility: Maintaining a Scientific Public Domain*, 80 ST. JOHN'S L. REV. 519, 519 (2006) ("[Subject matter rejections'] relative dormancy should not be mistaken for obsolescence.").

40. *See, e.g.*, U.S. Patent No. 6,567,790 (filed Dec. 1, 1999).

41. *See, e.g.*, John M. Conley & Roberte Makowski, *Back to the Future: Rethinking the Product of Nature Doctrine as a Barrier to Biotechnology Patents (Part II)*, 85 J. PAT. & TRADEMARK OFF. SOC'Y 371, 388 (2003) ("The PTO's current view of the product of nature doctrine can be stated succinctly: it is a dead letter."); Burton T. Ong, *Patenting the Biological Bounty of Nature: Re-examining the Status of Organic Inventions as Patentable Subject Matter*, 8 MARQ. INTELL. PROP. L. REV. 1, 23–24 (2004) (criticizing the court in *Merck & Co. v. Olin Mathieson Chem. Corp.*, 253 F.2d 156 (4th Cir. 1958), for ignoring patentable subject matter rules).

42. Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 868–69 (1990) (discussing the effect of patent scope on incentives).

III and IV, strict application of the patentability criteria should not have a significant deleterious effect on the patent system.

II. Patentable Subject Matter Through a Different Lens

Virtually all of the important[43] historical patentable subject matter cases may be explained by applying each of the other requirements for patentability. When viewed through this lens, subject matter concerns are at bottom patentability concerns.[44] The cases can be grouped into three broad categories: obviousness, specification/inventorship, and novelty/utility.

A. Obviousness

Several cases that otherwise appear to be subject matter cases instead apply rigorous obviousness thresholds, barring patentability of combinations that do not create synergies—a whole greater than the sum of the combined parts.[45] Some might argue that strict application of nonobviousness standards fell by the wayside after passage of the 1952 Patent Act; however the cases discussed herein continue to be cited well after passage of the 1952 Patent Act.[46] Furthermore, the Supreme Court's recent opinion in *KSR International Co. v. Teleflex Inc.* revived a stricter requirement of nonobviousness.[47]

---

43. Because appellate cases often cite to statements made in Supreme Court cases in applying or extending the law, "important" here refers to Supreme Court cases only. There are several important lower court cases, such as *Parke-Davis & Co. v. H.K. Mulford & Co.*, 189 F. 95 (C.C.S.D.N.Y. 1911) (L. Hand, J). Many of these cases will be discussed in Part III below.

44. *Cf.* United States Patent and Trademark Office, *Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility* (Nov. 22, 2005) (conflating subject matter question with other patentability criteria), *available at* http://www.uspto.gov/web/offices/pac/dapp/opla/preognotice/guidelines101_20051026.pdf. *But see* Kane, *supra* note 39, at 546:

> It could be argued that there are shadow doctrines behind each exclusion from patentable subject matter which amplify why they cannot be patented. The patenting of natural phenomena and laws of nature most directly implicates issues of novelty, while the patenting of abstract ideas would be most immediately objected to on disclosure grounds. This is not to suggest that these exclusions are redundant to existing doctrines—the Supreme Court certainly adheres to the categorical exclusions from patentable subject matter as the meaningful components of a public domain.

45. *See, e.g.*, KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1727 (2007); Funk Bros. Seed Co. v. Kato Inoculant Co., 333 U.S. 127, 127 (1948); Am. Fruit Growers, Inc. v. Brodex Co., 283 U.S. 1, 1 (1931).

46. *See, e.g.*, Giles S. Rich, *The Principles of Patentability*, 14 FED. CIR. B.J. 135, 145 (2004), *reprinted from* 42 J. PAT. OFF. SOC'Y 75 (1960) (arguing that 1952 Patent Act dispensed with the requirement of "invention" in favor of a less restrictive standard of obviousness to one with skill in the art).

47. *See, e.g.*, *KSR Int'l Co.*, 127 S. Ct. at 1742 ("The same constricted analysis led the Court of Appeals to conclude, in error, that a patent claim cannot be proved obvious merely by showing

For example, in *American Fruit Growers, Inc. v. Brogdex Co.*, the Court addressed a claim for fruit that had been soaked in a borax solution, creating mold-resistant fruit.[48] The Court ruled that a fruit combined with borax was simply the combination of two known raw materials, not something new, and therefore was not a "manufacture" under the statute.[49]

The Court followed similar reasoning in *Funk Brothers Seed Co. v. Kato Inoculant Co.*, in which the applicant sought to patent a combination of different bacteria.[50] The Court did not consider the claimed method of selecting bacteria to combine; it only determined whether the final combination could be patented.[51] The Court essentially held that an end-product combination of preexisting products is obvious if the individual functions of the combined parts do not change.[52] In so holding, the Court emphasized that the qualities of bacteria were properties of nature and thus could not be patented.[53] This pronouncement was hardly required to reach the holding—the Court could have simply ruled that a combination of parts is obvious if no new product features are created. Some might argue that this view of *Funk Brothers* pushes obviousness too far. The case cites *Cuno Engineering Corp. v. Automatic Devices Corp.*,[54] a case whose "flash of genius" holding was expressly overruled by statute in the 1952 Patent Act.[55] Despite *Funk Brothers's* dubious reliance on *Cuno*, the Court's obviousness jurisprudence soon reached this very rule, which is still valid today: a combination of known pieces that adds nothing new is obvious.[56]

Justice Frankfurter's concurring opinion in *Funk Brothers* pointed out the problems with the majority's focus on the naturalness of the components:

---

that the combination of elements was 'obvious to try.'").

48. Am. Fruit Growers, Inc. v. Brogdex Co., 283 U.S. 1, 6 (1931).

49. *Id.* at 11–12, 14.

50. Funk Bros. Seed Co. v. Kalo Inoculant Co., 333 U.S. 127, 130 (1948).

51. *Id.* at 130–31.

52. *See id.* at 131 ("The combination of species produces no new bacteria, no change in the six species of bacteria, and no enlargement of the range of their utility. Each species has the same effect it always had. The bacteria perform in their natural way. Their use in combination does not improve in any way their natural functioning. They serve the ends nature originally provided and act quite independently of any effort of the patentee."); Conley & Makowski, *supra* note 41, at 379 ("So, in the biological context, it is clearly insufficient to bring about, without more, an unprecedented combination of existing species.").

53. *Funk Bros.*, 333 U.S. at 130.

54. 314 U.S. 84, 91 (1941).

55. 35 U.S.C. § 103(a) (2000) ("Patentability shall not be negatived by the manner in which the invention was made.").

56. Great Atl. & Pac. Tea Co. v. Supermarket Equip. Corp., 340 U.S. 147, 152–53 (1950) ("A patent for a combination which only unites old elements with no change in their respective functions, such as is presented here, obviously withdraws what already is known into the field of its monopoly and diminishes the resources available to skillful men."), *cited with approval in* KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1739 (2007), *and in* Graham v. John Deere Co., 383 U.S. 1, 11–12 (1966). *Cf. Graham*, 383 U.S. 1, 15–17 (1966) (1952 Patent Act did not change pre-1953 patentability standards other than "flash of genius" requirement of *Cuno*).

> It only confuses the issue, however, to introduce such terms as "the work of nature" and the "laws of nature." For these are vague and malleable terms infected with too much ambiguity and equivocation. Everything that happens may be deemed "the work of nature," and any patentable composite exemplifies in its properties "the laws of nature."[57]

Consider what would have happened if the results of the bacterial combination had resulted in previously unknown but naturally occurring effects, such as the generation of electricity. The Court's focus on "natural phenomena" subject matter would have required the PTO to reject such a hypothetical patent covering a method for generating bacterial electricity no matter how novel and nonobvious it may have been.[58]

B. Specification/Inventorship

A second case grouping in which the Court cited subject matter issues as the rationale for its decisions, but which really turned on other factors, relates to specification and inventorship. In this group, the Court was concerned with whether the claimed "natural" invention was described or enabled in the patent[59] and thus whether the inventor actually invented the claimed invention in the first place.[60]

In *O'Reilly v. Morse*, the Supreme Court considered a claim that encompassed any process for transmitting printed information by an electromagnetic signal by any means.[61] While the Court had no problem affirming a particular form of such transmission—the telegraph—it invalidated the broader claim based on a failure to describe or enable the particular invention:[62]

> If this claim can be maintained, *it matters not by what process or machinery the result is accomplished.* For aught that we now know some future inventor, in the onward march of science, may discover a mode of writing or printing at a distance by means of the electric or galvanic current, *without using any part of the process or combination set forth in the plaintiff's specification* . . . .

---

57. *Funk Bros.*, 333 U.S. at 134–35 (Frankfurter, J., concurring).

58. *Cf.* Diamond v. Chakrabarty, 447 U.S. 306, 310 (1979) ("the patentee has produced a new bacterium with markedly different characteristics from any found in nature . . . ").

59. In fact, the Court in *Funk Brothers* could have ruled on specification grounds. Justice Frankfurter pointed out in a concurrence that the patent did not actually describe which strains of bacteria could be combined and thus was unpatentable for lack of disclosure. *Funk Bros.*, 333 U.S. at 133–34.

60. *See, e.g.*, Corning v. Burden, 56 U.S. 252, 256 (1854).

61. O'Reilly v. Morse, 56 U.S. 62, 77–78 (1853).

62. Robert A. Kreiss, *Patent Protection for Computer Programs and Mathematical Algorithms: The Constitutional Limitations on Patentable Subject Matter*, 29 N.M. L. REV. 31, 69 (1999) (agreeing that the Court believed "that Morse's claim was too broad" but attributing the ruling to "constitutional theories" rather than to statutory claiming requirements); Kane, *supra* note 15, at 748.

> . . . In fine he claims an exclusive right to use a manner and process *which he has not described and indeed had not invented, and therefore could not describe* when he obtained his patent.[63]

Despite the relatively narrow language that implicates both the specification and novelty requirements, it did not take long for *Morse* to be reinterpreted as a subject matter ruling that simply invalidated a claim for the natural phenomenon of electromagnetic communications.[64]

Similarly, in *Corning v. Burden*, the Supreme Court considered a method for making iron malleable.[65] The trial court instructed the jury that the patent covered *any* method of creating malleable iron, so long as that method performed the same steps as the patentee's machine.[66] The Court, however, made clear that the proper subject matter of the patent at bar was the *specific* machine described, and not the known process employed by the machine.[67] *Corning* is the analogue of *Morse*: if a particular means for achieving an end is invented, then the means may be patented, but the general end may not be patented if it is not new.[68] Here too, however, the Court focused on a specification and novelty issue; the inventor was limited only to the particular machine invented and described[69] and was not allowed a patent for a general process that he did not discover or describe.[70]

In *Powder Co. v. Powder Works* the patentee initially claimed methods for "exploding nitro-glycerin[]" but later attempted to modify the claims to include new explosive products (such as nitroglycerin combined with gunpowder) in reissue proceedings.[71] The case has been cited as relevant to the patentability of

---

63. *Morse*, 56 U.S. at 113 (emphasis added); *see also id.* at 118–19 (describing the importance of description of the patented invention). *Morse* may be the first famous use of re-issue to broaden a claim to cover later invented technology. *Id.* at 114. *See also* Mackay Radio & Tel. Co. v. Radio Corp. of Am., 306 U.S. 86, 98 (1939) (stating that the patentee cannot claim an invention that was not disclosed in the patent simply by broadening claims to cover competition).

64. The Telephone Cases, 126 U.S. 1, 534 (1888) ("The effect of [*Morse*] was, therefore, that the use of magnetism as a motive power, without regard to the particular process with which it was connected in the patent, could not be claimed, but that its use in that connection could. In the present case the claim is not for the use of a current of electricity in its natural state . . . .").

65. Corning v. Burden, 56 U.S. 252, 252 (1854).

66. *Id.* at 253.

67. *Id.* at 268 ("[I]t is well settled that a man cannot have a patent for the function or abstract effect of a machine, but only for the machine which produces it.").

68. *Id.* at 269.

69. *Id.*("In fine, his specification sets forth the 'particulars' of his invention, in exact accordance with its title in the patent, and in clear, distinct, unequivocal, and proper phraseology.").

70. *Id.* at 268 ("It is clear that Burden does not pretend to have discovered any new process by which cast iron is converted into malleable iron.").

71. Powder Co. v. Powder Works, 98 U.S. 126, 133 (1878). Reissue is a process by which a patentee may amend the claims of an issued patent after it issues. 35 U.S.C. § 251 (2000); 37 C.F.R. §§ 1.171, 1.173 (2007); *see also*, 3 JOHN GLADSTONE MILLS, III, DONALD CRESS REILEY, III, & ROBERT CLARE HIGHLEY, PATENT LAW FUNDAMENTALS § 15.109 (2d ed. 2008) (explaining requirements and procedure for reissue).

certain processes.[72] However, a simpler reading of *Powder Co.* shows that the Court ruled that a process claim (covering steps to a process) cannot reissue as a composition of matter claim (covering physical objects) if the initial patent specification did not disclose the newly claimed physical matter.[73]

C. Novelty/Utility

Novelty issues help illustrate the Court's construction of the "new and useful" requirement of 35 U.S.C. § 101.[74] These issues can be divided into two sub-categories that emanate from the claiming of natural phenomena—preexisting materials[75] and non-useful claims.[76]

1. Preexisting Materials and Methods

In *Cochrane v. Badische Anilin & Soda Fabrik*, the Court considered a composition of matter made by a new process.[77] It held that even though the composition was "artificial" when made by the new process, the composition had the same chemical make-up as a naturally occurring product and thus could not be novel.[78] One of the Court's key concerns was specification because the patent application described only the process and not the composition.[79]

---

72. William T. Goglia, Annotation, *Supreme Court's Views as to What is Patentable Subject Matter Under Federal Law as "Process," "Machine," "Manufacture," or "Composition of Matter,"* 65 L. Ed. 2d 1197, 1205 (1981).

73. *Powder Co.*, 98 U.S. at 135 ("[I]n all this specification there is not a hint of any new mixture or new composition of matter having been invented by the patentee.").

74. 35 U.S.C. § 101 (2000). The general consensus is that "new" is now subsumed by "novelty" under 35 U.S.C. § 102. *In re* Bergy, 596 F.2d 952, 960 (C.C.P.A. 1979) ("Notwithstanding the words "new and useful" in § 101, the invention is not examined under that statute for novelty because that is not the statutory scheme of things or the long-established administrative practice."). However, some argue that inventions and discoveries should be "new" independent of § 102 requirements. *In re* Nuijten, 500 F.3d 1346, 1363–64 (Fed. Cir. 2007) (Linn, J., dissenting) (arguing that "new" has meaning in addition to novelty provisions of § 102); Linda J. Demaine & Aaron Xavier Fellmeth, *Reinventing the Double Helix: A Novel and Nonobvious Reconceptualization of the Biotechnology Patent*, 55 STAN. L. REV. 303, 361, 364 (2002) (arguing that "new" in § 101 and "novel" in § 102 are distinct, even though "courts and commentators have been assuming that novelty and newness were the same since the 1952 Patent Act was passed"); *cf.* Glue Co. v. Upton, 97 U.S. 3, 6–7 (1877) (no "new" composition is created by breaking a known composition into pieces).

75. 35 U.S.C. § 102(a) bars a patent where the invention is "known or used by others in this country." The language of § 102 becomes important in inherency analysis as well as analysis relating to natural products not known or found in the U.S., as discussed in Part IV.C, *infra*.

76. 35 U.S.C. § 101. Creations must be "useful."

77. Cochrane v. Badische Anilin & Soda Fabrik, 111 U.S. 293, 311–12 (1884). The patent related to an improved process of preparing alizarine. *Id.* at 294.

78. *Id.*

79. *Id.* at 310 ("Every patent for a product or composition of matter must identify it so that it

In *The Wood-Paper Patent*, the Court considered extracts from wood to create paper and explicitly ruled that extracts from a known product cannot be novel because they already exist.[80] The Court stated, "'What the law looks to . . . is the inventor and discoverer who finds out and introduces a manufacture which supplies the market for useful and economical purposes with an article which was previously little more than the ornament of a museum.' But this is no such case."[81] Likewise, in *Glue Co. v. Upton*, the Court considered an "instant" glue created by crushing large, previously known glue flakes that were then grated into small, uniform grains of glue.[82] The inventor claimed the glue product but did not attempt to patent the method used to create the glue.[83] Unfortunately for the inventor, the Court ruled that breaking up a known substance into smaller parts was insufficiently novel to patent those parts as a separate product: "There is nothing new in the fact that the solution of a soluble substance is accelerated by increasing its fragmentary division."[84] The Court ruled that a new compound "must be more or less efficacious, or possess new properties by a combination with other ingredients; not from a mere change of form produced by a mechanical division."[85]

Finally, in *Parker v. Flook*, the Supreme Court considered a claim related to automobile catalytic converters.[86] The claimed method was for determining the level of temperature, pressure, or flow rate necessary to trigger an alarm and included a mathematical algorithm to determine the proper "alarm limit."[87] The Court ruled that the only allegedly "new" part of the three-step method was the mathematical algorithm.[88] The Court then held that discovery of a mathematical

---

can be recognized aside from the description of the process for making it, or else nothing can be held to infringe the patent which is not made by that process.").

80. The Wood-Paper Patent, 90 U.S. 566, 593 (1874). ("[I]t is equally clear, in cases of chemical inventions, that when, as in the present case, the manufacture claimed as novel is not a new composition of matter, but an extract obtained by the decomposition or disintegration of material substances, it cannot be of importance from what it has been extracted."). The Court also took issue with (but did not decide) the notion that "purification" of a product creates a new product. *Id.* at 594.

81. *Id.* at 596 (citation omitted).

82. Glue Co. v. Upton, 97 U.S. 3, 4–5 (1878). At the time of the invention, glue was sold in solid form, and soaked in water to create a malleable substance. *Id.* at 4. The invention apparently shortened the time it took for the glue to become viscous. *Id.* at 4–5.

83. *Id.* at 5.

84. *Id.* at 6.

85. *Id.* at 6–7 ("Where certain properties are known to belong generally to classes of articles, there can be no invention in putting a new species of the class in a condition for the development of its properties similar to that in which other species of the same class have been placed for similar development; nor can the changed form of the article from its condition in bulk to small particles, by breaking or bruising or slicing or rasping or filing or grinding or sifting, or other similar mechanical means, make it a new article, in the sense of the patent law.").

86. Parker v. Flook, 437 U.S. 584, 584 (1978).

87. *Id.*

88. *Id.* at 585–86. This is a "point of novelty" analysis that is generally disfavored. *See id.* at

algorithm cannot be novel even if the algorithm was previously unknown: "Whether the algorithm was in fact known or unknown at the time of the claimed invention, as one of the 'basic tools of scientific and technological work' it is treated as though it were a familiar part of the prior art."[89] In other words, the Court ruled that a scientific principle could not be novel because it must have existed in nature.[90] On the other hand, the decision could have been decided as a matter of disclosure, such that the patent claim was not enabled or properly described because the inventor omitted details about selection of the alarm limit.[91]

2. *Non-useful Claims*

While utility is generally considered a separate requirement for patentability, novelty and utility tend to merge with respect to claims for mathematical algorithms and similar methods that involve no "action."[92]

The utility/novelty nexus appeared in the early Supreme Court case, *Le Roy v. Tatham.*[93] Though the patent at issue related to mundane machines used to

---

599–600 (stating that "a claimed process [should] lose[] its status at subject matter patentability simply because *one step* in the process would not be patentable subject matter if considered in isolation") (Stewart, J., dissenting). It may, however, be relevant in obviousness analysis. Graham v. John Deere Co., 383 U.S. 1, 17 (1966) (discussing the step of comparing the claim to the prior art).

89. *Id.* at 591–92 (citation omitted); *see* Cohen, *supra* note 16, at 1169 (noting that most objections to computer software patents are lack of novelty and obviousness). Of course, another way to view the case is that *Flook* was a subject matter case disguised as a novelty case. This is not an unreasonable view; the difficulty of reconciling *Flook* with other precedents is discussed further below.

90. *See Flook*, 437 U.S. at 584 (citing Funk Bros. Seed Co. v. Kalo Inoculant Co., 333 U.S. 127, 130 (1948)) (explaining that a phenomena of nature is not novel by itself). Because *Flook* undertook a "point of novelty" analysis, another way to look at the issue is that the entire claim was obvious because the use of the algorithm added nothing to the remainder of the elements, which were previously known. *See Flook*, 437 U.S. at 594 n.16 (applying § 103 type analysis to § 101); Arrhythmia Research Tech., Inc. v. Corazonix Corp., 958 F.2d 1053, 1056–57 (Fed. Cir. 1992) ("In accordance with *Flook*, the claims were analyzed [as a whole] to determine whether the process itself was new and useful, assuming the mathematical algorithm was 'well known.'").

91. Richard S. Gruner, *In Search of the Undiscovered Country: The Challenge of Describing Patentable Subject Matter*, 23 SANTA CLARA COMPUTER & HIGH TECH. L.J. 395, 406 (2007) ("The Court held that this invention did not constitute patentable subject matter because it involved only a formula for computing an alarm limit without associated details on how to 'select the appropriate margin of safety, the weighing factor, or any of the other variables' and did not 'contain any disclosure relating to the chemical processes at work, the monitoring of process variables, or the means of setting off an alarm or adjusting an alarm system.'") (citations omitted); Osenga, *supra* note 5, at 1120 ("To satisfy enablement under § 112, the application must disclose the claimed invention sufficiently to allow a person having ordinary skill in the art to practice the invention without undue experimentation—the very essence of repeatability or predictability.").

92. 35 U.S.C. § 101 (2000) (inventions or discoveries must be "new and useful"); *In re* Bergy, 596 F.2d 952, 960 (C.C.P.A. 1979).

93. Le Roy v. Tatham 55 U.S. 156, 171 (1852).

manufacture metal pipes, *Le Roy* was one of the first cases to assert in dicta that laws of nature cannot be patented:

> [A] principle is not patentable. A principle, in the abstract, is a fundamental truth; an original cause; a motive; these cannot be patented, as no one can claim in either of them an exclusive right. . . . The same may be said of electricity, and of any other power in nature, which is alike open to all, and may be applied to useful purposes by the use of machinery.[94]

Despite this sweeping statement, the Court was concerned with the utility requirement as it related to nature: "In all such cases, the processes used to extract, modify, and concentrate natural agencies, constitute the invention. The elements of the power exist; the invention is not in discovering them, but in applying them to useful objects."[95] As in *Morse*, the Court also linked enablement and novelty to consider whether the applicant actually invented the claimed invention.[96] Despite the Court's dicta, the holding relied on obviousness and had little to do with natural phenomena; previewing *Funk Brothers* and other cases, the Court held that a combination of parts must be new.[97]

More recently, in *Gottschalk v. Benson*, the Supreme Court considered a patent relating to the mathematical conversion of binary coded decimals into pure binary format, a conversion that was known and could be done by pencil and paper.[98] *Gottschalk* is often cited for the notion that pure mathematical algorithms are unpatentable subject matter,[99] but the opinion implies that the Court was more concerned with the inventor's failure to describe the process in such a way that it was clear that the applicant actually invented the claimed invention.[100] The real concern appeared to be that the claim fell short of the

---

94. *Id.* at 174–75; *see also* Eames v. Andrews, 122 U.S. 40, 54 (1887) ("The novelty of the process under consideration does not lie in a mechanical device . . . . It consists in the new application of a power of nature . . . .").

95. *Le Roy*, 55 U.S. at 175.

96. *Id.* ("A new property discovered in matter, when practically applied, in the construction of a useful article of commerce or manufacture, is patentable; but the process through which the new property is developed and applied, must be stated, with such precision as to enable an ordinary mechanic to construct and apply the necessary process.").

97. *Id.* at 177.

98. Gottschalk v. Benson, 409 U.S. 63, 65 (1972); *see id.* at 67.

99. *See id.* at 71–72. While *Benson* does say this, the dicta is, by its own terms, a "nutshell" of the actual holding, which is that one may not patent a non-useful algorithm where the particular method for carrying out the process is neither described nor novel. This is an example of unchallenged dicta later interpreted as a bright-line rule. *See also* Parker v. Flook, 437 U.S. 584, 585 (1978).

100. *Benson*, 409 U.S. at 68 ("Here the 'process' claim is so abstract and sweeping as to cover both known and *unknown* uses of the BCD to pure binary conversion. The end use may (1) vary from the operation of a train to verification of drivers' licenses to researching the law books for precedents and (2) *be performed through any existing machinery or future-devised machinery or without any apparatus.*") (emphasis added); *see also id.* at 69–70 (discussing other cases in terms

specification and novelty requirements.[101] Furthermore, a pure algorithm with no practical purpose was not "useful" as required by § 101.[102]

In *Diamond v. Diehr*, the Court again considered whether a patent should issue if a claim included a mathematical algorithm, this time in a method for processing and curing rubber.[103] The process included a well-known algorithm, which calculated the time required to cure rubber.[104] The patent applicant argued, and the Court agreed, that the process could be novel and useful because the claimed invention described a process for accurately measuring the temperature that was later used in the mathematical algorithm.[105] Thus, the Court ruled that the patent could not be rejected on subject matter grounds.[106] The decision did not turn on the mathematical nature of one of the steps; indeed, the process could have contained a *non-mathematical* step that was well known, so long as it was only one step in the process.[107]

D. Rigorous Patentability

An alternate lens focused on patentable subject matter leads to a different view of patentability, which this Article calls "rigorous patentability." Under rigorous patentability, concerns about patentable subject matter are addressed primarily by the application of the patent requirements on a case-by-case basis. These requirements must be (a) systematic, logical, and as consistent as possible; (b) based on adherence to the statutory language; and (c) applied with a goal that only patents deserving of protection are issued.[108] Attention to rigorous

---

of definiteness). The Court does go on to define a process as one that changes matter from one state to another, but such language is not necessarily about subject matter. Rather, the language concerns definiteness where a particular step of the process was not defined. *Id.* at 70. In making this ruling, the Court cites to *Morse*, which invalidated Morse's broad claim based on specification issues as well. *Benson*, 409 U.S. at 68; O'Reilly v. Morse, 56 U.S. 62, 112–13 (1853). In another example, the Court forgave the failure to claim a particular method for grinding an ingredient to a powder because that step is just one in a chain that transforms a material. Cochrane v. Deener, 94 U.S. 780, 787–88 (1876). The claim was not simply for a process whereby an ingredient is ground to a powder; such processes have been around since the dawn of humankind. *Id.* at 788.

101. *Benson*, 409 U.S. at 71.

102. *Id.*; *see also* Kreiss, *supra* note 62, at 68. Note that all statutes are citing to 35 U.S.C. unless otherwise specified.

103. Diamond v. Diehr, 450 U.S. 175, 177 (1981).

104. *Id.*

105. *Id.* at 178–79, 187 ("Their process admittedly employs a well-known mathematical equation, but they do not seek to pre-empt the use of that equation.").

106. *Id.* at 191.

107. *See, e.g.*, *id.* at 181 ("The respondents' claims were not directed to a mathematical algorithm or an improved method of calculation but rather recited an improved process for molding rubber articles *by solving a practical problem* which had arisen in the molding of rubber products.") (emphasis added). The Court did not reach the question of whether the patent claim satisfied the statutory requirements of novelty or nonobviousness. *Id.* at 191.

108. Rigorous patentability may also resolve patent policy concerns unrelated to subject

application of the patentability standards would replace unclear and undefined subject matter rules based on unsupportable statutory interpretations of the Patent Act.[109]

What does rigorous patentability require? For the most part, the requirements are already set forth in the statute as interpreted by the courts.[110] The following is a short discussion of what rigorous patentability means with respect to the elements of patentability discussed in Part I.

•*Statutory Category*: A claimed invention must fit into one of the statutory categories: "process, machine, manufacture, or composition of matter, or any new and useful improvement thereof."[111] While rare, inventions do exist that may fall outside these categories.[112] The issue should not be whether a claim is, for example, a law of nature, but instead whether the claim falls into a statutory category.

•*Utility*: A claimed invention must meet practical utility standards.[113] Process and product claims must lead to a result that can be used to some substantial and specific practical end. Patents should not issue on inventions that are simply useful for further study.[114]

•*Novelty*: A claimed invention must be new.[115] No patent should issue for compositions that exist either artificially[116] or naturally,[117] unless they are purified

---

matter; however, such policy concerns are outside the scope of this Article.

109. *See In re* Bergy, 596 F.2d 952, 960–61 (C.C.P.A. 1979) ("Section 101 states three requirements: novelty, utility, and statutory subject matter. The understanding that these three requirements are separate and distinct is long-standing and has been universally accepted.... Thus, the questions of whether a particular invention is novel or useful are questions wholly apart from whether the invention falls into a category of statutory subject matter. . . ." (emphasis omitted)).

110. Burk & Lemley, *Policy Levers*, *supra* note 17, at 1590–93. Burk and Lemley point out that many of these criteria are more strictly enforced in some industries rather than others. *Id.* Rigorous patentability standards would dictate that such standards be applied equally in all industries, though there may be some costs to such uniformity.

111. 35 U.S.C. § 101 (2000).

112. *See In re* Nuijten, 500 F.3d 1346, 1357 (Fed. Cir. 2007) ("signal" does not fall into statutory category); Kreiss, *supra* note 62, at 58 ("[F]or the most part, laws of nature, natural phenomena, and abstract ideas do not fall within any of the four classes of patentable subject matter listed in § 101."); Rich, *supra* note 46, at 135 ("Russian is not a patentable invention because it is outside of the enumerated categories . . . ."). Note that Judge Rich believed that certain business models were also outside the statute. *Id.* at 135 ("Also outside [the statute] is one of the greatest inventions of our times, the diaper service.").

113. Brenner v. Manson, 383 U.S. 519, 534–35 (1966) (requiring specific and substantial utility); 35 U.S.C. §§ 101, 112 (requiring that an invention be "new and useful" and that "[t]he specification shall contain . . . the manner and process of making and using it").

114. *In re* Fisher, 421 F.3d 1365, 1371 (Fed. Cir. 2005) ("It thus is clear that an application must show that an invention is useful to the public as disclosed in its current form, not that it may prove useful at some future date after further research."). Part IV of this Article discusses why this requirement satisfies rigorous patentability requirements, but subject matter restrictions do not.

115. 35 U.S.C. § 102(a) ("A person shall be entitled to a patent unless (a) the invention was known or used by others in this country...."); *Id.* § 102(f) (no patent awarded if inventor "did not himself invent the subject matter sought to be patented"); Evans v. Eaton, 16 U.S. 454, 513–14

to a point that the invention is different in kind from what exists in nature.[118] As discussed in Part III, this standard may be difficult to apply,[119] but the focus of patentability decisions should be on novelty and not subject matter.

•*Obviousness*: A claimed invention must be nonobvious, and the determination of obviousness should not be limited to any particular test.[120] Instead, the court must have broad latitude to find an invention obvious.

(1818) ("[T]he 6th section of the general patent act . . . declares, that if the thing was not originally discovered by the patentee, but had been in use, or had been described in some public work, anterior to the supposed discovery of the patentee, judgment shall be rendered for the defendant, and the patent declared void." (emphasis omitted)); WILLARD PHILLIPS, THE LAW OF PATENTS FOR INVENTIONS; INCLUDING THE REMEDIES AND LEGAL PROCEEDINGS IN RELATION TO PATENT RIGHTS 150 (American Stationers Co. 1837) ("It is an essential requisite that the invention shall be *new*."). Note that under the 1952 Patent Act, only § 102 governs novelty and not the requirement of a "new" invention, as stated in § 101. *In re* Bergy, 596 F.2d 952, 961 (C.C.P.A. 1979) ("new" in § 101 defined solely under § 102), *cited with approval in* Diamond v. Diehr, 450 U.S. 175, 190 (1981).

116. *See* The Wood-Paper Patent, 90 U.S. 566, 593 (1874) ("When . . . the manufacture claimed as novel is not a new composition of matter, but an extract obtained by the decomposition or disintegration of material substances, it cannot be of importance from what it has been extracted."); *see also* Glue Co. v. Upton, 97 U.S. 3, 6 (1878) ("[T]o render the article new in the sense of patent law, it must be more or less efficacious, or possess new properties by a combination with other ingredients; not from a mere change of form produced by a mechanical division.").

117. *In re* Ridgway, 76 F.2d 602, 603 (C.C.P.A. 1935) ("[W]hile appellants might be entitled to a patent on a method of purifying alpha alumina, they would not be entitled to a patent on the article alpha alumina, a natural product, merely because of the degree of purity of the article.").

118. Parke-Davis & Co. v. H.K. Mulford Co., 189 F. 95, 103 (C.C.S.D.N.Y. 1911).

119. *Id.* ("The line between different substances and degrees of the same substance is to be drawn rather from the common usages of men than from nice considerations of dialectic.").

120. KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1738 (2007). The Supreme Court's recent opinion in *KSR* makes implementation of this standard much more likely. For example, "brute force" inventions that are the result of computer processing time or repetitive combinatorial experimentation rather than invention would usually be obvious. *See Funk Bros.*, 333 U.S. at 131 (noting that the combination of species may have led to a discovery, but "[i]t is no more than the discovery of some of the handiwork of nature and hence is not patentable"); John M. Golden, *Biotechnology, Technology Policy, and Patentability: Natural Products and Invention in the American System*, 50 EMORY L.J. 101, 115 (2001) ("However, advances in technology and in laboratory techniques have eased and automated much of this process, substantially routinizing a variety of tasks that had previously required considerable effort and ingenuity. . . . [T]he sequencing of [species' genomes] has become only a matter of attention and time."). Compositions created through the application of known processes to known materials would also be obvious. *See also* Great Atl. & Pac. Tea Co. v. Supermarket Equip. Corp., 340 U.S. 147, 152–53 (1950); Sakraida v. Ag Pro, Inc., 425 U.S. 273, 282 (1976), *both cited with approval, KSR Int'l Co.*, 127 S. Ct. at 1739–40. Inventions that were "obvious to try" would be obvious. *Id.* Inventions that are combinations of known elements that do not provide for functionality beyond the known elements would be obvious. *Id.* Of course, not all inventions meeting the above criteria would be obvious—those determinations would have to be made on a case-by-case basis.

Discretion is counterintuitive to a rigorous requirement; however, the ability to reject patents as obvious requires flexibility.[121]

•*Specification*: A claimed invention must be supported not only by a detailed disclosure enabling one skilled in the art to make and use the invention, but also by a full description of the invention[122] such that the PTO, courts, and other interested parties can determine whether the inventor actually invented the fullest scope of the claimed subject matter[123] and whether the inventor "possesses" all the elements of the claimed invention.[124]

III. Applying Rigorous Patentability

One test of this Article's proposal is whether application of rigorous patentability standards satisfactorily answers new or controversial questions of patentability without regard to non-statutory subject matter bars. This section applies the standards to a few areas of concern.

A. Business Methods

The United States Court of Appeals for the Federal Circuit's express sanction allowing business method patents in 1998[125] led to an increase in patent applications for processes divorced from physical transformations.[126] Critics

---

121. *See KSR Int'l Co.*, 127 S. Ct. at 1739 (finding that obviousness is a difficult standard to apply consistently and categorically stating, "[r]igid preventative rules that deny factfinders recourse to common sense, however, are neither necessary under our case law nor consistent with it").

122. Regents of the Univ. of Cal. v. Eli Lilly & Co., 119 F.3d 1559, 1566–67 (Fed. Cir. 1997); O'Reilly v. Morse, 56 U.S. 62, 118 (1853).

123. *But see* Christopher M. Holman, *Is Lilly Written Description a Paper Tiger?: A Comprehensive Assessment of the Impact of Eli Lilly and Its Progeny in the Courts and PTO*, 17 ALB. L.J. SCI. & TECH. 1, 80–82 (2007) (reviewing written description cases and finding that a strict rule is not broadly or consistently applied by the PTO or courts).

124. *See Eli Lilly & Co.*, 119 F.3d at 1567 ("Whether or not [the specification] provides an enabling disclosure, it does not provide a written description of the cDNA encoding human insulin, which is necessary to provide a written description of the subject matter of claim 5.").

125. State St. Bank & Trust Co. v. Signature Fin. Group, Inc., 149 F.3d 1368, 1375 (Fed. Cir. 1998) ("Since the 1952 Patent Act, business methods have been, and should have been, subject to the same legal requirements for patentability as applied to any other process or method."). Though the Federal Circuit has since disapproved of the "tangible, concrete and useful" test of *State Street*, it reaffirmed that business methods may be patented. *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *10 ("We rejected just such an exclusion in *State Street*, noting that the so-called 'business method exception' was unlawful . . . . ").

126. Anne H. Chasser, *Developments at the United States Patent and Trademark Office*, 19 TEMP. ENVTL. L. & TECH. J. 27, 28, 31 (2000) (discussing growth of patent applications and numbers of examiners in business methods and software); Mark A. Lemley & Bhaven M. Sampat, *Is the Patent Office a Rubber Stamp?*, Stanford Public Law and Legal Theory Workshop Paper Series Research Paper No. 999098, at 41 (July 2007), *available at* http://ssrn.com/abstract=999098 (Class 705 for business methods receives the most patent

assert that business methods have no place among patentable subject matter.[127] However, business methods have been approved as patentable subject matter for at least 150 years.[128] Samuel Morse claimed, and the Court upheld, "the system of signs, consisting of dots and spaces . . . in combination with machinery for recording them, as signals for telegraphic purposes."[129] This claim is nothing more than a particular business method for communicating by telegraph. Modern internet-based business methods are different from the telegraph only in medium, and are usually much narrower than Morse's claim.

In short, business methods should be patentable if they otherwise meet rigorous patentability standards. They are processes under § 100, which defines a process to include a new use of existing machines, compositions of matter, manufactures—even existing processes—and they are not otherwise barred by the statute.[130] Thus, if a business method is novel, nonobvious, and adequately described, no bar to patentability should exist, whether or not the process is tied to a machine or transforms something physical.[131]

One criticism of business method patents is that the PTO grants patents to otherwise obvious methods.[132] Yet, this is not a problem of subject matter, but of

---

applications); *see also, e.g.*, eBay, Inc. v. MercExchange, L.L.C., 547 U.S. 388, 390 (2006); U.S. Patent No. 6,085,176 (filed Mar. 8, 1999) (claiming "[a] computer-implemented method of searching for an item in a plurality of independently operated electronic auctions interconnected by a computer network, each electronic auction having an associated data repository, the method comprising: receiving input identifying an item; and instructing a software search agent to search for the item on the computer network in the respective data repositories of one or more of the electronic auctions").

127. Robert P. Merges, *As Many as Six Impossible Patents Before Breakfast: Property Rights for Business Concepts and Patent System Reform*, 14 BERKELEY TECH. L.J. 577, 581 (1999) (discussing "bad business concept patents"); Kreiss, *supra* note 62, at 52 ("[T]he *State Street* panel treats § 101 as if it could be read literally. . . . A little common sense shows that this cannot possibly be true.").

128. O'Reilly v. Morse, 56 U.S. 62, 101 (1853) ("The art is distinct from the means employed in its exercise; both may be, and under this patent are patented."). *See also Bilski*, 2008 WL 4757110, at *37 (Newman, J., dissenting) (listing business methods patents from 18th century England).

129. *Morse*, 56 U.S. at 86.

130. 35 U.S.C. § 100(b) (2000) (process includes any method, including a new use for a machine). In fact, they were expressly recognized by Congress after the *State Street* decision. *Id.* § 273 (prior users of business methods not liable for infringement).

131. *Bilski*, 2008 WL 4757110, at *10–14 (rejecting categorical exclusions for non-technological arts and business methods, but creating a new "machine or transformation" test for patentability of processes).

132. Nuno Pires de Carvalho, *The Primary Function of Patents*, 2001 U. ILL. J.L. TECH. & POL'Y 25, 59 (2001) ("The problems with some business ideas that have been granted patents are not peculiar to business ideas but respect all inventions: the problems of obviousness and utility."). The case that gave rise to the "business method exception" was a novelty case. *See* Hotel Security Checking Co. v. Lorraine Co., 160 F. 467, 472 (2d Cir. 1908) (holding that a method of hotel bookkeeping was non-novel, and stating in dicta that such processes cannot be patented in any event).

examination. Over time, the PTO has applied patentability criteria more strictly, granting fewer business methods patents.[133]

Another concern with business method patents is that they protect methods that are widely, even publicly, practiced, but that the PTO cannot discover such prior art in order to reject patents.[134] This too is not an issue of subject matter; patent law has never barred patents because others have used methods secretly.[135] Indeed, one of the rationales for patent law is to encourage disclosure of trade secrets.[136] Inventors who fail to patent a secret invention risk having their later use of that invention be found to infringe another's patent, even when the inventor secretly practiced it first.[137] In any event, in 1999, Congress considered the question and decided that business methods patents may still issue, but departed from earlier law by protecting prior, secret users of a business method from infringement suits.[138]

B. Tax Methods

Tax minimization claims are particular types of business methods that recently caused considerable concern. Policymakers and scholars question whether methods for tax minimization should be patentable.[139] As with other

133. Dennis Crouch, Evidence Based Prosecution IV: Business Method PAIR Entries (Oct. 19, 2006), http://www.patentlyo.com/patent/2006/10/evidence_based__1.html (business methods patents being examined more closely in the PTO); Dennis Crouch, Evidence Based Prosecution V: Business Method Rejections (Oct. 22, 2006), http://www.patentlyo.com/patent/2006/10/evidence_based__2.html (business method patents subject to more resistance in PTO than "general population").

134. Merges, *supra* note 127, at 589 ("There is every reason to believe that there is a vast volume of non-patent prior art in the software-implemented business concept field, as is widely believed to be the case with software patents in general.").

135. W.L. Gore & Assocs., Inc., v. Garlock, Inc., 721 F.2d 1540, 1550 (Fed. Cir. 1983), *cert. denied*, 469 U.S. 851 (1984).

136. Kewanee Oil Co. v. Bicron Corp., 416 U.S. 470, 487–88 (1974); *see also* Bonito Boats, Inc. v. Thunder Craft Boats, Inc., 489 U.S. 141, 151 (1989); WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW 328–29 (2003).

137. *See* Macbeth-Evans Glass Co. v. Gen. Elec. Co., 246 F. 695, 707 (6th Cir. 1917); 5 DONALD CHISUM, CHISUM ON PATENTS § 16.03[4] (Supp. 2005) ("One of the purposes of the patent system is to encourage prompt disclosure of new innovations. Innovators who decline to seek patents on innovations and, instead, utilize them as trade secrets can be said to act contrary to that purpose.").

138. Pub. L. 106-113, div. B, § 1000(a)(9) (codified at 35 U.S.C. § 273(b)(1) (2000)) (effective date Nov. 29, 1999). Admittedly, the protection of § 273 reduces the penalties for failing to patent a method.

139. Hearing on Issues Relating to the Patenting of Tax Advice Before the Subcomm. on Select Revenue Measures of the H. Comm. on Ways & Means, 109th Cong. 109–77 (2006) ("We at the USPTO recognize that the patenting of tax planning strategies has raised a number of concerns in Congress, the IRS, and the financial services community.") (statement of James A. Toupin, General Counsel, United States Patent and Trademark Office); William A. Drennan, *The Patented Loophole: How Should Congress Respond to this Judicial Invention?*, 59 FLA. L. REV.

business methods, tax methods fit into the process subject matter category under § 101.[140]

Under a rigorous patentability test, however, most tax minimization methods would be considered obvious.[141] While creative, pure tax methods are merely an obvious combination of transactions that are considered nontaxable under the Internal Revenue Code.[142] Tax methods may even be automated, but automation alone is not patentable unless the means for automation are novel and nonobvious.[143]

C. DNA and Other "Natural Products"

The patentability of naturally occurring biotechnological products, such as DNA,[144] is one area where abandoning subject matter restrictions[145] and applying rigorous patentability standards would likely disallow many patents that are currently allowed.[146] The Supreme Court has not recently considered the patentability of products derived from nature.[147] The Court had the opportunity to consider one such patent in *In re Bergy*,[148] a companion case to *Chakrabarty*.[149]

---

229 (2007); Matthew A. Melone, *The Patenting of Tax Strategies: A Patently Unnecessary Development*, 5 DEPAUL BUS. & COM. L.J. 437, 438 (2007).

140. Ironically, tax methods might be patentable under the Federal Circuit's stringent "machine or transformation" test if they were carried out by a machine.

141. Melone, *supra* note 139, at 459 ("Moreover, tax strategy patents invariably involve the combination of well known tax techniques that, when used in isolation, are patently obvious."). *But see* Hearing, *supra* note 139 (statement of Richard Gruner) ("In short, a patent mediated world of tax planning may be one in which greater efforts are devoted to the types of innovative tax planning methods that are nonobvious advances over prior methods and that can qualify for patents."); Drennan, *supra* note 139, at 257–59 (proving lack of novelty of tax patents may be difficult due to non-public prior use).

142. *See, e.g.*, U.S. Patent No. 6,567,790, col.2 l.43-67, col.3 l.1-16 (filed Dec. 1, 1999) (describing invention in terms of transactions that satisfy Tax Code and IRS regulations).

143. Dann v. Johnston, 425 U.S. 219, 230 (1976) (holding that automation of bookkeeping system is obvious).

144. *See In re* O'Farrell, 853 F.2d 894 (Fed. Cir. 1988) (providing a relatively simple description of DNA technology); Michael Davis, *The Patenting of Products of Nature*, 21 RUTGERS COMPUTER & TECH. L.J. 293, 309 (1995) (providing the same).

145. Stephen McKenna, Note, *Patentable Discovery?*, 33 SAN DIEGO L. REV. 1241, 1253–54 (1996) (arguing that legislative history shows Congress knew that natural compositions of matter were "discovered" and that they are thus patentable).

146. *But see* Dan L. Burk, *Biotechnology in the Federal Circuit: A Clockwork Lemon*, 46 ARIZ. L. REV. 441, 441–42, 449–50 (2004) (arguing that the combination of theories applied by the Federal Circuit are not beneficial to the biotech industry).

147. Helen M. Berman & Rochelle C. Dreyfuss, *Reflections on the Science and Law of Structural Biology, Genomics, and Drug Development*, 53 UCLA L. REV. 871, 889 (2006) ("Unlike other issues that arise in patent litigation, the status of gene and protein discoveries as statutory subject matter has managed to escape review at all adjudicatory levels."). *Chakrabarty* related to artificial entities, and *Funk Bros.* related to a combination, rather than a derivation.

148. *In re* Bergy, 596 F.2d 952 (1979).

However, Bergy withdrew his application prior to the ruling, and the Court dismissed the appeal as moot.[150]

Products derived from nature fit within the § 101 categories as compositions of matter or manufactures and should not be barred on subject matter grounds.[151] Three primary tests exist, however, that might limit claiming natural products under rigorous patentability standards: novelty, obviousness, and disclosure.[152] Each criterion sets important, but different, limits on the patentability of natural derivatives.[153]

1. Novelty/Utility

Since Judge Learned Hand's opinion in *Parke-Davis*, the test for novelty of naturally occurring products has been whether a product has been "isolated and purified" from its state in nature.[154] In general, novelty and utility are intertwined[155] such that determination of patentability depends on just how useful

149. Diamond v. Chakrabarty, 447 U.S. 303, 306 (1980).

150. Diamond v. Chakrabarty, 444 U.S. 1028 (1980).

151. Merck & Co. v. Olin Mathieson Chem. Corp., 253 F.2d 156, 161–62 (4th Cir. 1958) ("There is nothing in the language of the [1952 Patent] Act which precludes the issuance of a patent upon a 'product of nature' when it is a 'new and useful composition of matter' and there is compliance with the specified conditions for patentability. All of the tangible things with which man deals and for which patent protection is granted are products of nature in the sense that nature provides the basic source materials. The 'matter' of which patentable new and useful compositions are composed necessarily includes naturally existing elements and materials. A product of nature which is not a 'new and useful . . . machine, manufacture, or composition of matter' is not patentable, for it is not within the statutory definition of those things which may be patented. Even though it be a new and useful composition of matter it still may be unpatentable if the subject matter as a whole was obvious within the meaning of § 103, or if other conditions of patentability are not satisfied. In dealing with such considerations, unpatentable products have been frequently characterized as 'products of nature.' But where the requirements of the Act are met, patents upon products of nature are granted and their validity sustained." (citations omitted)).

152. *See* 35 U.S.C. §§ 102, 103, 112 (2000).

153. *Id.*

154. Parke-Davis & Co. v. H.K. Mulford Co., 189 F. 95, 103 (C.C.S.D.N.Y. 1911). Judge Hand's decision was not the first making this ruling, but it is the most famous. Kuehmsted v. Farbenfabriken of Elberfeld Co., 179 F. 701, 705 (7th Cir. 1910) (upholding the appellee's patent because the compound, though similar in make-up to appellant's, significantly increased Aspirin's therapeutic benefits); *see also In re* Bergstrom, 427 F.2d 1394, 1402 (C.C.P.A. 1970) ("[P]ure materials necessarily differ from *less* pure . . . materials, and if the latter are the only ones existing and available as a standard of reference . . . perforce the 'pure' materials are 'new' with respect to [the existing materials]."); Conley & Makowski, *supra* note 41, at 387 ("[T]he CCPA made clear that Bergstrom was a section 102 novelty case, not a section 101 patentable subject matter case.").

155. Demaine & Fellmeth, *supra* note 74, at 338 ("A doctrinal problem with the test is that, as one commentator has suggested, it mistakes utility for newness."); *cf.* GEORGE TICKNOR CURTIS, A TREATISE ON THE LAW OF PATENTS FOR USEFUL INVENTIONS IN THE UNITED STATES OF AMERICA (C.C. Little and J. Brown 1849) (novelty and utility discussed together in a single chapter).

the isolated and purified natural product may be.[156] This is true for DNA as well.[157] However, over time, the "purification" requirement has become less important in judicial decision-making.[158]

Rigorous application of patentability standards implies that "isolation from nature" should be re-examined with respect to novelty.[159] An illustrative example is *General Electric v. DeForest Radio*, in which the Third Circuit determined

---

156. *Kuehmsted*, 179 F. at 705 (determining that the patent for aspirin is valid: "And it makes no difference, so far as patentability is concerned, that the medicine thus produced is lifted out of a mass that contained, chemically, the compound; for, though the difference between [the patent and the prior art] be one of purification only – strictly marking the line, however, where the one is therapeutically available and the others were therapeutically unavailable – patentability would follow. In the one case the mass is made to yield something to the useful arts; in the other case what is yielded is chiefly interesting as a fact in chemical learning."); *Parke-Davis*, 189 F. at 103 ("But even if it were merely an extracted product without change, there is no rule that such products are not patentable. Takamine was the first to make it available for any use by removing it from the other gland-tissue in which it was found, and, while it is of course possible logically to call this a purification of the principle, it became for every practical purpose a new thing commercially and therapeutically."); Kane, *supra* note 15, at 739–40 (arguing that the *Merck* Court's decision eliminated the objection that products of nature are not new for purposes of patentability). *But see In re* Bergy, 596 F.2d 952, 960–61 (C.C.P.A. 1979) (novelty, utility, and subject matter should be separately considered); Conley & Makowski, *supra* note 41, at 374 ("Thus, the subject matter inquiry is whether the claimed invention is or is not a statutory machine, manufacture, or composition of matter, and the answer should not be influenced by the presence or absence of novelty or utility.").

157. Amgen, Inc. v. Chugai Pharm. Co., 927 F.2d 1200, 1206 (Fed. Cir. 1991) ("It is important to recognize that neither Fritsch nor Lin invented EPO or the EPO gene. The subject matter of claim 2 was the novel *purified and isolated* sequence which codes for EPO, and neither Fritsch nor Lin knew the structure or physical characteristics of it and had a viable method of obtaining that subject matter until it was actually obtained and characterized." (emphasis in original)).

158. Conley & Makowski, *supra* note 41, at 390 ("'Pure,' in other words, simply meant 'isolated.'") (quoting Schering Corporation v. Amgen, Inc., 35 F. Supp. 2d 375, 399 (D. Del. 1999)). Even *Parke-Davis* allowed the patentability of adrenalin salts because the particular method of making such salts had been attempted without success in the past and the resulting isolated product had great benefit. *Parke-Davis*, 189 F. at 103.

159. Rebecca S. Eisenberg, *Patenting the Human Genome*, 39 Emory L.J. 721, 723 (1990) ("An intuitively appealing objection to patent protection for DNA sequences in the human genome is that the sequences themselves are not new. The human genome resides in every cell of every human being. DNA sequences within this genome exist quite apart from the inventive efforts of the private parties who might seek to patent them, and thus no one may claim to have invented them."); *see also* Davis, *supra* note 144, at 331 ("Although one can forgive Judge Hand in *Parke-Davis* for his self-admitted ignorance as to scientific matters, particularly in view of the general lack of knowledge in the field of biochemistry at the turn of the century, the perpetuation of this legal fiction [of isolation] within the field of intellectual property is somewhat less understandable."); Oskar Liivak, *The Forgotten Originality Requirement: A Constitutional Hurdle for Gene Patents*, 87 J. PAT. & TRADEMARK OFF. SOC'Y 261, 282 (2005) ("[P]urifying and isolating a gene is nothing more than just copying it."); *see, e.g.*, Conley & Makowski, *supra* note 41, at 392 ("Fourth, words such as 'isolated,' 'purified,' and 'synthesized,' should not be accorded talismanic status.").

that purification of tungsten did not create a "new" composition of matter.[160] The inventor admittedly did not create the metal, nor did the inventor create the properties of the metal.[161] While *General Electric* is generally considered to be an outlier,[162] the court's analysis is an application of the rigorous standard of novelty for "new" compositions[163] and highlights how novelty may be a clearer way to handle isolated products of nature than subject matter.[164]

The difficult novelty question for compositions of matter is whether the purification of a substance is in degree or in kind—that is, determining whether the inventor has transformed the starting (preexisting) natural materials into something new rather than simply removed imperfections from what existed before.[165] Simple extraction cannot suffice.[166] If extraction alone were enough to render matter novel, then a person could patent blood because it was isolated from a body through the use of a needle and syringe.[167]

---

160. Gen. Elec. Co. v. De Forest Radio Co., 28 F.2d 641, 643 (3d Cir. 1928). The process to create a purer but non-novel composition may very well be novel, however, as it was in the *General Electric* case. *Id.*

161. *Id.*

162. ROBERT PATRICK MERGES & JOHN FITZGERALD DUFFY, PATENT LAW AND POLICY: CASES AND MATERIALS 111–12 (4th ed. 2007); Burk & Lemley, *Inherency*, *supra* note 39, at 408 n.172.

163. Conley & Makowski, *supra* note 41, at 392 (citing *General Electric* with approval: "[D]espite the absence of bright-line tests, the clear import of more than a hundred years of precedent is that, where a claimed invention has a natural precursor or variant, the differences must be quite robust.").

164. While Conley & Makowski consider biotechnological advances in terms of "product of nature" doctrine rather than "novelty," they do provide a very thorough discussion about how several biotech advances should be viewed under a "strict" comparison with pre-existing materials. Conley & Makowski, *supra* note 41, at 393–98.

165. *Parke-Davis & Co.*, 189 F. at 103; Merck & Co. v. Olin Mathieson Chem. Corp., 253 F.2d 156, 164 (4th Cir. 1958) ("[I]f the process produces an article of such purity that it differs not only in degree but in kind it may be patentable. If it differs in kind, it may have a new utility in which invention may rest."), *quoting In re* Merz, 97 F.2d 599, 601 (C.C.P.A. 1938). *See also* Conley & Makowski, *supra* note 41, at 386 ("Purity, in other words, is a basis for patentability only if it creates a material difference between the claimed product and its natural precursor."); Berman & Dreyfuss, *supra* note 147, at 891 (isolated genes and proteins are not different "in kind" from their natural sources); Michael Greenfield, Note, *Recombinant DNA Technology: A Science Struggling with the Patent Law*, 44 STAN. L. REV. 1051, 1069 (1992) ("In short, the central question is whether the greater degree of purity has resulted in such a significant change that a new and useful composition of matter has been created."). *Compare In re* Ridgway, 76 F.2d 602, 603 (C.C.P.A. 1935) (pure alpha alumina not novel) *and In re* King, 107 F.2d 618, 620 (C.C.P.A. 1939) (purified Vitamin C not novel) *with Kuehmsted*, 179 F. at 705 (aspirin is novel because it provides previously unknown and unavailable therapeutic benefit).

166. *See Parke-Davis*, 189 F. at 103.

167. *See, e.g.*, Stephanie Arcuri, Note, *They Call That Natural? An Analysis of the Term "Naturally Occurring" and the Application of Genes to the Patent Act*, 40 VAL. U.L. REV. 743, 745 (2006) (comparing gene extraction to plucking a blade of grass). *But see* McKenna, *supra* note 145, at 1270 ("The *Dennis* court recognized the absurdity of denying patent protection to the discoverer while rewarding the mechanic. There would seem to be no valid reason or sound support

Those in favor of DNA patents may argue that an extraction analogy is inapt. They would argue that the issue is not in the isolation of the blood, but instead that blood cannot be patented because people have known how to extract blood from humans and animals for an eternity. However, the issue is not one of timing: the first human to extract blood with a spear did not invent blood, even though throwing the spear at an animal may have been novel. Likewise, if the method for isolating DNA is novel and nonobvious then the method used would be patentable whether or not the resultant DNA is patentable. Further, using complementary DNA (cDNA) to create proteins was novel the first time the process was used, but after that the process is part of the prior art.

Patenting the actual DNA sequence is another matter. Though more difficult to achieve than extraction, mere isolation and purification of a nucleotide sequence may not be sufficient to render a strand of DNA novel where such DNA exists in human or animal bodies.[168] Fundamentally, DNA is an encoding of information[169] that allows for the production of certain proteins.[170] Thus, a purified cDNA sequence will produce the same protein that is produced by the gene in the human or animal body in the same way.[171] Isolation and purification simply may not create something that is novel even if the product provides a previously nonexistent use or benefit.[172] However, placing the cDNA into a bacterium might create something that is novel, though such a combination may be obvious.[173] Furthermore, if the cDNA were modified, spliced, or otherwise changed to behave in a way that it did not in a human body, then it might very

---

for a position which would deny to discoveries . . . the protection of our patent laws when such discovery is that an old, or at least well-known chemical product, will . . . produce new, unknown, and unexpected results, whereas one who puts together at least two old and well-known chemical substances . . . and gets new results helpful to man may receive patent protection.") (citing Dennis v. Pitner, 106 F.2d 142, 144 (7th Cir. 1939)).

168. Berman & Dreyfuss, *supra* note 147, at 891.

169. Rebecca S. Eisenberg, *Re-examining the Role of Patents in Appropriating the Value of DNA Sequences*, 49 EMORY L.J. 783, 797 (2000) (discussing DNA as an information repository and concerns about using the patent system to protect information).

170. Arti K. Rai, *Intellectual Property Rights in Biotechnology: Addressing New Technology*, 34 WAKE FOREST L. REV. 827, 836 (1999).

171. Berman & Dreyfuss, *supra* note 147, at 891 (stating that "[f]or genes, the information is identical whether the gene is isolated or not; for proteins, the shape in a crystal is no different from the shape in nature"); Conley & Makowski, *supra* note 41, at 394. ("[D]espite its nominal chemical distinctiveness, what is patented is functionally indistinguishable from natural DNA and RNA. It contains exactly the same genetic information as its natural counterpart. It can do precisely the same work as a naturally occurring gene-protein synthesis-and it employs precisely the same processes to do it, whether in the body or in the laboratory."); Eisenberg, *Patenting the Human Genome*, *supra* note 159, at 724.

172. Titanium Metals Corp. v. Banner, 778 F.2d 775, 782 (Fed. Cir. 1985) (claim is anticipated even if useful new properties of old composition are discovered). See *infra* Part IV, for a discussion of inherency and public benefit.

173. *See infra* Part III.C.2.

well be novel even when not combined with a bacterium.[174] Thus, patenting of genetically modified proteins and new ways to create those proteins would not be foreclosed.[175]

Here, distinguishing novelty from products of nature is important. Isolated DNA and synthesized cDNA are not available directly in nature.[176] Instead, non-coding segments of DNA,[177] called introns, are removed during the creation of isolated or synthesized DNA.[178] As such, they are not products of nature, and every extraction is to some extent purified despite no change in functionality.[179] Even so, they might not be novel.

A related and intertwined bar to patentability of some biotechnology inventions is lack of utility.[180] For example, an inventor may isolate a gene, but may not know what the gene does or how that gene might be used in the future.[181] As a result, the simple discovery of a gene sequence is not practically useful and cannot be patented.[182] The Federal Circuit has previously rejected patents on gene fragments for a lack of utility.[183] A rigorous application of utility standards would bar patentability of non-useful discoveries without a need to rely on product of nature subject matter requirements.[184]

---

174. Liivak, *supra* note 159, at 291 ("Amgen takes the naturally occurring human DNA and then changes the codons so that preferred expression codons for *E. coli* bacteria are used instead. Some person has decided that this change to the naturally occurring sequence is worth pursuing. Amgen has created an original DNA sequence.").

175. Davis, *supra* note 144, at 339–40; *cf.* Genentech, Inc. v. Wellcome Foundation, Ltd., 29 F.3d 1555 (Fed. Cir. 1994) (genetically engineered protein is not equivalent of naturally derived protein).

176. Jonathan Kahn, *Race-ing Patents, Patenting Race*, 92 IOWA L. REV. 353, 407–408 (2007) ("The PTO constructs cDNA as isolated . . . in the sense of separating the genetic material itself from nature. This is not a scientific process but a legal one.").

177. That is, segments that do not generate amino acids to be used in a protein.

178. Conley & Makowski, *supra* note 41, at 393–94; Eisenberg, *supra* note 159, at 727 n.25.

179. Conley & Makowski, *supra* note 41, at 393 ("[A]n inventor can justifiably say that the invention is not, and cannot be, a product of nature."); Golden, *supra* note 120, at 127–28 (stating that "with respect to biotechnology, the century-old 'purification exception' tends to swallow the rule").

180. Berman & Dreyfuss, *supra* note 147, at 889.

181. *Id.* at 886 ("After all, it is not only important that the compound exhibit the biochemical function sought; the drug must also be efficacious in humans and not harm the patient in unanticipated ways.").

182. Utility Examination Guidelines, 66 Fed. Reg. 1092, 1098 (Jan. 5, 2001); Berman & Dreyfuss, *supra* note 147, at 889 (stating that "[n]onetheless, a strong argument can be made that raw information about biological endowments should not be considered patentable unless the advance has end-product functionality"); Rebecca S. Eisenberg & Robert P. Merges, *Opinion Letter as to the Patentability of Certain Inventions Associated with the Identification of Partial cDNA Sequences*, 23 AIPLA Q.J. 1, 16–19 (1995).

183. *In re* Fisher, 421 F.3d 1365, 1371, 1379 (Fed. Cir. 2005) (denying gene fragment claims for failure to show use other than for further study).

184. Golden, *supra* note 120, at 129 ("*Brenner* [*v. Manson*]*'s* demand that a patentable invention provide a 'currently available' and 'specific' benefit could be used to block patents for

2. *Obviousness*

Strict application of obviousness rules may present another bar to patentability of natural derivations both in the final product and in the method used. For example, once the method of putting cDNA into a plasmid is known, then the combination of the two preexisting items should be considered obvious under *Funk Brothers*.[185] Further, where a known[186] method is used to isolate and purify a composition, "brute force" experimentation[187] will usually not be enough to render a newly isolated segment of DNA patentable.[188]

Purified, mutated, or otherwise modified DNA could still be patented if such a modification was not obvious.[189] If one were to create a novel and nonobvious automated technique, the end result might be nonobvious. Such DNA would be no different from the modified bacteria in *Chakrabarty*.[190] However, over time one would expect the number of patentable compositions to decrease as methods for their creation become well known.[191]

Application of an "obvious to try"[192] standard would most likely require a reversal of a number of cases, including *In re Bell*,[193] *In re Deuel*,[194] and *Amgen*,

---

DNA sequences for which 'practical utilities' are more posited than proven – a description that might apply to most existing DNA patent claims." (citations omitted)).

185. *Cf. In re* Mayne, 104 F.3d 1339, 1341, 1343 (Fed. Cir. 1997) (recombination of two proteins obvious); Eisenberg, *supra* note 159, at 726–27 (discussing application of combined DNA and plasmids with respect to *Funk Bros.*).

186. "Known" here means a method used under license from a third party, in the public domain, or even by the patentee if such method is not novel under § 102.

187. Berman & Dreyfuss, *supra* note 147, at 883 ("In today's world, however, this usually involves a variety of automated and computational techniques for screening compounds that are potentially bioactive.").

188. Pfizer, Inc. v. Apotex, Inc., 480 F.3d 1348, 1366 (Fed. Cir. 2007) (stating that the end result is obvious even though verification required testing). *But see* Andrew Chin, *Artful Prior Art and the Quality of DNA Patents*, 57 ALA. L. REV. 975, 976–77 (2006) (noting that poor quality prior art may impede novelty and obviousness analysis). As a policy matter society might want to incentivize investment in costly "brute force" development using known techniques, but that incentive should not come from the patent system; patents should instead promote the development of new techniques. Part IV discusses this issue in more detail.

189. Kahn, *supra* note 176, at 408 ("purification involves stripping the genetic material of its identity as a part of nature—ridding it of its natural associations").

190. Diamond v. Chakrabarty, 447 U.S. 303, 305 (1980).

191. Demaine & Fellmeth, *supra* note 74, at 306–07 (stating "a few decades ago it might have taken ten years to find a particular gene, but, with modern gene maps, a gene can now often be found with a fifteen second computer search"); Eisenberg, *Patenting the Human Genome*, *supra* note 159, at 730 ("The fact that the Patent and Trademark Office has issued patents on some DNA sequences thus does not necessarily portend that such patents will continue to issue in the future.").

192. *See* KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1742 (2007) ("When there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was

Inc. v. Chugai Pharmaceutical Co.[195] In these cases, the patentee began with some information and starting materials and applied known processes for isolating genetic material or for creating new chemical compositions.[196] The Federal Circuit held in each that so long as the inventor did not know what the result would be, the new compound would not be rendered obvious.[197]

Denying patents that are obvious to try is a realistic option in light of the stricter nonobviousness standard announced in *KSR*.[198] *KSR* even cites *Deuel* unfavorably, noting that failure to consider whether a patent was "obvious to try" is error.[199] The ruling has already had some effect on gene patents. In *Ex Parte Kubin*, the Board of Patent Appeals and Interferences ruled that a claim for cDNA was obvious to try in light of known processes.[200] Similarly in *PharmaStem Therapeutics, Inc. v. Viacell, Inc.*, the Federal Circuit ruled that confirming the suspicions of prior theorists cannot justify a patent: "[T]he inventors merely used routine research methods to prove what was already believed to be the case. Scientific confirmation of what was already believed to be true may be a valuable contribution, but it does not give rise to a patentable invention."[201]

---

obvious to try might show that it was obvious under § 103.").

193. *In re* Bell, 991 F.2d 781, 785 (Fed. Cir. 1993).

194. *In re* Deuel, 51 F.3d 1552, 1559 (Fed. Cir. 1995) (stating that cDNA may be patented even though it was obvious to try a known procedure for isolating DNA). For a thorough discussion of obviousness relating to DNA technology prior to *Deuel*, *see* Eisenberg, *Patenting the Human Genome*, *supra* note 159, at 729–35.

195. Amgen, Inc. v. Chugai Pharm. Co., 927 F.2d 1200, 1209 (Fed. Cir. 1991) ("obvious to try" known methods in different combinations does not render DNA claim obvious).

196. *Deuel*, 51 F.3d at 1555–56; *Bell*, 991 F.2d at 782–83; *Amgen*, 927 F.2d at 1209.

197. *Deuel*, 51 F.3d at 1559 (stating that "a conceived method of preparing some undefined DNA does not define it with the precision necessary to render it obvious over the protein it encodes"); *Bell*, 991 F.2d at 784; *Amgen*, 927 F.2d at 1209.

198. KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1742 (2007). ("When there is a design need or market pressure to solve a problem and there are a finite number of identified, predictable solutions, a person of ordinary skill has good reason to pursue the known options within his or her technical grasp. If this leads to the anticipated success, it is likely the product not of innovation but of ordinary skill and common sense. In that instance the fact that a combination was obvious to try might show that it was obvious under § 103.").

199. KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1739 (2007). The Federal Circuit relied on *Deuel* in the opinion reversed by the Supreme Court. Teleflex, Inc. v. KSR Int'l Co., 119 Fed. App'x 282, 289 (Fed. Cir. 2005), *rev'd* KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727 (2007).

200. *Ex Parte* Kubin, 83 U.S.P.Q.2d (BNA) 1410, 1414 (B.P.A.I. 2007). "The 'problem' facing those in the art was to isolate NAIL cDNA, and there were a limited number of methodologies available to do so. The skilled artisan would have had reason to try these methodologies with the reasonable expectation that at least one would be successful." *Id.* (quoting KSR Int'l Co. v. Teleflex Inc., 127 S. Ct. 1727, 1742 (2007)). *In re Kubin* is currently pending before the Federal Circuit. *Id.*, appeal docketed No. 08-1184 (Fed. Cir. Jan 31, 2008).

201. PharmaStem Therapeutics, Inc. v. Viacell, Inc., 491 F.3d 1342, 1363–64 (Fed. Cir. 2007). *But see* Takeda Chem. Indus., Ltd. v. Alphapharm Pty., Ltd., 492 F.3d 1350, 1356–57 (Fed. Cir. 2007) (citing *Deuel* with approval and ruling that use of routine methods need not render a

One difficulty with a stricter nonobviousness standard in biotechnology is potential conflict with legislative pronouncements on the obviousness of biotechnological inventions. For example, § 103(b) declares that minor biotechnological process improvements are nonobvious by fiat where the result of the process is a patentable composition.[202] If, however, a minor process improvement[203] was obvious to try and thus resulted in a novel but obvious composition, then under *KSR*, both the process and the composition would be obvious.[204] Section 103(b) could be read to mean that the modified process is nonobvious because it led to a new composition, and the new composition is nonobvious because it was created by a nonobvious process. This circular reading would lead to patentability of the composition without evidence of nonobviousness. The purpose of § 103(b) was to protect potentially obvious processes that yield novel and nonobvious compositions, not to protect obvious compositions created by obvious processes.[205]

The solution to the § 103(b) conundrum is to focus first on the end composition, whose novelty and nonobviousness is a condition precedent in the statute.[206] If the composition is unpatentable under a rigorous novelty test, then the process may not be patented under § 103(b).[207] If the composition is novel, the composition should be considered independently to determine whether it is obvious, including the range of production processes available, the starting materials, the techniques used, and how much was already known about the composition's chemical family. If the result is that the composition is nonobvious, then it is patentable, as is its creation process under § 103(b). If not, then the process must withstand the nonobviousness test on its own.

---

compound "obvious to try" where selection of the starting materials is not obvious).

202. 35 U.S.C. § 103(b) (2000). "[A] biotechnological process using or resulting in a composition of matter that is novel under section 102 and nonobvious under subsection (a) of this section shall be considered nonobvious if . . . claims to the process and the composition of matter are contained in either the same application for patent or in separate applications having the same effective filing date . . . ." *Id.* This statute was enacted in response to *In re Durden*, which held that a process for creating a composition was obvious given disclosures of the starting materials and a patent on the final composition. *In re* Durden, 763 F.2d 1406, 1411 (Fed. Cir. 1985); *see also* 2 DONALD S. CHISUM, CHISUM ON PATENTS § 5.04(8)(b)(ii)(B) (2004). The statute is largely irrelevant today because the Federal Circuit considers most process improvements to be nonobvious. *In re* Ochiai, 71 F.3d 1565, 1569 (Fed. Cir. 1995); Chisum, *supra* note 202, at § 5.04(8)(b)(ii)(B). However, the *KSR* decision may rejuvenate the use of § 103(b).

203. Existing processes would arguably be non-novel and also render the composition obvious.

204. *KSR*, 127 S. Ct. at 1742–43.

205. *See* 141 CONG. REC. S11201-03, S11207 (1995) (statement of Sen. Hatch) ("[T]he current patent law is not adequate to protect our creative American inventors who are on the cutting edge of scientific experimentation . . . . ").

206. *See* 35 U.S.C. § 103.

207. The process may, of course, be nonobvious pursuant to § 103(a). 35 U.S.C. § 103(a) (2000).

3. Description

In addition to claiming novel and nonobvious compositions, inventors must fully describe claimed compositions in order to obtain a patent.[208] As discussed in Part II, the description/specification requirement was critical in early Supreme Court jurisprudence that implicated patentable subject matter.[209] The Federal Circuit recently revived a "strict" description requirement in biotechnology and chemistry areas.[210] For example, in *Regents of the University of California v. Eli Lilly & Co.* the patent disclosed the nucleotide sequence for the gene that produces insulin in rats.[211] However, the claim was broader than the disclosure, claiming technology relating to cDNA in humans and in all vertebrates.[212] The Federal Circuit affirmed invalidation of the broad claims because the specification did not describe anything other than rat DNA, despite the fact that the patent described how non-rat DNA could be obtained.[213] Insisting that the applicant demonstrate "possession" of the invention by fully describing it will tend to reduce the number of broad patents covering basic biological functions.[214]

---

208. 35 U.S.C. § 112 (2000).

209. *See, e.g.*, O'Reilly v. Morse, 56 U.S. 62, 113 (1854).

210. Regents of the Univ. of Cal. v. Eli Lilly & Co., 119 F.3d 1559 (Fed. Cir. 1997).

211. *Id.* at 1562–63.

212. *Id.* at 1567.

213. *Id.* at 1568 ("A written description of an invention involving a chemical genus, like a description of a chemical species, 'requires a precise definition, such as by structure, formula, [or] chemical name,' of the claimed subject matter sufficient to distinguish it from other materials.") (quoting Fiers v. Revel, 984 F.2d 1164, 1171 (Fed. Cir. 1993)). In technical patent law terms, the court ruled that even if the claim is enabled, it must still be described. *Id.* at 1567.

214. Berman & Dreyfuss, *supra* note 147, at 899–900. "[T]his approach may be appropriate to prevent patentees from gaining control over products that they have not in fact discovered . . . ." *Id.* Berman & Dreyfuss argue that strict written description arguments may have the negative consequence of barring patents for those engaged in fundamental research, and creates incentives for creative claiming to avoid restrictions, among other negative consequences. *Id. See also* Rai, *supra* note 170, at 839 (strict written description may deny patents that should be granted). *But see* Burk, *supra* note 146, at 442–43 (arguing that a narrow written description requirement can lead to the patenting of obvious advances and suggesting that "[i]f conception requires detailed knowledge and revelation about the structure or detailed physical qualities of the molecule, then in order for a molecule to be 'obvious,' it needs to meet the same criteria – the same degree of detail within the prior art is required for obviousness that is needed in the mind of the inventor for conception"). These criticisms need not be true, however, if each patentability requirement is addressed separately, especially given that compositions might be "obvious to try." *See, e.g.*, Lockwood v. Am. Airlines, Inc., 107 F.3d 1565, 1571–72 (Fed. Cir. 1997) (invention is not disclosed even if it would be obvious from the disclosure).

D. Mathematical Algorithms and Computer Software

Like those in other areas, subject matter rejections of software patents should be rare because software falls within the statutory categories.[215] If combined with a useful process or device, mathematical and other algorithms could be patentable,[216] as discussed in the well-known cases of *Diamond v. Diehr,*[217] *In re Alappat,*[218] and *AT&T Corp. v. Excel Communications.*[219] The Federal Circuit's most recently announced test, whether the software or algorithm is tied to a machine or transformation, affirms that computer software should not be subject to a categorical exclusion.[220]

However, this Article's proposal would abandon the "machine or transformation" test and all attempts to define software as distinct from abstract ideas, mathematical algorithms, or any "post-solution activity."[221] Instead, all such claims would be patentable only if they meet rigorous standards for patentability.[222]

For example, a standalone mathematical algorithm would not be patentable because it does not have practical utility, even if the algorithm was a process

---

215. Osenga, *supra* note 5, at 1109 ("A software-related invention will nearly always be a process and will almost never, so long as it has a practical application, fall within one of the three exclusions for law of nature, natural phenomenon, or abstract idea."). Osenga identifies articles that take issue with the statutory categories for software. *Id.* at 1107 n.156. Note that, contrary to Osenga, the proposal in this Article would allow software inventions – on subject matter grounds, at least – even if they included laws of nature, natural phenomenon, or abstract ideas. *See also* Donald S. Chisum, *The Patentability of Algorithms*, 47 U. PITT. L. REV. 959, 972–92 (1986) (discussing and criticizing the *Benson* opinion); Arrhythmia Research Technology, Inc. v. Corazonix, Corp., 958 F.2d 1053, 1056 (Fed. Cir. 1992) (confirming that *Benson* does not require all mathematical algorithms to be unpatentable).

216. Chisum, *supra* note 215, at 997 (arguing that it is well settled that mathematical algorithms have been patentable as part of a larger claim) (citing MacKay Radio & Telegraph Co. v. Radio Corp. of Am., 306 U.S. 86 (1939)).

217. Diamond v. Diehr, 450 U.S. 175, 191 (1981).

218. *In re* Alappat, 33 F.3d 1526, 1540–41 (Fed. Cir. 1994) (en banc).

219. AT&T Corp. v. Excel Commc'ns, 172 F.3d 1352, 1355–56 (Fed. Cir. 1999) (software patentable as a method).

220. *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *10 n.23 ("[W]e decline to adopt a broad exclusion over software or any other such category of subject matter beyond the exclusion of claims drawn to fundamental principles set forth by the Supreme Court.").

221. *See id. at* *8.

222. State St. Bank & Trust Co. v. Signature Fin. Group, Inc., 149 F. 3d 1368, 1375 (Fed. Cir. 1998) ("Section 101 specifies that statutory subject matter must also satisfy the other 'conditions and requirements' of Title 35, including novelty, nonobviousness, and adequacy of disclosure and notice.") (citations omitted); Vincent Chiappetta, *Patentability of Computer Software Instruction as an "Article of Manufacture:" Software as Such as the Right Stuff*, 17 J. MARSHALL J. COMPUTER & INFO. L. 89, 93 (1998) ("In addition, the resulting failure to clearly and properly define the actual nature of software inventions by applying the patentable subject matter analysis leads to inadequate identification of prior art and insufficiently stringent review for novelty and nonobviousness.").

under § 101.[223] While such an algorithm may allow for new, faster, or more accurate computation of real world effects, it does not act unless coupled with some physical process or device.[224] Practical utility requires some "action" beyond the possibility of calculation.[225] In general, determining whether a process has practical utility should be less difficult than determining whether a process is solely a mathematical algorithm.[226]

With respect to software patents, rigorous patentability requires a complete specification: complete written description, enablement, and specific source code or pseudo-code to fulfill the best mode requirement.[227] Rigorous patentability also requires extensive obviousness analysis to ensure that a claimed series of steps was not only unknown but also sufficiently inventive to separate it from the prior art.[228]

Professors Burk and Lemley point out that software engineers are considered extremely skilled for purposes of enablement.[229] This level of sophistication has the effect of making obviousness findings more likely because of the small leap it would take for a highly skilled engineer to improve an algorithm.[230] Higher

---

223. Chiappetta, *supra* note 222, at 106 ("Focusing on 'usefulness/utility' of a software invention in a 'useful arts' sense contains the key to resolving the software as patentable subject matter conundrum.") (citing Examination Guidelines for Computer-Related Inventions, 61 Fed. Reg. 7,478 (1996)); *cf. State St. Bank & Trust Co.*, 149 F.3d at 1375 ("The question of whether a claim encompasses statutory subject matter should not focus on *which* of the four categories of subject matter a claim is directed to – process, machine, manufacture, or composition of matter – but rather on the essential characteristics of the subject matter, in particular, its practical utility.") (footnotes omitted).

224. Kreiss, *supra* note 62, at 68 ("Purely mathematical algorithms provide one illustration of the theory that abstract ideas are not patentable.").

225. *In re* Schrader, 22 F.3d 290, 295 (Fed. Cir. 1994) (some sort of transformation is required, even if not physical); *Bilski*, 2008 WL 4757110, at *12 (transformation of data about physical objects required).

226. *In re* Warmerdam, 33 F.3d 1354, 1359 (Fed. Cir. 1994) (difficult to determine whether a process is an algorithm).

227. Dan L. Burk & Mark A. Lemley, *Is Patent Law Technology-Specific?*, 17 BERKELEY TECH. L.J. 1155, 1162 (2002) [hereinafter Burk & Lemley, *Technology-Specific?*] ("[A] series of recent Federal Circuit decisions has all but eliminated the enablement and best mode requirements. In recent years, the Federal Circuit has held that software patentees need not disclose source or object code, flow charts, or detailed descriptions of the patented program."); Lawrence D. Graham & Richard O. Zerbe, Jr., *Economically Efficient Treatment of Computer Software: Reverse Engineering, Protection, and Disclosure*, 22 RUTGERS COMPUTER & TECH. L.J. 61, 96–97 (1996) (discussing problems with enforcement of written description, enablement, and best mode).

228. Cohen, *supra* note 16, at 1169 ("Intuitively, the most troubling aspect of many computer program-related patents is that they appear to reward the inventor for recognizing the obvious—that a given function may be performed more efficiently or more accurately if computerized—and using general purpose computer equipment and standard programming techniques to computerize it."); Burk & Lemley, *Technology-Specific?*, *supra* note 227, at 1167–68 (recent Federal Circuit cases "viewed obviousness as a rather substantial hurdle to patenting software").

229. Burk & Lemley, *Technology-Specific?*, *supra* note 227, at 1168, 1170.

230. *Id.*

rejection rates for software patent claims may not be a social detriment given concerns about patenting software generally.[231] Regardless of the optimal level of patenting, when software claims do issue, their disclosures will be opaque because highly skilled engineers require less detail for enablement.[232] Perhaps a way to avoid the conflict between enablement and obviousness is to focus on the description and best mode requirements.[233] Even if a bare-bones specification enables one skilled in the art to make and use the claimed software, a patentee should still be required to fully describe the software and to disclose the best way of making and using the invention, two requirements that do not allow gap-filling by one skilled in the art.[234] While it is true that courts have assumed that *any* programmer only needs a broad functional description to write a program,[235] rigorous enforcement of description and best mode requirements would require more proof that the patentee actually possessed a particular invention.[236] The resolution of these questions should be based on examination of patentability criteria, not on an absolute subject matter bar.

E. Natural Phenomena

Like products of nature, natural phenomena should not be unpatentable *per se* because most patents are based in part on such phenomena.[237] The most recent case to focus attention on this issue is *Laboratory Corp. of America Holdings v.*

---

231. *See* Cohen, *supra* note 16, at 1169.

232. Burk & Lemley, *Technology-Specific?*, *supra* note 227, at 1168, 1170.

233. 35 U.S.C. § 112 ¶ 1 (2000) (requiring inventors to describe the best way to practice an invention).

234. 35 U.S.C. § 112 (2000); Robotic Vision Sys. v. View Eng'g, 112 F.3d 1163, 1165–66 (Fed Cir. 1997); Lockwood v. Am. Airlines, 107 F.3d 1565, 1571–72 (Fed. Cir. 1997) ("Entitlement to a filing date does not extend to subject matter which is not disclosed, but would be obvious over what is expressly disclosed . . . . [A] prior application itself must describe an invention, and do so in sufficient detail that one skilled in the art can clearly conclude that the inventor invented the claimed invention as of the filing date sought.").

235. Burk & Lemley, *Technology-Specific?*, *supra* note 227, at 1164–65; *see, e.g.*, *Robotic Vision Sys.*, 112 F.3d at 1166 ("[I]t is generally sufficient if the functions of the software are disclosed, it usually being the case that creation of the specific source code is within the skill of the art."). *But see* Scott Elengold, Note, *An Inquiry into Computer System Patents: Breaking Down the "Software Engineer,"* 61 N.Y.U. ANN. SURV. AM. L. 349, 372 (2005) (suggesting courts to consider different types of programmers and the skills they have in specialized areas such as graphic design).

236. Chemcast Corp. v. Arco Indus. Corp., 913 F.2d 923, 928 (Fed. Cir. 1990) (stating that best mode is a factual determination).

237. *See, e.g.*, MacKay Radio & Telegraph Co. v. Radio Corp. of Am., 306 U.S. 86, 94 (1939) (radio antenna could be patentable subject matter even though its dimensions directly correspond to a natural phenomenon); Wright Co. v. Paulhan, 177 F. 261, 263–64 (C.C.S.D.N.Y 1910) (L. Hand, J.), *rev'd on other grounds*, 180 F. 112 (2d Cir. 1910). The Wright Brothers' invention, for example, was based primarily on the discovery that a rudder could be used for stabilization of airlift caused by wing "warping." *Id.*

Metabolite Laboratories, Inc., in which the Supreme Court first granted certiorari and then dismissed the petition as improvidently granted, which left scholars and practitioners wondering how patent claims relating to medical tests should be treated.[238] In *Metabolite*, the patentee discovered that an elevated homocysteine level was an indicator of a Vitamin B deficiency.[239] The claim at issue involved two steps: first, measure homocysteine levels; second, correlate the results and diagnose a vitamin deficiency if levels are elevated.[240] Metabolite alleged that any laboratory performing homocysteine level measurements, whether or not such measurements were patented, contributorily infringed the claim relating to diagnosing vitamin deficiencies.[241]

Three justices dissented from the dismissal, arguing that certiorari was proper, and that the method claim "amounted to" an unpatentable natural phenomenon.[242] However, invalidating the claim as a phenomenon of nature, as the dissent might have done, would draw a poorly defined line.[243] Even if *Metabolite* clearly involved a natural phenomenon as the dissent asserted, the proposed ruling would have done nothing to aid the PTO, the courts, or inventors as to proper patentable subject matter in the future.[244] Many natural phenomena are simple to apply, both inventively and usefully, once the natural phenomenon is discovered.[245]

---

238. Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 126 S. Ct. 2921 (2006) (per curium).

239. Metabolite Labs, Inc. v. Lab. Corp. of Am. Holdings, 370 F.3d 1354, 1358 (Fed. Cir. 2004), *cert. dismissed*, 126 S. Ct. 2921 (2006).

240. *Id.* at 1358–59.

241. *Id.* Metabolite made this argument pursuant to 35 U.S.C. §§ 271(b) and (c). The patent also included a novel claim to a method for performing the homocysteine measurement, but this claim was not at issue. *Metabolite*, 370 F.3d at 1365.

242. *Metabolite*, 126 S. Ct. at 2922, 2927.

243. *Id.* at 2926 ("I concede that the category of non patentable '[p]henomena of nature,' like the categories of 'mental processes,' and 'abstract intellectual concepts,' is not easy to define.") (Breyer, J., dissenting) (citing Parker v. Flook, 437 U.S. 584, 589 (1978)).

244. *See* Kevin Emerson Collins, *Propertizing Thought*, 60 SMU L. REV. 317, 353 (2007) ("Phrased in terms of a preemption analysis as suggested in *Benson*, the argument in favor of the patentability of claim 13 has merit . . . ."); *cf.* Gruner, *supra* note 91, at 400 ("Much of the current uncertainty in the law of patentable subject matter stems from the failure of the Supreme Court to articulate clear principles for separating patentable applications from unpatentable abstract ideas. The Court has, for the most part, dealt with what are essentially easy cases . . . . What the Court's analyses have generally lacked is a clear discussion of what minimum features must be present in order for an implementation of an idea to be considered a practical application rather than just an unpatentable abstract idea.").

245. *See* Corning v. Burden, 56 U.S. 252, 268 (1854) ("As, for instance, A has discovered that by exposing India rubber to a certain degree of heat, in mixture or connection with certain metallic salts, he can produce a valuable product, or manufacture; he is entitled to a patent for his discovery . . . ."); Mackay Radio & Telegraph Co. v. Radio Corp. of Am., 306 U.S. 86, 94 (1939) (radio antenna could be patentable subject matter even though its dimensions directly correspond to a natural phenomenon); Diamond v. Diehr, 450 U.S. 175, 189 n.12 (1981) ("To accept the analysis proffered by the petitioner would, if carried to its extreme, make all inventions unpatentable because

Instead, with rigorous patentability, the analysis should focus on the patentability criteria rather than on nebulous subject matter definitions. Metabolite, the patent assignee, argued that the inventors not only were the first invent a way to measure homocysteines, but also the first to discover a particular method for finding vitamin deficiency by measuring homocysteine levels; they did so long before anyone else had even discovered how to measure homocysteines in the first place.[246] In other words, invention lies not only in the solution, but also in discovery of the specific cause of the problem.[247] Here, the purportedly "simple" solution of correlating homocysteine levels was preceded by more complex problems, discovering how to measure homocysteine levels and then discovering how newly measurable homocysteine levels relate to vitamin deficiencies.[248]

A broad patent claim covering a "simple" solution is not improper, so long as the disclosure describes and enables the broad claim, showing that the inventor truly found the broad but simple solution.[249] Thus, the fact that others later

---

all inventions can be reduced to underlying principles of nature which, once known, make their implementation obvious."); Burk & Lemley, *Inherency*, *supra* note 39, at 406–07 (noting "the problem of characterizing this subject matter category in terms of human intervention); Collins, *supra* note 244, at 353 (noting that "law of nature" analysis does not answer the question posed in *Metabolite).*

246 *Metabolite*, 126 S. Ct. at 2923 (Breyer, J., dissenting); *see* Cochrane v. Deener, 94 U.S. 780, 788 (1877) ("The machinery pointed out as suitable to perform the process may or may not be new or patentable; whilst the process itself may be altogether new, and produce an entirely new result."); *see also* Collins, *supra* note 244, at 333 (difference between claim 1, the homocysteine test, and claim 13, the vitamin deficiency test, was the generality of the method used and not the "natural phenomena").

247. *See, e.g.*, Eibel Process Co. v. Minnesota & Ontario Paper Co., 261 U.S. 45, 68 (1923) ("The invention was not the mere use of a high or substantial pitch to remedy a known source of trouble. It was the discovery of the source not before known and the application of the remedy . . . ."); Varu Chilakamarri, *Structural Nonobviousness: How Inventiveness is Lost in the Discovery*, 10 VA. J.L. & TECH. 7, 11 (2005) (noting the distinction between "invention" and "discovery").

248. Metabolite Labs., Inc. v. Lab. Corp. of Am. Holdings, 370 F.3d 1354, 1358 (Fed. Cir. 2004).

249. Parke-Davis & Co. v. H.K. Mulford Co., 189 F. 95, 102 (C.C.S.D.N.Y. 1911) ("There is nothing improper, so far as I can see, in first putting your claims as broadly as in good faith you can, and then, *ex abundanti cautela*, following them successively with narrower claims designed to protect you against possible anticipations of which you are not yet aware."). *But see* Michael Meehan, *The Handiwork of Nature: Patentable Subject Matter and Laboratory Corporation v. Metabolite Labs*, 16 ALB. L.J. SCI. & TECH. 311, 317–22 (2006) (describing several diagnostic tests that could have been patentable under Federal Circuit's decision in *Metabolite*). Two of the three examples by Dr. Meehan differ from *Metabolite* in that (a) the diagnoses are easily determined through visual inspection, and (b) do not require analysis by a laboratory. *Id.* Congress has made clear that doctors cannot be held liable for infringing such patents. 35 U.S.C. § 287(c)(1) (2000). As such, to say categorically that diagnostic measures should not be patented through these examples is unpersuasive. Additionally, the examples Dr. Meehan describes relate to enablement and novelty – did the person who claimed a diagnostic test really invent *all* such incarnations of that test, or instead just a particular one. *See, e.g.*, Incandescent Lamp Patent, 159 U.S. 465, 475 (1895).

discover different ways to measure homocysteine levels does not necessarily mean that the Metabolite inventors did not inventively solve a different but related problem. Similarly, in *Arrhythmia Research Technology, Inc. v. Corazonix Corp.*, the Federal Circuit ruled that a process for analyzing electrocardiograph data was patentable subject matter even though the relationship between the data and health was "natural."[250] The Federal Circuit recently expounded on this rule, stating that if a process is tied to a machine or transforms physical objects or data about physical objects then it does not preempt a "fundamental principle."[251]

In this sense, the *Metabolite* claim is like any patent claim covering a new use for a known composition or process.[252] A new use patent claims the natural phenomenon that a medicine has a certain effect on the body (or, as in *Metabolite*, that certain test process results reflect a certain condition), and the patentee is the first to discover the previously unknown effect.[253] Furthermore, the *Metabolite* test transforms blood to perform a homocysteine test and manipulates data about that physical phenomenon to determine whether there is a vitamin deficiency.

Nonetheless, patentability is not ensured. Strict application of specification requirements under *Morse* might have invalidated the *Metabolite* patent on the basis that the inventors did not "possess"[254] *all* homocysteine correlation tests and did not describe or enable such in the specification.[255] The Federal Circuit did not consider the broad scope on § 112 enablement grounds; it focused on the "correlating" step rather than the use of *any* test, whether or not invented by the applicants.[256] In fact, Laboratory Corp. of America did not even make the

---

250. Arrhythmia Research Tech., Inc. v. Corazonix Corp., 958 F.2d 1053, 1058–60 (Fed. Cir. 1992) ("Arrhythmia Research argues that the claims are directed to a method of detection of a certain heart condition by a novel method of analyzing a portion of the electrocardiographically measured heart cycle.").

251. *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *7, 12 (Fed. Cir. Oct. 30, 2008).

252. 35 U.S.C. § 100(b) (2000) ("The term 'process' means process, art or method, and includes a new use of a known process, machine, manufacture, composition of matter, or material."); *see, e.g.*, Rohm & Haas Co. v. Roberts Chems., Inc., 245 F.2d 693, 697 (4th Cir. 1957) (upholding patent for new use of a chemical as a fungicide).

253. 35 U.S.C. § 100(b) ("The term 'process' . . . includes a new use of a known process, machine, manufacture, composition of matter, or material.").

254. LizardTech, Inc. v. Earth Res. Mapping, Inc., 424 F.3d 1336, 1345 (Fed. Cir. 2005) (holding that the specification "must describe the invention sufficiently to convey to a person of skill in the art that the patentee had possession of the claimed invention at the time of the application, i.e., that the patentee invented what is claimed") (citing O'Reilly v. Morse, 56 U.S. 62, 112–13 (1853))).

255. The Federal Circuit opinion notes that the PTO rejected an initial attempt by the applicants to patent "[a] method for detecting a deficiency" by assaying a body fluid for elevated levels of homocysteines for a failure to describe the method. Metabolite Labs., Inc. v. Lab. Corp. of Am. Holdings, 370 F.3d 1354, 1362 (Fed. Cir. 2004) (quoting from the prosecution history of U.S. Patent No. 4,940,658 (filed Nov. 20, 1986)).

256. *Id.* at 1366–67.

argument that the claim to the use of *any* test was broader than enabled.[257] As discussed below, one harm caused by inordinate focus on subject matter is that such focus detracts from rigorous consideration of patentability criteria.

Regardless of how one would resolve the question of whether the inventors described and enabled such a broad claim,[258] a specification question should be answered in place of an unprincipled and potentially unanswerable question of patentable subject matter.

F. Mental Steps and Human Action

More than 35 years ago, the courts stopped barring patent claims simply because they included steps that could be performed by a human,[259] but this area remains controversial.[260] The Federal Circuit recently reinvigorated the doctrine in *In re Comiskey*, ruling that patent claims based *solely* on human thought processes are not patentable subject matter.[261] This ruling was extended under *Bilski's* machine or transformation test: mental processes not tied to machines or transformative of matter are barred.[262] Even so, there is little principled discussion in the literature or in case law about when and how "mental steps" should be allowed in patents.[263]

---

257. Brief for Petitioner at 38–52, Metabolite Labs., Inc. v. Lab. Corp. of Am. Holdings, 370 F.3d 1354 (Fed. Cir. 2004) (No. 03-1120).

258. *Cf.* Collins, *supra* note 244, at 331–32 ("The recitation of an act of thinking is harmless to the public when that act has been appended onto an otherwise patentable method claim. In this situation, the thinking merely restricts the scope of a patentee's right to exclude. . . . More specifically, a claim is exempted from thought-propertizing status if the steps other than the acts of thinking recite a novel, nonobvious, and useful method.").

259. *In re* Musgrave, 431 F.2d 882, 893 (C.C.P.A. 1970) ("We cannot agree with the board that these claims (all the steps of which can be carried out by the disclosed apparatus) are directed to non-statutory processes merely because some or all the steps therein can also be carried out in or with the aid of the human mind or because it may be necessary for one performing the processes to think."); *see* Collins, *supra* note 244, at 321 ("However, the courts abandoned the mental steps doctrine over a quarter-century ago, and the doctrine was notoriously ill-defined and under-theorized even in its heyday."); *id.* at 355–57 (summarizing the history of the mental steps doctrine).

260. *See, e.g.*, *In re* Schrader, 22 F.3d 290, 291 (Fed. Cir. 1994) (holding that auction process is not patentable subject matter); Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 126 S. Ct. 2921, 2923 (2006) (Breyer, J., dissenting) (asserting that mental processes are not patentable subject matter); *see also In re* Comiskey, 499 F.3d 1365, 1376 n.11 (Fed. Cir. 2007) (arguing that Supreme Court's decision in *Gottschalk v. Benson* undercut *Musgrave*). Even the court in *Musgrave* required that any mental steps be part of "technological arts." *Musgrave*, 431 F.2d at 893.

261. *In re* Comiskey, 499 F.3d 1365, 1376 (Fed. Cir. 2007).

262. *In re* Bilski, 2007-1130, 2008 WL 4757110, at *12 (Fed. Cir. Oct. 30, 2008).

263. Collins, *supra* note 244, at 344 ("The mess that resulted from the Supreme Court proceedings in Laboratory Corp. demonstrates that there is no well-established approach for bringing Section 101 and its restriction on the subject matters eligible for patent protection to bear on the propertization of thought.").

Primary concerns with mental steps are that human intervention fails the definiteness,[264] usefulness[265] or nonobviousness tests.[266] Such concerns, however, do not mean that every invention that involves human thought fails to meet these tests—each claim can be tested for definiteness, usefulness, or nonobviousness independently.[267] In fact, virtually every method requires human intervention at some point, if only to push a button on a machine that will carry out the method.[268]

Other theories for barring protection of mental steps might be advanced. For example, some might claim that protecting mental processes would limit free speech and thought, conflicting with the First Amendment.[269] However, to the extent that people may think about and discuss—but not practice—the contents of a patent, the First Amendment seems unlikely to be implicated. Further, the Intellectual Property Clause is constitutional as well.[270] To the extent that a law is constitutional under the Intellectual Property Clause, First Amendment protection is lessened.[271] For example, copyright law limits speech by barring the distribution of copyrighted works, but arguments that such a bar is unconstitutional *per se* are unpersuasive.[272]

---

264. Norman D. McClaskey, *The Mental Process Doctrine: Its Origin, Legal Basis, and Scope*, 55 IOWA L. REV. 1148, 1165–69, 1195 (1969) (discussing cases relating to indefiniteness of human activity).

265. *In re* Comiskey, 499 F.3d at 1376 ("[W]hen an abstract concept has no claimed practical application, it is not patentable.").

266. Thomas F. Cotter, *A Burkean Perspective on Patent Eligibility*, 22 BERKELEY TECH. L.J. 855, 886 (2007) ("[I]f technological arts and mental steps are to perform a modest but non-negligible function in preventing patents from intruding upon liberty and other important interests, it might be more fruitful to reconsider application of a point of novelty approach.").

267. *In re* Musgrave, 431 F.2d 882, 893 (C.C.P.A. 1970) ("Of course, to obtain a valid patent the claim must also comply with all the other provisions of the statute, including definiteness under 35 USC § 112. A step requiring the exercise of subjective judgment without restriction might be objectionable as rendering a claim indefinite, but this would provide no statutory basis for a rejection under 35 USC § 101."); McClaskey, *supra* note 264, at 1151–52 (asserting that *O'Reilly v. Morse* stands for the proposition that any useful art— including mental steps—is patentable, so long as it may be described in a definite manner and so long as it leads to predictable results).

268. *See, e.g.*, Wright Co. v. Paulhan, 177 F. 261, 264 (C.C.S.D.N.Y 1910), *rev'd on other grounds*, 180 F. 112 (2d Cir. 1910). In *Wright*, the patent claimed a system of ropes and pulleys to automatically adjust the tail rudder in response to wing "warping." *Id.* at 264. Competitors discovered that they could design around the patent if the pilot performed the rudder adjustment manually. *Id.* Judge Learned Hand ruled that the substitution of a person instead of the automatic system was an equivalent and thus infringing. *Id.* at 264.; *see also* Merges & Duffy, *supra* note 162, at 821–25. *But see* Collins, *supra* note 244, at 329–30 (distinguishing human activity from human thought). Even Collins's test, which asks whether "thought" is a necessary element of a method or merely an additional but non-critical step, would be difficult to administer in practice. *Id.*

269. *See* U.S. CONST. amend I.

270. U.S. CONST. art. I, § 8, cl. 8.

271. Roth v. United States, 354 U.S. 476, 492 (1957) (stating that the postal power provides wider latitude relating to First Amendment).

272. Harper & Row, Publishers v. Nation Enters., 471 U.S. 539, 556–57 (1985) ("But

Others might argue that protection of mental steps allows for protection of information that the patent laws intended to dedicate to the public.[273] However, it is difficult to square any definition of "thought" with any statutory intention discernable in the categories in § 101. Additionally, the public dedication theory specifically embraces a "point of novelty" approach, where thought is the point of novelty.[274] However, the "point of novelty" argument has been expressly rejected even where the Court has seemed to apply such analysis.[275] As such, reliance on point of novelty, however elegant for mental steps theory, is unlikely to yield consistent results in practice, which is a goal of this Article's proposal.

The *Metabolite* case brought newfound attention to the mental steps issue because any doctor could determine the correlation between homocysteine levels and vitamin deficiencies, nearly automatically.[276] Applying a mental steps subject matter test, however, draws the wrong lines on patentability. For example, no mental step would be necessary if the method instead claimed an electronic test that flashed a "vitamin deficiency" sign if homocysteines exceeded a particular level (similar to a pregnancy test).[277] Further, the *Bilski* machine or transformation test is of little help; drawing and testing blood is a transformation and machines are most certainly used to perform the tests. Thus, determination of the claim's patentability should not hinge on whether the mental step was (or could be) carried out by a device.[278] Furthermore, with a binary test the mental process is surely definite and not prone to variation.[279]

Following *Metabolite*, *In re Comiskey* reasserted the mental steps exclusion in cases where the *entire* claim can be performed by the human mind, but even

---

copyright assures those who write and publish factual narratives . . . that they may at least enjoy the right to market the original expression contained therein as just compensation for their investment.").

273. *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *3, 15 (Fed. Cir. Oct. 30, 2008) (noting that mental steps fall under the bar against phenomena of nature and abstract ideas); Collins, *supra* note 244, at 357–60 (arguing that propertizing "thought" allows for removal of too much from the public domain).

274. Collins, *supra* note 244, at 357; Bilski, 2008 WL 4757110, at *8 (quoting Diamond v. Diehr, 450 U.S. 175, 191–92 (1981)) (noting that "insignificant post-solution activity" is to be disregarded in determining whether a claim is tied to a machine).

275. Parker v. Flook, 437 U.S. 584, 594 n.16 (1978).

276. Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 126 S. Ct. 2921, 2924 (2006) ("Hence, in reviewing the test results, doctors would look at the [test results] and automatically reach a conclusion about whether or not a person was suffering from a vitamin deficiency."). *But see* Collins, *supra* note 244, at 321 (noting that courts and commentators did not focus on mental steps aspect of *Metabolite* case).

277. In fact, the claim at issue could have been performed by a machine. *See Metabolite*, 126 S. Ct. at 2924 (stating that the inventor's claim required a "correlating" step but that the correlation was nothing more than a binary process; the patient either had or did not have a vitamin deficiency based upon his or her homocysteine levels).

278. *But see* Collins, *supra* note 244, at 330–31 (holding that "thought" steps that can be performed by a machine should still be considered thought steps).

279. The answer would be "yes" or "no."

that definition is problematic.[280] The claims at issue in *Comiskey* included enrollment of a person and "unilateral documents" in a mandatory arbitration "system" which incorporated "conducting" arbitration.[281] The document step is certainly not mental, and conducting an arbitration must include some human interaction.[282] While the *Comiskey* court dismissed these objections with little discussion, the case provides little future guidance for determining which claims are *entirely* mentally performed.[283]

The Federal Circuit attempts to shore up these failings in *Bilski* by recasting *Comiskey* under the "machine or transformation" standard.[284] The recasting adds little certainty, as the "non-mental" aspects of *Comiskey* are now disregarded because they do not recite a machine or transformation.[285] This too draws a poor line—Comiskey might simply claim that the agreements are stored on a computer to recite a machine. Determining whether such a claim is "insignificant post-solution activity"[286] is no more predictable than determinations under any other test. Instead of having to be re-explained mere months after its issuance, *Comiskey* could have been (and indeed was originally by the PTO) decided on obviousness grounds.[287] Because it cannot be consistently applied, the mental steps doctrine should remain unused in lieu of other patentability criteria.[288]

G. Signals

The recent case of *In re Nuijten* brought further attention to patentable subject matter.[289] The applicant sought to patent not only the means for creating and using a signal,[290] but also the signal itself divorced from any tangible

---

280. *In re* Comiskey, 499 F.3d 1365, 1379 (Fed. Cir. 2007).

281. *Id.* at 1368–69.

282. *See id.* at 1379.

283. *Id.* ("Comiskey's independent claims 1 and 32 seek to patent the use of human intelligence in and of itself.").

284. *In re* Bilski, 2008 WL 4757110, at *10 ("[W]e actually applied the machine-or-transformation test to determine whether various claims at issue were drawn to patent[-]eligible subject matter.").

285. *Id.* at *10 ("As a result, even a claim that recites 'physical steps' but neither recites a particular machine or apparatus, nor transforms any article into a different state or thing, is not drawn to patent-eligible subject matter.").

286. *Id.* at *8 (quoting Diamond v. Diehr, 450 U.S. 175, 191–92 (1981)).

287. *Comiskey*, 499 F.3d at 1370 (PTO rejected claims as obvious).

288. Part IV discusses why particular subject matter bars do not fall under reasonable statutory interpretation. However, interpretation of "process" to exclude claims where every step can *only* be performed in the mind or with pencil and paper (and faithful application of that interpretation, which was missing in *Comiskey*) might be a reasonable reading of 35 U.S.C. §100(b). Such an interpretation would also exclude "pure" mathematical algorithms divorced from any other steps. The "machine or transformation" test of *Bilski* does not provide a similarly grounded test.

289. *In re* Nuijten, 500 F.3d 1346, 1348 (Fed. Cir. 2007).

290. *Id.* at 1348 ("Nuijten's patent application discloses a technique for reducing distortion induced by the introduction of 'watermarks' into signals. In the context of signal processing,

medium.[291] The reason the applicant cared about patentability of the signal was so that the patent owner could sue not only on the senders of such signals for infringement, but also the carriers of such signals, such as internet service providers, even though they might unknowingly carry the signal.[292]

Some have commented that signals should be yet another specific unpatentable subject matter exception.[293] A problem with this approach is that many patent claims are in some sense related to ordered information.[294] Others have focused on the statutory requirements for patentable categories.[295] This is how the *Nuijten* Court proceeded, ruling that a signal does not fit one of the statutory categories.[296]

Nuijten's rigorous categorical analysis is more in line with this paper but is still difficult to apply to abstract ordered information, such as signals.[297] In *O'Reilly v. Morse*, the Supreme Court upheld a claim for "the system of signs . . . in combination with machinery for recording them, as signals for telegraphic purposes."[298] The *Nuijten* signal could very well fall in the category of dots and

---

watermarking is a technique by which an original signal (such as a digital audio file) is manipulated so as to embed within it additional data.").

291. *Id.* at 1350–51. In broadest terms, a signal like Nuijten's is a quantum of information organized in a specific way but separated from any physical medium. *In re* Foster, 438 F.2d 1011, 1016 (C.C.P.A. 1971) (explaining that a signal is "[a] visual, aural, or other indication used to convey information" or "[a]n event or occurrence that transmits information from one location to another.") (citations omitted); *see, e.g.*, Sam S. Han, *Analyzing the Patentability of "Intangible" Yet "Physical" Subject Matter*, 3 COLUM. SCI. & TECH. L. REV. 2, 55 (2002) ("Webster's dictionary provides that a 'signal' is: in radio, etc. the electrical impulses transmitted or received. Although that definition only provides for 'electrical impulses,' other types of signals may be encompassed in our analysis (e.g., magnetic impulses, continuous waves, etc.)." (footnote omitted)).

292. *See Nujiten*, 500 F.3d at 1353 (stating that "any tangible means of information carriage will suffice for all of the claims at issue"). It is not clear that such a theory would work in practice, as one must "use[]" the invention to be liable. 35 U.S.C § 271(a).

293. Cotter, *supra* note 266, at 872 n.100 (describing *Nuijten* as pushing the limits of Federal Circuit subject matter jurisprudence).

294. Kevin Emerson Collins, *Claims to Information Qua Information and a Structural Theory of Section 101*, 4 I/S: A JOURNAL OF LAW AND POLICY 11, 26–29 (2008) (discussing the difficulty of determining what is and is not information, and the further difficulty of determining which information is patentable and which is not).

295. John F. Duffy, *Nuijten: Patentable Subject Matter, Textualism and the Supreme Court*, PATENTLY-O (Feb. 5, 2007) http://patentlyo.com/patent/2007/02/in_re_nuijten_p.html (discussing statutory categories and applying historical meaning of terms to determine whether signals are "compositions of matter"); Han, *supra* note 291, at 56 (embodiment of a signal in a tangible medium creates patentable subject matter).

296. *Nuijten*, 500 F.3d at 1357.

297. *See, e.g.*, Han, *supra* note 291, at 65–67 (arguing that signals fall within statutory category of § 101 and that patentability should be determined by other patentability criteria). Han relies only on *method* claims that *use* signals to support his argument that a signal, standing alone, is statutory subject matter; he does not rely on any case holding that a signal is actually a composition of matter, for example. *Id.*; *see also* Osenga, *supra* note 5, at 1111–12 (noting the same confusion).

298. O'Reilly v. Morse, 56 U.S. 62, 86 (1853); *see also Nuijten*, 500 F.3d at 1353 (signals are

dashes claimed by Morse. On the other hand, the claim at issue in *Morse* covered a system of signals combined with specific machinery.[299] In other words, Morse's claim was for a particular method of using a signal and not for the signals themselves transmitted by other media.[300]

Another way to consider signals is through the application of requirements for patentability other than category. Fundamentally, a signal is information. The question, then, is whether a particular combination of information can be novel and nonobvious; the apparent answer is no. Under the rationale of *Funk Brothers*, the combination of known elements into something that is not more than the sum of parts cannot be novel and nonobvious.[301] Further, such signals—in the abstract at least—have no practical utility; they fail to "do" anything until coupled with a storage medium or process.[302] Under this analysis, a signal should not be patentable on grounds unrelated to subject matter.

H. Books, Art, and Music

Books, art, music, and pictures are extensions of signals: they are other forms of "ordered information." Such works are generally not patentable for a variety of reasons.[303]

First, books, art, and music are not processes, compositions of matter, or machines.[304] Books are a manufacture, but art and music stretch the interpretation of manufacturing, which requires raw materials to take a new form.[305] Further, even though art takes a new form when paint is combined on a canvas, and

---

physical); Arrhythmia Research Tech., Inc. v. Corazonix Corp., 958 F.2d 1053, 1059 (Fed. Cir. 1992) ("The view that 'there is nothing necessarily physical about 'signals' is incorrect." (quoting *In re* Taner, 681 F.2d 787, 790 (C.C.P.A. 1982))); U.S. Patent No. 5,568,202 (filed Sept. 22, 1992) (in which inventor Koo claimed "[a]n electronic reference signal in a system for minimizing the effects of ghosts occurring during the transmission and reception of a television signal over a communications path . . ."). Both Morse and Koo claimed signals as part of a specific system that used the signals in a particular way rather than the signals themselves. *Morse*, 56 U.S. at 86; U.S. Patent No. 5,568,202 (filed Sept. 22, 1992).

299. *See Morse*, 56 U.S. at 86.

300. *See id.*

301. Funk Bros. Seed Co. v. Kalo Inoculant Co., 332 U.S. 755 (1947).

302. *In re* Nuijten, 500 F.3d 1346, 1365 (Fed. Cir. 2007) (Linn, J., dissenting) (unless abstract information is applied, such information is not useful).

303. In addition to the reasons discussed below, constitutional limitations may limit the patenting of such "writings." *See infra* Part IV.

304. MANUAL OF PATENT EXAMINING PROCEDURE, § 2106.01 (2007) ("Certain types of descriptive material, such as music, literature, art, photographs, and mere arrangements or compilations of facts or data, without any functional interrelationship is not a process, machine, manufacture, or composition of matter."). Of course, "performing music" or "showing art" might be a method for entertaining. Similarly, "blowing into a flute" might be a method for making sounds.

305. Am. Fruit Growers, Inc. v. Brogdex Co., 283 U.S. 1, 11–12 (1931) (fruit dipped in borax does not take a new form).

photographs are printed from film, only the *first* piece of art and the *first* photograph were novel manufactures.[306]

Second, such copyrightable subject matter would often be obvious under *Funk Brothers* because it is made of preexisting materials and information without any new effects based on the combination.[307] Under this analysis, textual writing in the abstract, no matter how creative, is not patentable because it is a combination of known letters and words and thus, obvious.[308] Furthermore, ordered information can be viewed as a set of instructions and finding defendants liable for transmission of such instructions would make little sense[309]—it would be infringement to read, copy, or transmit the patent document itself.[310]

Third, such works would not be "practically useful" under *Brenner v. Manson* because their sole use would be for visual or auditory examination.[311] In *Brenner*, the Court determined that a process for making steroids lacked utility because the resulting steroid had no known practical use beyond further study.[312] Like the steroid in *Brenner*, visual and auditory art does not *do* anything and is only useful for static viewing.[313] It may be that the first book was a novel, nonobvious, and useful medium used to convey information, but successors that

---

306. *See In re* Gulack, 703 F.2d 1381, 1384 (Fed. Cir. 1983) (affirming ruling that a hatband with material printed on it was statutory subject matter as a manufacture, but only patentable if writing made it novel and nonobvious).

307. Funk Bros. Seed Co. v. Kalo Inoculant Co., 333 U.S. 127 (1948); Great Atl. & Pac. Tea Co. v. Supermarket Equip. Corp., 340 U.S. 147, 152–153 (1950); *see also In re* Ngai, 367 F.3d 1336, 1339 (Fed. Cir. 2004) (printed matter must have some functional relationship to whatever the information is printed on in order to create a "new" product); *Gulack*, 703 F.2d at 1385 ("Where the printed matter is not functionally related to the substrate, the printed matter will not distinguish the invention from the prior art in terms of patentability.").

308. A new language, however, might be patentable as a communication method if the symbols therein are nonobvious. Morse code is an example of such a language, though such symbols were tied to the particular telegraph hardware.

309. Collins, *supra* note 244, at 318 (patent law makes instructions part of public domain).

310. *Cf.* Microsoft Corp. v. AT&T Corp., 127 S. Ct. 1746, 1755 (2007) ("[Software] abstracted from a tangible copy no doubt is information—a detailed set of instructions—and thus might be compared to a blueprint (or anything containing design information, e.g., a schematic, template, or prototype). A blueprint may contain precise instructions for the construction and combination of the components of a patented device, but it is not itself a combinable component of that device."); Pellegrini v. Analog Devices, Inc., 375 F.3d 1113, 1117–19 (Fed. Cir. 2004) (transmission of instructions is not the same thing as transmission of the object the instructions describe).

311. Brenner v. Manson, 383 U.S. 519, 534–35 (1966); *see In re* Nuijten, 500 F.3d 1346, 1365 (Fed. Cir. 2007) (Linn, J. dissenting) (unless abstract information is applied, it is not useful).

312. *Brenner*, 383 U.S. at 520.

313. *Ngai*, 367 F.3d at 1339 (printed matter must have some functional relationship to whatever the information is printed on in order to create a "new" product); Kreiss, *supra* note 62, at 79 (arguing for a "functional" requirement for printed matter). Of course, some forms of art may be mechanical and thus have such practical utility, but the use would be the mechanical structure and not the non-practical viewing value.

change only the information contained therein would not be patentable due to the failure to create a new use.[314]

Art, music, and other copyrightable subject matter, however, could be patentable if they otherwise met the criteria for patentability. Examples include a rain dance that actually produced rain, a method for consistently inducing sleep through the singing of a particular lullaby, a new medium for artistic expression (e.g. holographic technology), a compact disc with novel sounds that could operate a machine, or other useful forms of copyrightable subject matter. Such examples, however, would not bar ordinary, unpatented art and music and thus floodgate concerns are not significant.

IV. Potential Criticism and Concerns

Discarding specialized subject matter restrictions will undoubtedly raise concerns. First, some might argue that subject matter restrictions are constitutionally warranted, or alternatively, might argue that there is no reason to question judicial statutory interpretation of patentable subject matter categories. Second, some might be concerned about the competitive or ethical harm that broadly construed patentable subject matter might create, and as a fallback might argue that there is no good reason to reject the status quo. Third, some might argue that the proposal is unworkable in practice because the statutory criteria do not coincide with the rigorous patentability criteria discussed previously.

A. Constitutional and Statutory Concerns

1. Potential Constitutional Bar

Some concerned with this Article's proposal might contend that allowing patents on all subject matter is unconstitutional.[315] For example, in *Graham v. John Deere Co.*, the Supreme Court stated, "Congress may not authorize the issuance of patents whose effects are to remove existent knowledge from the public domain, or to restrict free access to materials already available."[316] Even if one accepts the statement as a normatively appropriate reading of the constitution,[317] *Graham* does not necessarily apply to patentable subject

314. Signals differ from books in that a signal in a medium can have nonobvious effects that render the end product as something different. For the same reason, the first "electronic book" might be patentable.

315. *See, e.g.*, Kreiss, *supra* note 62, at 58–66 (describing constitutional limits on subject matter); Liivak, *supra* note 159, at 273–74 (U.S. Constitution requires "originality").

316. Graham v. John Deere Co., 383 U.S. 1, 6 (1966).

317. *See, e.g.*, Special Equip. Co. v. Coe, 324 U.S. 370, 378 (1945) (stating that Congress has wide latitude in determining just how to promote the progress); Rite-Hite Corp. v. Kelley Co., 56 F.3d 1538, 1544–45 (Fed. Cir. 1995) (stating that Congress has broad damages authority under constitutional mandate). Indeed, removing pre-existing knowledge from the public domain could

matter.[318] Instead, this quote reads much more like a constitutional requirement for novelty and nonobviousness: that which is in the public domain should not be protected because it does not "promote the progress of [the] useful arts."[319]

In any event, this Article's focus on rigorous patentability complies with *Graham*: if something sought to be patented preexists and is publicly known, then it will not be novel or nonobvious, and a patent should not issue.[320] Even under a theory that every patentability prerequisite must separately "promote the progress,"[321] it is not at all clear that absolute bars to a particular subject matter will separately promote progress.[322]

Finally, one might argue that any protection that is unrelated to the "technological arts" is unconstitutional.[323] The PTO has recently used the "technological arts" limitation as a justification of patentable subject matter rejections.[324] However, no clear precedential consensus mandates that "useful arts"—under which "technological arts" would fall—be so narrowly construed.[325]

---

promote the progress of the "useful arts" depending on the terms of such removal. For example, patents are granted on inventions that others have used only secretly. 35 U.S.C. § 102(g)(2) (2000). The patent grant removes this knowledge from the public domain (even as to the prior users), but the corresponding public disclosure of the invention may very well promote progress generally. Similarly, patenting of natural phenomena might provide an incentive to fund basic research, which would in turn promote the "useful arts."

318. *See, e.g.*, Merges, *supra* note 127, at 587 ("Given a constitutional provision rooted in a blind faith in 'progress,' we cannot read in historically contingent limitations on patentable subject matter. Put simply, there are no plausible subject matter limits, express or implied, in this broad, enabling clause."); *cf.* Chisum, *supra* note 215, at 1011 (arguing that opponents of the particular subject matter have burden of proving that the subject matter falls outside the constitutional mandate).

319. U.S. CONST. art. I, § 8, cl. 8. However, to the extent a product of nature does not qualify as prior art under § 102, something that is not new might be patentable. Such a rule may not necessarily be contrary to promoting progress.

320. *Graham*, 383 U.S. at 6; *see* McKenna, *supra* note 145, at 1253 ("Today, the Patent Act's requirements of utility, novelty, and nonobviousness ensure that the constitutional purpose is met. A narrow reading of the statutory classes of subject matter is unnecessary to meet this constitutional purpose.").

321. U.S. CONST. art. I, § 8, cl. 8. *But see* Cont'l Paper Bag Co. v. E. Paper Bag Co., 210 U.S. 405, 422–23 (1908) (rejecting constitutional argument that failure to exploit patented invention does not promote the progress of useful arts).

322. *See, e.g.*, Chisum, *supra* note 215, at 1015–16 (describing how protection for algorithms might create incentives for innovation); *see also* Parts II, III for examples of innovation despite supposed subject matter bars.

323. *See In re* Musgrave, 431 F.2d 882, 893 (C.C.P.A. 1970) ("All that is necessary, in our view, to make a sequence of operational steps a statutory 'process' within 35 USC 101 is that it be in the technological arts so as to be in consonance with the Constitutional purpose to promote the progress of 'useful arts.'" (citations omitted)).

324. Steven M. Greenberg, *The Inconsistent Treatment of Computer Software as Patentable Subject Matter*, 11 J. TECH. L. POL'Y 77, 88 (2006) (basis for rejections are mental steps and technological arts rejections rather than rejections based on other patentability criteria).

325. *Cf.* Corning v. Burden, 56 U.S. 252, 267 (1854) ("useful art" is a general term). *But see*

For example, in *Jacobs v. Baker*, the Court assumed a broad meaning of "art" as anything that did not fall into the other subject matter categories.[326] In another case, the Court implied that the goals of the constitution are adaptable.[327] Finally, the Court included bookkeeping in the categories of useful arts, along with use of medicine, construction of ploughs, and mixing paints, implying that "useful art" in the constitutional sense is not necessarily "mechanical" or "technological."[328] The brief congressional authorization in the Intellectual Property Clause does not warrant subject matter limits.[329]

2. Statutory Concerns

Even if this Article's proposal passes constitutional muster, another potential criticism is that judicial common law limitations on patentable subject matter are simply a matter of statutory interpretation. For example, critics might wonder why the mental steps exclusion under "process" is improper statutory interpretation,[330] but the "practical utility" interpretation of "useful"[331] is acceptable.

The statutory interpretation concern, however, assumes that courts limiting patentable subject matter are actually performing statutory interpretation. While the *Brenner* Court explicitly noted that it was interpreting the statute,[332] many

---

Paulik v. Rizkalla, 760 F.2d 1270, 1276 (Fed. Cir. 1985) (en banc) ("The exclusive right, constitutionally derived, was for the national purpose of advancing the useful arts – the process today called technological innovation.").

326. Jacobs v. Baker, 74 U.S. 295, 298 (1868) ("But waiving all these difficulties as hypercritical, and assuming the correctness of the positions taken, that whatever is neither a machine, nor a manufacture, nor a composition of matter, must (*ex necessitate*) be 'an art;' that a jail is a thing 'made;' and that the patent is for the '*process* of making it' . . . .").

327. Kendall v. Winsor, 62 U.S. 322, 328 (1858) ("The true policy and ends of the patent laws enacted under this Government are disclosed in [Article I] of the Constitution, the source of all these laws, viz: 'to promote the progress of science and the useful arts,' contemplating and necessarily implying their extension, and increasing adaptation to the uses of society." (quoting U.S. CONST., art. I, § 8, cl. 8)).

328. Baker v. Selden, 101 U.S. 99, 102 (1879) (ruling that copyright cannot protect bookkeeping forms); *cf.* Bleistein v. Donaldson Lithographing Co., 188 U.S. 239, 249 (1903) ("We shall do no more than mention the suggestion that painting and engraving unless for a mechanical end are not among the useful arts, the progress of which Congress is empowered by the Constitution to promote. The Constitution does not limit the useful to that which satisfies immediate bodily needs.").

329. U.S. CONST. art. 1, § 8, cl. 8.

330. *Compare* 35 U.S.C. § 101 (2000) ("any . . . process"), *with In re* Comiskey, 499 F.3d 1365, 1375 (Fed. Cir. 2007) ("Specifically, Supreme Court decisions after the 1952 Patent Act have rejected a 'purely literal reading' of the process provision and emphasized that not every 'process' is patentable.").

331. *Compare* 35 U.S.C. § 101 (2000) ("any . . . useful"), *with* Brenner v. Manson, 383 U.S. 519, 534–35 (1966) (specific and substantial utility required).

332. *Brenner*, 383 U.S. at 532 ("Since we find no specific assistance in the legislative

patentable subject matter opinions simply assume that certain subject matter should not be patentable.[333] Of course, courts do sometimes interpret the statute. In *Brogdex*, the Supreme Court determined that dipping an orange in borax did not create a new manufacture.[334] Similarly, in *Nuijten*, the Federal Circuit considered whether a signal fell into any of the particular categories set forth in § 101 and, in doing so, considered the meanings of each.[335] Such analysis, however, is not the norm; as discussed in Parts II and III, most patentable subject matter decisions were based in large part on the parroting of dicta from prior cases with little or no actual statutory interpretation.[336]

The most recent example is the Federal Circuit's en banc *In re Bilski* decision.[337] The court gives lip service to the notion of statutory interpretation, but then it discards the actual words of the statute in a footnote, with almost no analysis.[338] The court then goes on to apply what it sees as Supreme Court precedent about "fundamental principles" when neither *Benson*, *Flook*, nor *Diehr* used that terminology, and certainly not in the context at issue.[339] Even if those cases had used that terminology, the discussion in Parts II and III shows that Supreme Court pronouncements are not statutorily based either. Finally, the court settled on the "machine or transformation" test as the *only* test, despite the fact that *Benson* and *Flook* explicitly state that the Court was not ruling on that question and no other decision has so held.[340] This is not what statutory

---

materials underlying § 101, we are remitted to an analysis of the problem in light of the general intent of Congress, the purposes of the patent system, and the implications of a decision one way or the other.").

333. *See, e.g.*, Funk Bros. Seed Co. v. Kalo Inoculant Co., 333 U.S. 127, 130–31 (1948) (waxing poetically about how certain natural phenomena cannot be patented, without any explicit interpretation of the statute).

334. Am. Fruit Growers, Inc. v. Brogdex Co., 283 U.S. 1, 11–12 (1931) (considering the meaning of the word "manufacture" in the dictionary and in other statutory contexts).

335. *In re* Nuijten, 500 F.3d 1346, 1354–57 (Fed. Cir. 2007) (considering the meaning of each term and applying to the proposed claim).

336. *See, e.g., In re* Comiskey, 499 F.3d 1365, 1377 (Fed. Cir. 2007) (stating that "mental processes" are not patentable based on list of dicta categories from *Gottschalk v. Benson*); *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *13 (Fed. Cir. Oct. 30, 2008) ("Thus, while we agree with Applicants that the only limit to patent-eligibility imposed by Congress is that the invention fall within one of the four categories enumerated in § 101, we must apply the Supreme Court's test to determine whether a claim to a process is drawn to a statutory 'process' within the meaning of § 101.").

337. No. 2007-1130, 2008 WL 4757110 (Fed. Cir. Oct. 30, 2008).

338. *Id.* at *62 n.3; *see also id.* at *25 (Newman, J., dissenting) ("The definition of 'process' provided at 35 U.S.C. §100(b) is not 'unhelpful,' as this court now states . . . but rather points up the errors in the court's new statutory interpretation. Section 100(b) incorporates the prior usage 'art' and the term 'method,' and places no restriction on the definition. This court's redefinition of 'process' as limiting access to the patent system to those processes that use specific machinery or that transform matter, is contrary to two centuries of statutory definition.").

339. *Id.* at *3 (majority opinion).

340. *Id.* at *6–7.

interpretation, or even faithful interpretation of Supreme Court precedent, is made of.

Instead, application of basic statutory interpretation principles calls into question interpreted limits on patentable subject matter. The section at issue is short:

> Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.[341]

To start, a plain language reading of the statute[342] yields a few insights that support this Article's proposal. First, the term "any" is unambiguous: virtually every definition of the word states that it denotes a quantity without limitation.[343] Thus, any limitation must come from the remainder of the section.[344] The first limitation is "new" and the second limitation is "useful," both of which are described in more detail in Parts II and III.[345]

The next limitations are of category.[346] These terms might very well be vague or ambiguous; as discussed above, some of the category terms have been interpreted and others might need interpretation with respect to specific claims.[347]

However, ambiguity does not mean that the term should be interpreted without reference to possible meanings, statutory definitions, congressional intent, historical precedent, and other bases for statutory interpretation. For example, "process" has an extremely broad statutory definition.[348] Further, "process" generally means a series of definite steps taken to achieve some end.[349] The term "process" was inserted into the statute in replacement of the term "art."[350] As discussed above, "art" was historically considered anything that did

---

341. 35 U.S.C. § 101 (2000).

342. Plain language is a preferred method of statutory interpretation where terms are not ambiguous. Randall v. Loftsgaarden, 478 U.S. 647, 656 (1986) ("Here, as in other contexts, the starting point in construing a statute is the language of the statute itself.").

343. 1 OXFORD ENGLISH DICTIONARY 538–39 (1989) (defining "any" as "[a]n indeterminate derivative of *one*").

344. *See, e.g.*, Roth v. United States, 354 U.S. 476, 479, 484 (1957) (finding that obscenity is exempted from "no law" language of First Amendment because it is not protected speech under the amendment).

345. 35 U.S.C. § 101 (2000).

346. *Id.* (i.e. process, machine, manufacture, or composition of matter).

347. For example, because the constitution separates protection of writings from protection of discoveries, any ambiguities in the categories of § 101 might exclude writings. *Cf.* Baker v. Selden, 101 U.S. 99, 102–03 (1879) (separating writings from the "useful arts" that the writings describe).

348. 35 U.S.C. § 100(b) ("The term "process" means process, art, or method and includes a new use of a known process, machine, manufacture, composition of matter, or material.").

349. 12 OXFORD ENGLISH DICTIONARY 545–48 (1989) (defining "process" as "an action or series of actions; progress, course").

350. *In re* Schrader, 22 F.3d 290, 295 n.11 (Fed. Cir. 1994).

not fit into one of the other categories.[351] It is true that early case law discussed a process in terms of the physical, but none of the cases examined any dictionaries, statutory definitions, congressional intentions, patent policies, or other principled bases for interpretation.[352] Unlike the interpretation of "manufacture" in *Brogdex*, sweeping limiting interpretations of "process" and other statutory categories have generally parroted dicta.[353] However, close analysis shows that "process" was never intended to be limited to physical processes.[354]

No apparent statutory basis exists, therefore, to exclude a business method or even a mathematical algorithm from the process category based on an arbitrary test. The Supreme Court recognized this even as it struck down patents on mathematical algorithms; it did so on grounds other than a narrow interpretation of "process."[355] As discussed in Part II, most nominal subject matter decisions were not really about patentable subject matter; however, to the extent that courts

---

351. Jacobs v. Baker, 74 U.S. 295, 298 (1868). *But see* Sean M. O'Connor, *Using Insights From the History of Science to Redefine Patentable Subject Matter under the IP Clause* 14 (Oct. 8, 2007), *available at* http://ssrn.com/abstract=1104899 ("art" historically considered broad in history of science, but ultimately arguing that "useful" arts should be understood much more narrowly).

352. Cochrane v. Deener, 94 U.S. 780, 788 (1877) ("A process is a mode of treatment of certain materials to produce a given result. It is an act, or a series of acts, performed upon the subject-matter to be transformed and reduced to a different state or thing.").

353. *See, e.g.*, *In re* Comiskey, 499 F.3d 1365, 1376 (Fed. Cir. 2007) ("In that context, the Supreme Court has held that a claim reciting an algorithm or abstract idea can state statutory subject matter only if, as employed in the process, it is embodied in, operates on, transforms, or otherwise involves another class of statutory subject matter, i.e., a machine, manufacture, or composition of matter.").

354. *Schrader*, 22 F.3d at 295 n.12 ("subject matter" in *Cochrane* was not limited to physical transformation, and transformation of "intangibles" is also statutory subject matter, or else the method used by the telephone would not have been patentable). *Comiskey* does not discuss this aspect of *Schrader*. *See Comiskey*, 499 F.3d 1365; *see also In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *11 (Fed. Cir. Oct. 30, 2008) (transformation of data about physical objects is sufficient transformation).

355. Parker v. Flook, 437 U.S. 584, 588 n.9 (1978) ("The statutory definition of 'process' is broad. An argument can be made, however, that this Court has only recognized a process as within the statutory definition when it either was tied to a particular apparatus or operated to change materials to a 'different state or thing.' . . . [W]e assume that a valid process patent may issue even if it does not meet one of these qualifications of our earlier precedents." (citations omitted)); *see also* Gottschalk v. Benson, 409 U.S. 63, 71 (1972) ("It is argued that a process patent must either be tied to a particular machine or apparatus or must operate to change articles or materials to a 'different state or thing.' We do not hold that no process patent could ever qualify if it did not meet the requirements of our prior precedents. It is said that the decision precludes a patent for any program servicing a computer. We do not so hold. It is said that we have before us a program for a digital computer but extend our holding to programs for analog computers. We have, however, made clear from the start that we deal with a program only for digital computers. It is said we freeze process patents to old technologies, leaving no room for the revelations of the new, onrushing technology. Such is not our purpose.").

rendered such decisions on the basis of subject matter limitations, the courts did not conduct statutory interpretation of the § 101 categories.

The final limitation, "subject to the conditions and requirements of this title,"[356] supports this Article's proposal. If *any* claimed invention falls into one of the categories, it is patentable, but *only if* it meets the rigorous criteria set forth in the Patent Act. To exclude inventions that otherwise fall into a category and satisfy the criteria of the Patent Act essentially reads those provisions out of the statute, something to be avoided in statutory interpretation.[357]

B. Competitive and Ethical Harm

One potential criticism of this Article's *laissez-faire* approach to patentable subject matter is the risk that more patents will issue that bar the use of "fundamental truths," and that such patents will lead to anti-competitive or unethical results. However, these concerns can be addressed. First, evidence shows that competition would not be harmed, and the judiciary is not in a position to avoid harm in any event. Second, there are strong policy reasons to adopt the proposal.

1. Competition Would Not be Harmed

Concern about the effect of broad subject matter patentability can be allayed in two ways: (a) such concern is empirically unsupported, and (b) the judiciary is poorly equipped to address subject matter policy.

First, whether unwarranted growth in the number of patents will occur is unclear.[358] Careful claims drafting can avoid many subject matter limitations, which means that focus on subject matter rather than on the underlying invention should have little effect on unwarranted patent claims.[359] Not surprising, then, is the fact that evidence does not indicate excessive growth in patenting of "suspect" subject matter.[360] For example, in the eight years since *State Street Bank*[361] condoned business method patents, only 4% of business method patent applications actually issued as patents.[362] Similarly, in computer software, only

---

356. 35 U.S.C. § 101 (2000).

357. Robinson v. United States, 324 U.S. 282, 285 (1945).

358. *See, e.g.*, Gruner, *supra* note 91, at 429 (arguing that even if subject matter is not considered, patents must still overcome several hurdles before being granted); Lemley & Sampat, *supra* note 126, at 32, 34, 41 (stating that only 13% of patent applications are in software and that only 3% are in business methods).

359. There are exceptions. In *In re* Nuijten, 500 F.3d 1346 (Fed. Cir. 2007), the PTO allowed a patent on a signal in a "storage medium," and the focus on whether a signal in the abstract was patentable subject matter was critical because the patentee wanted to assert patent infringement against those who transmitted, but did not store, the signal. *Id.* at 1351.

360. Lemley & Sampat, *supra* note 126, at 31, 41.

361. State St. Bank & Trust Co. v. Signature Fin. Group, Inc., 149 F.3d 1368 (1998).

362. Lemley & Sampat, *supra* note 126, at 31, 41 (additionally, 52% of business method

51% of applications matured into patents, compared to an overall 69% grant rate.[363] Even these statistics do not address whether the allowed claims are broad or narrow.

While the grant rate in biotechnological and chemical subject matter is 60%,[364] only 4% of all applications are in biotechnology and organic chemistry.[365] Studies have found that biotechnology patents have resulted in few "anti-commons."[366] Applicants in these areas may be cognizant of the prior art—which is more easily discernable than software and business method patents—resulting in patent applications that are filed on specific and discrete inventions that carefully avoid the prior art. Furthermore, as discussed above, a rigorous application of the patent rules should lead to fewer issued patents in the biotechnology area.[367]

Of course, the contrary concern may apply—rigorous patentability may result in too few patents issuing, especially in biotechnology research, which might be costly even if such research does not yield patentable results.[368] This question is commercially important.[369] For example, isolating a protein from its natural

---

applications are abandoned); *see also* John R. Allison & Emerson H. Tiller, *The Business Method Patent Myth*, 18 BERKELEY TECH. L.J. 987, 1081 (2003) (finding that business method patents show indicators of high quality, such as several prior art references).

363. Lemley & Sampat, *supra* note 126, at 32. *But see* Greenberg, *supra* note 324, at 88 (asserting that basis for rejections are mental steps and technological arts rejections rather than rejections based on other patentability criteria).

364. Lemley & Sampat, *supra* note 126, at 30.

365. *Id.* at 27.

366. Timothy Caulfield et al., *Evidence and Anecdotes: an Analysis of Human Gene Patenting Controversies*, 24 NATURE BIOTECHNOLOGY 1091, 1092–93 (2006) (summarizing studies and arguing that lack of access is more related to market price and terms other than exclusion: "The empirical research suggests that the fears of widespread anticommon effects that block the use of upstream discoveries have largely not materialized."); Christopher M. Holman, *The Impact of Human Gene Patents on Innovation and Access: A Survey of Human Gene Patent Litigation*, 76 UMKC L. REV. 295, 318, 352 (2007) (noting that many "gene" patents do not actually claim genes *per se*, and finding that, to date, enforcement of human gene patents does not appear to have a had substantial negative impact on innovation or access to gene-based technologies); *see also* Caulfield, *supra* note 366, at 1092. Caulfield noted that gene patents relating to diagnostic testing have been more exclusive. *Id.* Whether "sole source" availability of genetic testing is harmful is unclear. *Id.* at 1092–93. So long as new therapies are being researched, Congress and not the courts should evaluate the evidence and legislate accordingly. John P. Walsh et al., *Where Excludability Matters: Material Versus Intellectual Property in Academic Biomedical Research*, 36 RESEARCH POLICY 1184, 1199–1201 (2007) (concluding that despite the existence of patents, intellectual property rights have little effect on research and finding the inability to obtain physical materials is a bigger hindrance to research).

367. *Ex parte* Kubin, Appeal 2007-0819, 83 U.S.P.Q.2D (BNA) 1377 (B.P.A.I. May 31, 2007) (rejecting patent claim as "obvious to try").

368. To the extent that a process is expensive to discover and implement, it is less likely to be found obvious.

369. *But see In re* Fisher, 421 F.3d 1365, 1378 (Fed. Cir. 2005) ("Congress did not intend for these practical implications to affect the determination of whether an invention satisfies the

source can be far more expensive than replicating the protein from a derived cDNA sample.[370] However, this too should not be the province of judicially mandated subject matter rules: patentability is based on invention, not expense.[371] For example, many manufacturing methods may be cheaper than the alternatives, but unless such methods meet the patentability criteria they are not entitled to a patent.[372] A court considering a specific patent cannot consider such facts. Perhaps biotechnology research should be incentivized by one means or another, but such incentives should be narrowly tailored and legislatively mandated.[373]

To generalize, judicial attempts to create subject matter policy in a particular industry would likely be doomed by focus on the trees rather than the forest. First, such attempts would likely not take into account variations within that industry. Second, focus on the subject matter of any particular invention can cause unintended effects in that and related industries.

Thus, neither competitiveness nor ethics should be the province of patentable subject matter determinations by courts[374] or the PTO without an express legislative bar.[375] In addition to recognizing and protecting certain biotechnology patents,[376] Congress has also passed laws that countenance business method

---

requirements set forth in 35 U.S.C. §§ 101, 102, 103, and 112.").

370. Davis, *supra* note 144, at 336.

371. 35 U.S.C. §§ 101–103, 112 (2000).

372. *See* State St. Bank & Trust Co. v. Signature Fin. Group, Inc., 149 F. 3d 1368, 1376 (1998).

373. *See, e.g.*, 35 U.S.C. § 103(b) (2000) (providing for special protection for biotechnology processes). *See also supra* Part III.

374. Parker v. Flook, 437 U.S. 584, 595 (1978) ("Difficult questions of policy concerning the kinds of programs that may be appropriate for patent protection and the form and duration of such protection can be answered by Congress on the basis of current empirical data not equally available to this tribunal."). The *Flook* Court, however, opted for *not* allowing patentability unless Congress took such action. *Id* at 596. *But see* Burk & Lemley, *Policy Levers*, supra note 17, at 1669 (arguing against judicial minimalism: "we think the solution is for the courts to get their decisions right, rather than for them to wash their hands of involvement in the calibration of policy."). This Article argues that courts cannot "get their decisions right" if those decisions are based on broad subject matter rules.

375. Diamond v. Chakrabarty, 447 U.S. 303, 315 (1980) ("Congress has performed its constitutional role in defining patentable subject matter in § 101; we perform ours in construing the language Congress has employed. . . . The subject-matter provisions of the patent law have been cast in broad terms. . . ."); Juicy Whip, Inc. v. Orange Bang, Inc., 185 F.3d 1364, 1368 (Fed. Cir. 1999) ("Of course, Congress is free to declare particular types of inventions unpatentable for a variety of reasons . . . ."); *In re* Fisher, 421 F.3d 1365, 1378 (Fed. Cir. 2005) ("public policy considerations . . . are more appropriately directed to Congress as the legislative branch of government, rather than this court as a judicial body responsible simply for interpreting and applying statutory law."); Chisum, *supra* note 215, at 1011 (opponents of the broad statutory construction have burden of proving that exclusions are necessary); Kiley, *supra* note 7, at 474 ("It is not the job of the patent system to regulate new technologies, but rather to bring them into being, and into view.").

376. 35 U.S.C. § 103(b) (2000) (protecting biotechnological processes, including genetic processes, if the resulting composition, including DNA, is novel and nonobvious).

patents[377] as patentable subject matter. The Court's caution in *Flook* that "[i]t is our duty to construe the patent statutes as they now read, in light of our prior precedents, and we must proceed cautiously when we are asked to extend patent rights into areas wholly unforeseen by Congress," [378] is belied by congressional action that has implicitly accepted that controversial subject matter may be patented.[379] Instead, because Congress has at least acquiesced to broad subject matter patentability of two controversial technologies, courts should be wary of imposing restrictions on the otherwise broad statutory language.[380]

Rather, Congress and other policymakers should study all the evidence,[381] and devise whatever reasonable and narrowly tailored limits[382] to the enforcement of patents are warranted by public policy.[383] Congress may address problematic areas in a variety of creative ways.[384] For example, patents on nuclear weapons

---

377. *Id.* § 273 (limiting damages for pre-existing users of patented business methods).

378. Parker v. Flook, 437 U.S. 584, 596 (1978).

379. *Id.* In fact, as discussed in note 6, *Flook*'s methodology relied on the narrow statutory reading in *Deepsouth Packing Co. v. Laitram Corp.*, 406 U.S. 518, 531 (1972), which was also overturned by Congress.

380. Part IV.A discusses why the statutory interpretation should be broad.

381. *Cf.* Caulfield et al., *supra* note 366, at 1093 ("The survey of policy reports reveals that the Myriad Genetics [breast cancer gene] controversy was used as primary tool for justifying patent reform – thus highlighting the potential of a single high-profile controversy to mobilize . . . policy makers."); Hearing, *supra* note 139 (statement of James Toupin) (discussing tax methods: "So, in terms of whether a patent would make a strategy more available or less available, it is a bit of a trade-off between whether the cost of the license that might be requested outweighs the cost of each tax adviser inventing the same strategy for each client. The second issue is if members of the tax advice community want to establish that certain strategies are well known, they will begin to publish the information about those strategies that they may not have published previously. So, the net effect – it is possible that the net effect of patenting is to make strategies more readily available to the public rather than less.").

382. Burk & Lemley, *Policy Levers*, *supra* note 17, at 1634–35 ("Even if industry-specific patent legislation is legal, we are not persuaded that it is a good idea . . . . [W]e are skeptical of the ability of a statute to dictate in detail the right patent rules for each industry. Many of the predictions of economic theory are fact-specific – they suggest different factors that should bear on the outcome of particular cases, but that require case-by-case application that cannot easily be captured in a statute.").

383. Kreiss, *supra* note 62, at 67 ("[T]he Court seems to have foreclosed [congressional investigation and legislation into the effect of subject matter rules on progress] through its pronouncements that laws of nature, natural phenomena, and abstract ideas are not patentable subject matter."); Kiley, *supra* note 7, at 473 (arguing that denial of patents on "products of nature" would be "without regard to the positive good resulting from [such products'] isolation. Patentees profit in general relation to the extent the public profits by their labors. Will there no longer be any profit in sifting the cornucopia of nature?").

384. *See, e.g.*, Drennan, *supra* note 139, at 329 (suggesting that tax patents be allowed, but limiting the types of damages that can be obtained). *But see* Rai, *supra* note 170, at 841–42 (legislature is a poor choice because problems are associated with poor judicial opinions, not the rules themselves). If it is true that the problem is caused by poor judicial implementation of existing law, it is unclear why one might expect the courts to correct themselves while the legislature could

are banned outright,[385] patents that may harm national defense may be kept secret or banned,[386] doctors may not be sued for infringement of certain types of patents,[387] and certain prior users of business methods are exempt from suit.[388] These are all different and reasonable solutions, only one of which includes a specialized subject matter ban.[389]

Further, Congress is not the only authoritative body that can act. For example, the Internal Revenue Service could issue rules relating to the use of patented tax strategies.[390] Also, technology consortiums routinely mandate cross-licensing of issued patents.[391] Additionally, rather than complex judicial rules about the patentability of sports methods, any sports league concerned with "anti-competitive" patents may make rules either banning or mandating compulsory free licensing in order to use a patent method.[392] This is no different than generally accepted sports league rules, such as the rule regulating the size and shape of goaltender gear by the National Hockey League.[393] Regardless of whether such rules are normatively justified, the interested regulatory authority is in a better position to implement the rules it considers best in a way more tailored to the authority's needs than a court would be.

2. Policy Supports Broad Subject Matter Eligibility

Even if potential harm is not an issue, another possible concern is that there is no policy reason to shift from the current system. However, policy dictates the opposite course: limited judicial bars on patentable subject matter because of the current system's costly failure of purpose.

---

not do so through clarifying amendments. Rai does discuss how special interest groups might impede congressional reform. *Id.* Industry capture is a more relevant concern. *See, e.g.*, John Boehner & Roy Blunt, Letter to House Speaker Nancy Pelosi (Aug. 20, 2007) (describing opposition to patent reform bill), *available at* http://www.patentlyo.com/patent/law/RepublicanReform.pdf.

385. 42 U.S.C. § 2181(a) (2000).

386. *Id.* § 181.

387. *Id.* § 287(c).

388. *Id.* § 273.

389. The above provisions may or may not have followed careful study. Of course, such study will lead to better policy outcomes. In all events, courts are not engaged in such study.

390. *See, e.g.*, Dennis Crouch, *Tax Strategy Patents*, PATENTLY-O (Nov. 13, 2007) http://www.patentlyo.com/patent/2007/11/tax-strategy-pa.html.

391. Mark A. Lemley, *Intellectual Property Rights and Standard-Setting Organizations*, 90 CAL. L. REV. 1889, 1904 (2002) (discussing standards setting organizations that require IP disclosure and/or licensing).

392. *See generally* MAJOR LEAGUE BASEBALL, OFFICIAL BASEBALL RULES (2008), http://mlb.mlb.com/mlb/official_info/official_rules/foreword.jsp; NATIONAL HOCKEY LEAGUE, OFFICIAL RULES 2006–07 (2006), *available at* http://cdn.nhl.com/rules/20062007rulebook.pdf.

393. NATIONAL HOCKEY LEAGUE, OFFICIAL RULES 2006–07 25 (2006) *available at* http://cdn.nhl.com/rules/20062007rulebook.pdf (stating rule 11 which limits the size of the goalie's equipment).

Changing the standards required to obtain a patent generally should be disfavored due to settled expectations.[394] Historically, this was not a problem; despite lip service to subject matter limitations, very few patent applications were actually rejected based solely on their subject matter.[395]

Now, however, patentable subject matter is as uncertain as ever. Consider, for example, *Metabolite*.[396] Patentable subject matter was not in dispute: the issue was never explicitly raised until the case reached the Supreme Court.[397] When the Supreme Court granted certiorari on a previously dormant subject matter issue, it injected uncertainty into the analysis and debate began.[398] Indeed, even if the dissent were the majority decision in *Metabolite*, the case would not have settled prospective questions about other new technological subject matters. The dissent admits that determining what is and is not patentable subject matter is extremely difficult.[399] The dissent "decision" would have likely created more uncertainty because it would have invalidated the patent as a so-called "phenomenon of nature" without providing any guidance about how such phenomena should be identified and analyzed.[400]

The Federal Circuit recently attempted to identify natural phenomena with little success by attempting to reconcile *Benson*, *Flook*, and *Diehr* in a new "machine or transformation" test.[401] In doing so, it explicitly disapproved of the "useful, concrete and tangible result" test that had been used in the ten years since *State Street*.[402] Indeed, the court did not even address if or how the patent in *State Street* would have been valid under the new test; under the new *Bilski* test, a process is patentable if it is tied to a machine, but not if the machine is "insignificant post-solution activity."[403] Every attempt to create a new subject matter test has the opposite effect of destabilizing the patentability landscape and unsettling expectations.[404]

---

394. LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW § 9-1 (2d ed. 1988) ("We deal here with the idea that government must respect 'vested rights' in property and contract – that certain settled expectations of a focused and crystallized sort should be secure against governmental disruption, at least without appropriate compensation.").

395. *See supra* Parts II & III.

396. Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc., 126 S. Ct. 2921 (2006).

397. *Id.* at 2925.

398. *Id.* at 2926.

399. *Id.* at 2926–27.

400. *Id.* at 2922.

401. *In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *11 (Fed. Cir. Oct. 30, 2008).

402. *Id.* at *9.

403. *Id.* at *8.

404. *Id.* at *42 (Newman, J. dissenting) ("Not only past expectations, but future hopes, are disrupted by uncertainty as to application of the new restrictions on patent eligibility. For example, the court states that even if a process is 'tied to' a machine or transforms matter, the machine or transformation must impose 'meaningful limits' and cannot constitute 'insignificant extra-solution activity.' . . . . We are advised that transformation must be 'central to the purpose of the claimed process,' . . . although we are not told what kinds of transformations may qualify . . . . These concepts raise new conflicts with precedent.").

Further, *Bilski* raises more questions than it answers: What is a fundamental principle? What is a machine? What does it mean to transform something to a different state? What about processes tied to compositions of matter or manufactures, or even other processes? What about low-tech processes that do not involve machines? How is one to determine whether a claim element is post-solution or insignificant? This last question—that of insignificant post-solution activity—swallows the entire test. The machine or transformation prong becomes irrelevant if it is impossible to determine whether such machine or transformation is significant.

Finally, even if barring preemption of "fundamental principles" were the optimal subject matter rule, the machine or transformation test fails to achieve a systematic resolution of that question. For example, *Bilski* cites to *Mackay Radio & Telegraph Co. v. Radio Corp. of America* for the proposition that an application of a natural principle is patentable.[405] This citation implies that the *Bilski* court approved of the patent in *Mackay Radio* case. The claim at issue in *Mackay Radio* was a radio antenna that implemented wire lengths that identically matched angles and lengths predicted by a well-known equation.[406] The Supreme Court found that this could be patentable because the equation was applied to a structure.[407] This was an apparent triumph of the machine or transformation test because the formula was tied to the antenna.

However, the result is not so clear. First, under *Bilski*, the antenna is a physical item rather than a process, so the court would have to determine whether the physical item is really just a pretext for the mathematical equation such that the test is implicated in the first place. Second, it is not clear that the antenna is a machine—it has no moving parts and performs no actions. *Bilski* implies that *only* a machine may satisfy the test.[408] Third, there could not be a more simple and direct application of this fundamental principle—it is certainly no more complex than the calculations performed by a machine in *Benson*, and there is little question that the antenna at issue pre-empted the fundamental principle associated with the antenna's dimensions.

As such, the machine or transformation test is unclear even in this simple case, and might very well call the implementation of the formula into an antenna

---

405. *Id.* at *4 (majority opinion) (citing Mackay Radio & Tel. Co. v. Radio Corp. of Am., 306 U.S. 86, 94 (1939)).

406. *Mackay Radio*, 306 U.S. at 92–93.

407. *Id.* at 94 ("While a scientific truth, or the mathematical expression of it, is not patentable invention, a novel and useful structure created with the aid of knowledge of scientific truth may be.").

408. *Bilski*, 2008 WL 4757110, at *62 n24 (Rader, J., dissenting) ("Our statement in *Comiskey* that 'a claim reciting an algorithm or abstract idea can state statutory subject matter only if, as employed in the process, it is embodied in, operates on, transforms, or otherwise involves another class of statutory subject matter, i.e., a machine, manufacture, or composition of matter,' . . . was simply a summarization of the Supreme Court's machine-or-transformation test and should not be understood as altering that test."). It is unclear whether *Bilski* is disapproving of the inclusion of manufactures or compositions of matter or not.

"insignificant postsolution activity"[409] despite the fact that the Supreme Court ruled that the antenna could be patented. *Bilski's* announcement of a new test that might (or might not) invalidate a patent claim explicitly allowed by the Supreme Court shows that judicial attempts to shape subject matter tests outside the statutory categories are uncertain and unhelpful.[410]

Because current patentable subject matter standards are in flux,[411] especially given cases like *Bilski* and *Metabolite*, the movement toward more subject matter limitations will impose unwarranted private and social costs[412] without producing any corresponding benefits by allowing fewer "bad" patents.

Like any bright line rule, fixed subject matter rules will lead to both over and under-allowance of bad or good patents respectively.[413] For example, where a subject matter is barred, the incentive to research and invent in a particular area, such as biotechnology, can be significantly reduced.[414] On the other hand, some

---

409. *Id.* at *8 (majority opinion) ("Therefore, even if a claim recites a specific machine or a particular transformation of a specific article, the recited machine or transformation must not constitute mere 'insignificant postsolution activity.'" (quoting *Flook*, 437 U.S. at 590 ('The notion that post-solution activity, no matter how conventional or obvious in itself, can transform an unpatentable principle into a patentable process exalts form over substance.'))).

410. *Id.* at *60 (Rader, J., dissenting) ("An abstract idea must be applied to (transformed into) a practical use before it qualifies for protection. The fine print of Supreme Court opinions conveys nothing more than these basic principles. Yet this court expands (transforms?) some Supreme Court language into rules that defy the Supreme Court's own rule.").

411. *See, e.g.*, *In re* Warmerdam, 33 F.3d 1354, 1359 (Fed. Cir. 1994) ("The difficulty is that there is no clear agreement as to what is a 'mathematical algorithm,' which makes rather dicey the determination of whether the claim as a whole is no more than that. An alternative to creating these arbitrary definitional terms which deviate from those used in the statute may lie simply in returning to the language of the statute and the Supreme Court's basic principles . . . .") (citation omitted); *In re* Musgrave, 431 F.2d 882, 891 (C.C.P.A. 1970) (describing how mental steps doctrine is a "morass"); Eisenberg, *Re-examining*, *supra* note 169, at 784 (noting persistent lack of clarity about patenting of DNA); Osenga, *supra* note 5, at 1093–1103 (describing inconsistent decisions).

412. *But see* Gruner, *supra* note 91, at 398 ("Disputes that still rage over the minimum physical features of patentable inventions . . . miss the point of keeping our patent system general and ensuring that this system encourages the broadest possible range of innovations of benefit to the public."). Rather than adopting the approach of this Article, Gruner later argues that subject matter rules should be determined on a case by case basis using a three-pronged test to maximize patent benefits for that type of technology. *Id.*

413. Michael W. Carroll, *One for All: The Problem of Uniformity Cost in Intellectual Property Law*, 55 AM. U. L. REV. 845, 857 (2006) (discussing the effect of patent scope on incentives); Gruner, supra note 91, at 428 ("The consequences of finding that a particular type of advance falls outside of patentable subject matter are particularly severe. These sorts of advances are never subject to patent rewards and incentives no matter how new and advantageous to society the advances might be.").

414. Robert P. Merges & Richard R. Nelson, *On the Complex Economics of Patent Scope*, 90 COLUM. L. REV. 839, 868–69 (1990) (discussing effect of patent scope on incentives); Gruner, *supra* note 91, at 428 ("The consequences of finding that a particular type of advance falls outside of patentable subject matter are particularly severe. These sorts of advances are never subject to patent rewards and incentives no matter how new and advantageous to society the advances might

might argue that no additional incentives are needed to encourage the development of business methods patents, leading to over-production of such patents if such patents are not barred. However, society might be better off barring certain biotechnology claims despite the negative effect on research. Similarly, society might fare better allowing certain business method claims despite the "positive" effect on future applications for such patents.[415]

In other words, although subject matter restriction can be a "policy lever,"[416] it is not a very effective lever because the rules cannot be applied narrowly or consistently.[417] Because poorly defined and mis-targeted subject matter exclusions cannot effectively answer fine-grained policy questions, resources should instead focus on the statutory patentability criteria. These criteria form a set of requirements developed over time that,[418] in combination, work more like a standard designed to grant only those patents deserving of protection.[419]

To be sure, standards are more costly to implement than rules, but the PTO is already engaged in applying such standards. Indeed, most PTO resources appear to be focused on the statutory patentability criteria because very few rejections are currently based on patentable subject matter.[420]

However, the institution of new rules, like that in *Bilski*, that would cause more subject matter rejections, or even more subject matter analysis,[421] may have

---

be.").

415. Burk & Lemley, *Policy Levers*, *supra* note 17 at 1634–35 (arguing that policy concerns are best handled on a case by case basis rather than through industry specific legislation).

416. *Id.* at 1642–44 (discussing "abstract ideas" as a potential area for judicial patent policy).

417. Carroll, *supra* note 413, at 893 ("The courts have resisted using discretion to sustain categorical exclusions from patentable subject matter, finding this to be too crude a filter.").

418. *Cf.* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW § 2.2 (4th ed. 1992) (common law rules developed over time are efficiency maximizing).

419. John A. Squires & Thomas S. Biemer, *Patent Law 101: Does a Grudging* Lundgren *Panel Decision Mean That the USPTO Is Finally Getting the Statutory Subject Matter Question Right?*, 46 IDEA 561, 582 (2006) ("Instead, the PTO should shift its focus to ensuring that only quality patents are granted. In order to do that, the PTO needs to better utilize the tools already at its disposal . . . . [O]nce the focus is properly placed on the quality of the patents it issues . . . it becomes clear that different tools, 35 U.S.C §§ 102 and 103, readily exist for the PTO to ensure such quality."); Burk & Lemley, *Policy Levers*, *supra* note 17, at 1639 (noting that standards "allow courts flexibility to accommodate different technologies within the general framework of patent law"); Carroll, *supra* note 413, at 893–94 (discussing rigorous patentability standards as reducing uniformity costs). *But see* Drennan, *supra* note 139, at 252–53 (arguing that applying patentability standards to tax patents is "messy" (quoting DONALD S. CHISUM, ET AL., PRINCIPLES OF PATENT LAW 847 (3d ed. 2004))); Burk & Lemley, *Policy Levers*, *supra* note 17, at 1639 (arguing that courts should sometimes implement bright-line rules).

420. Kane, *supra* note 39, at 519 (noting decrease in rejections based on subject matter).

421. *In re* Comiskey, 499 F.3d 1365, 1371 (Fed. Cir. 2007) ("[T]he *obligation* to determine what type of discovery is sought to be patented [so as to determine whether it is 'the kind of "discoveries" that the statute was enacted to protect'] *must precede* the determination of whether that discovery is, in fact, new or obvious." (quoting Parker v. Flook, 437 U.S. 584, 593 (1978)). *But see In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *62 n.1 (Fed. Cir. Oct. 30, 2008) (Rader, J., dissenting) (stating that *Comiskey* does not actually hold that subject matter

the effect of diverting examination resources away from statutory patentability requirements and toward judicial subject matter limitations.[422] Examination resources should instead be focused on whether inventions meet each of the relatively well-settled statutory patentability criteria.[423]

One cost of *ad hoc* patentable subject matter definition[424] is uncertainty and confusion in litigation and patent prosecution,[425] and cases like *Bilski* and *Metabolite* that encourage unpatentable subject matter defenses in litigation serve to increase this cost.[426] If courts impose further subject matter limits, uncertainty costs will continue to grow[427] due to more extensive and prioritized subject matter inquiries in the PTO examination process.[428] If nebulous subject matter requirements were actively enforced, then much more uncertainty would be injected into the patent system.[429] The PTO's subject matter guidelines have already hopelessly confused subject matter with other patentability criteria,[430] and

---

determinations must be done first).

422. Burk & Lemley, *Policy Levers*, *supra* note 17, at 1635 (noting administrative costs of industry specific rules); Cohen, *supra* note 16, at 1168 ("However, the intense focus on statutory subject matter ignores the existence of other statutory requirements for patentability."); Squires & Bierner, *supra* note 419, at 582 ("Special rules create special problems, particularly when a more general fix is required." (citing ADAM B. JAFFE & JOSH LERNER, INNOVATION AND ITS DISCONTENTS 203–05 (2004))).

423. *In re* Bergy 596 F.2d 952, 959–64 (C.C.P.A. 1979) (describing how most rigorous patentability standards are already in place).

424. That is, defining the rules themselves on a case by case basis.

425. *Bergy*, 596 F.2d at 961 ("The PTO, in administering the patent laws, has, for the most part, consistently applied § 102 in making rejections for lack of novelty. To provide the option of making such a rejection under either § 101 or § 102 is confusing and therefore bad law."); Conley & Makowski, *supra* note 41, at 378–79 ("The line between a product of nature, which does not constitute statutory subject matter, and a manmade machine, manufacture, or composition of matter, which does, has not been well defined . . . . The lower courts have been even less helpful in delineating the boundary between products of nature and patentable inventions. In the first place, courts have been inconsistent in deciding whether the product of nature problem is a section 101 subject matter issue, a section 102 novelty issue, a section 103 nonobviousness issue, or some combination of the three.").

426. *See*, Rochelle Cooper Dreyfuss, *In Search of Institutional Identity*, 23 BERKELEY TECH. L.J. 787, 789 (2008) ("[O]bservers of the patent system have voiced increasingly vociferous complaints about the state of patent jurisprudence . . . subjective elements in patent doctrine . . . increase costs and discourage inventors . . . .").

427. Kane, *supra* note 39, at 545–46 ("Yet, the prohibition on patenting laws of nature can result in an absurd kind of legal reductionism if a distinction is not made between the embodiments of physical laws and the laws themselves, such that all entities are judged to be the unpatentable expression of underlying natural laws.").

428. *In re* Comiskey, 499 F.3d at 1371 (subject matter to be determined *first*). *But see In re* Bilski, No. 2007-1130, 2008 WL 4757110, at *62, n.1 (Fed. Cir. Oct. 30, 2008) (Rader, J., dissenting) (arguing that *Comiskey* does not require such initial determination).

429. *Bergy*, 596 F.2d at 961.

430. *See generally* United States Patent and Trademark Office, *Interim Guidelines for Examination of Patent Applications for Patent Subject Matter Eligibility* (Nov. 22 2005),

more restrictions will only cause more confusion. Further, prioritizing the subject matter inquiry means that the PTO and courts would be forced to adjudicate difficult and poorly defined subject matter questions in patent applications that could otherwise be quickly disposed of on other grounds.[431] If § 101 determinations are to be made first, then such determinations should be straightforward and categorical.

A corollary concern also applies: given limited resources, excessive scholarly and popular focus on patentable subject matter detracts from focus on rigorous analysis of the other patentability criteria. A simple count of law review articles relating to the *Metabolite* opinion illuminates the attention given to uncertain patentable subject matter.[432] The time and effort spent thinking about and analyzing patentable subject matter could be better applied to the statutory requirements of patentability with more fruitful results in developing how statutory standards should apply to new technology.[433]

A further problem with broad, non-statutory subject matter restrictions is that such restrictions defeat the purpose of patent law.[434] The constitutional authorization of patent law is to "promote the progress" of useful arts.[435] The primary application of patent law is to determine whether a new technology is useful, novel, and nonobvious—even when the technology was not foreseen when the patent examination system was formed more than 150 years ago.[436] If the

---

available at http://www.uspto.gov/web/offices/pac/dapp/opla/preognotice/guidelines101_20051026.pdf; Osenga, *supra* note 5, at 1111 n.176 ("[I]t is unclear to me (and the 2005 Guidelines do not explain) why the examiner would need to conduct a thorough search of the prior art before determining subject matter eligibility under § 101."). *But see Bilski*, 2008 WL 4757110, at *8 ("whether a claimed process is novel or non-obvious is irrelevant to the § 101 analysis").

431. *See, e.g.*, Dann v. Johnston, 425 U.S. 219, 223–25 (1976). In *Dann*, the Supreme Court granted certiorari on both subject matter and obviousness issues, but only ruled on the obviousness issue, which was much more easily disposed of. *Id.*

432. Cotter, *supra* note 266, at 872 ("But perhaps the most anticipated development in the law of patent eligibility in recent years turned out to be something of a non-event."). Database searches show approximately 100 articles predismissal, and more than 500 secondary source/law journal references to the dissent.

433. Ironically, the time spent in this Article as well.

434. The Supreme Court, lower courts, and commentators often rely on language from the Congressional Report related to the 1952 Patent Act stating that "anything under the sun made by man" is patentable subject matter. *See, e.g.*, Diamond v. Chakrabarty, 447 U.S. 303, 309 (1979). However, the actual report only applies to two categories, "machine or manufacture." S. Rep. No. 82-1979 (1952), *reprinted in* 1952 U.S.C.C.A.N. 2394, 2399; H. R. Rep. No. 82-1923 (1952), *reprinted in* 1952 U.S.C.C.A.N. 2394, 2399. As such, the importance of this phrase on a policy basis may be overstated as to compositions of matter (*e.g.* biotech) and/or processes. In any event, the Report does support the rule proposed in this Article: "Section 101 sets forth the subject matter that can be patented, 'subject to the conditions and requirements of this title.'" *Id.*

435. U.S. CONST. art. I, § 8, cl. 8.

436. Gruner, *supra* note 91, at 396 ("The difficulty in defining patentable subject matter standards lies in describing a future range of potentially patentable technologies of value to the public, but which we cannot understand nor even remotely appreciate in concrete terms now.");

patentability of a new art is in doubt until the Supreme Court rules on such art's subject matter appropriateness, then the patent system cannot foster progress in that art. Instead, technology would have to thrive despite the patent system rather than because of it. As a result, there is little reason to maintain the *status quo* and even less reason to expand or prioritize subject matter bars.

C. *Potential Novelty Problems in Practice*

A final concern with this Article's proposed rule is that the practical application of rigorous patentability standards will not invalidate particular patents that would otherwise be barred by subject matter rules. Part III addressed some of these concerns with respect to particular subject matter areas. This part discusses how one particular issue—preexisting materials and their effect on naturally derived products—might be affected if there were no judicial subject matter limitations.

1. Foreign and Unchanged Prior Art

One concern with reliance on novelty in natural product patents is that preexisting materials unknown in the U.S. or derived from materials located only in foreign countries ordinarily would not be included in prior art under § 102(a), which only bars patents for inventions "known or used" in the United States.[437] This rule may lead to the patenting of derivatives of natural products such as a previously unknown plant or DNA that would not otherwise be patentable under this Article's proposal. However, to the extent that the natural product is truly unknown or available only overseas, then creating an incentive for inventors to seek out and disclose the utility of such compositions may be desirable.[438] Those who would rather see such materials remain in the public domain could bring them to the United States or publish materials about them in order to defeat the novelty claims of others.[439]

Additionally, rigorous patentability could extend to bar patenting of such substances. One alternative is to use § 102(f)[440] to bar patents derived from

---

Kreiss, *supra* note 62, at 66; Kiley, *supra* note 7, at 474 (noting that if patents are limited to what Congress knows at any given time, then there can be few new patents because current technologies would be obvious).

437. 35 U.S.C. § 102(a) (2000). Note, however, that § 102(a) does include foreign prior art if it is published or patented. *Id.*

438. *Cf.* The Incandescent Lamp Patent, 159 U.S. 465, 475–76 (1895) (use of Japanese bamboo in a light bulb).

439. This is a partial answer to the protection (or non-protection) of traditional knowledge. Under this Article's proposal, traditional U.S. knowledge could never be patentable, and those wanting to stop patents on foreign knowledge would seek that knowledge out and publicize it.

440. No patent may be granted if inventor "did not himself invent the subject matter sought to be patented." 35 U.S.C. § 102(f) (2000).

preexisting materials wherever located.[441] An unchanged natural product is not invented by anyone.[442] Even if the material is modified, materials derived from another under § 102(f) could be considered prior art for obviousness consideration.[443] Rigorous patentability implies that the literal terms of § 102(f) could apply to substances that are naturally occurring.

2. The Inherency Problem

One problem with the rigorous view of novelty proposed in this Article is that many patented or indisputably patentable inventions were merely extensions of preexisting natural phenomena.[444] Many "natural" inventions throughout history were discovered both through ingenuity and by accident; this alone should not be a bar to patentability.[445] For example, Pasteur invented germ-free yeast through the process of making beer.[446] Even Edison's light bulb was an extension of the discovery that a certain type of bamboo would glow brightly conducting an electrical current without burning out of existence.[447] Critics may then ask why these inventions should be considered novel while an isolated gene or test for a vitamin deficiency should not be patentable. As a result, gene patenting and *Metabolite* style cases are far more difficult from a policy standpoint than they may appear. Determining just what should and should not anticipate claims can be very difficult.

The answer lies in the inherency doctrine. Inherency is the "unintended, 'accidental' anticipation of an invention. [It] involve[s] the inherent, unintended production of a particular physical product." [448] In *Schering Corp. v. Geneva Pharmaceuticals*, the Federal Circuit ruled that the patent for Loratadine[449]

---

441. Liivak, *supra* note 159, at 265 ("The patent applicant cannot simply patent something they find even if they can prove that another *person* did not create it. Thus, secondly, originality also requires that the patent applicant did not copy or take the subject matter of the patent from *somewhere* else.") (emphasis in original).

442. *See* 35 U.S.C. § 102 (2000). Indeed, such a substance might fail § 101 scrutiny as not being "new" – this is an argument for giving that term independent meaning. See text accompanying note 74 *supra*.

443. OddzOn Prods., Inc. v. Just Toys, Inc., 122 F.3d 1396, 1403–04 (Fed. Cir. 1997).

444. *Cf.* Conley & Makowski, *supra* note 41, at 391 ("[T]he fact that an invention possesses novelty does not prove that it is not a product of nature, since new products of nature are discovered every day.").

445. 35 U.S.C. § 103(a) (2000) ("Patentability shall not be negatived by the manner in which the invention was made.").

446. U.S. Patent No. 141,072, at [3] (filed May 9, 1873) (claiming "[y]east, free from organic germs of disease, as an article of manufacture").

447. *See* The Incandescent Lamp Patent, 159 U.S. 465, 475–76 (1895).

448. ROBERT P. MERGES ET AL., INTELLECTUAL PROPERTY IN THE NEW TECHNOLOGICAL AGE 190 (rev. 4th ed. 2007). "Anticipation" means that pre-existing knowledge deprives a patent claim of novelty under 35 U.S.C. § 102. *Id.* at 188.

449. THOMPSON HEALTHCARE, INC., PHYSICIAN'S DESK REFERENCE: GUIDE TO INTERACTIONS, SIDE EFFECTS AND INDICATIONS 659 (2008) (stating that Loratadine is the generic name for

inherently anticipated a proposed patent claim for a composition generated by the human body while metabolizing Loratadine.[450] The court ruled that Loratadine necessarily caused the production of the metabolized composition, and thus the composition was not novel.[451]

Few cases have expressly invalidated a claim for inherent anticipation. Determining why is one of the problems with applying inherency doctrine.[452] In *Tilghman v. Proctor*, the Court considered the patentability of a process to separate component parts of fat by the use of "water at high temperature and pressure."[453] Unexpectedly, the normal operation of known steam engines necessarily ("inherently") created the same compound due to the application of steam to the tallow fat used to lubricate parts.[454] However, the Court did not invalidate the patent based on the steam engine; it held that a process not understood by ones skilled in the art could not bar a patent.[455]

However, the "lack of understanding" argument does not always hold true.[456] In *In re Seaborg*, the Court of Customs and Patent Appeals considered the patentability of a new element, called "Element 95."[457] Seaborg was the first to isolate the element. The patent specification discloses that the process for creating the element required a specific mix of raw materials, as well as a lengthy process for extracting the element from those materials.[458] The inventor's method for extracting the element thus appears to have created something different not just in degree, but in kind.

However, the PTO objected because the same element was believed to exist as part of prior particle acceleration experiments.[459] Seaborg argued that the theoretical existence of the element was less than one one-hundred millionth of a gram distributed among forty tons of uranium.[460] The Court ruled that even if scientists *understood* that Element 95 had in theory been previously created, the amount was undetectable.[461] *Seaborg* does not apply the "understanding" rule of

---

Claritin®).

450. Schering Corp. v. Geneva Pharms., Inc., 339 F.3d 1373, 1380 (Fed. Cir. 2003).

451. *Id.*

452. *See generally* Burk & Lemley, *Inherency*, *supra* note 39.

453. Tilghman v. Proctor, 102 U.S. 707, 709 (1881) (quoting U.S. Patent No. 11,766 (filed Oct. 3, 1854)).

454. *Tilghman*, 102 U.S. at 711–12.

455. *Id.* Note, though, that this was a process patent; it may very well be that the steam process would bar patenting of the product if it was not already known. There was also a real question about whether the process actually separated the fat. *Id.* ("*if* the scum which rose on the water issuing from the ejection pipe was fat acid . . .") (emphasis added).

456. Burk and Lemley argue that "lack of understanding" is not a dispositive factor in any inherency case. Burk & Lemley, *Inherency*, *supra* note 39, at 376–77.

457. *In re* Seaborg, 328 F.2d 996 (C.C.P.A. 1964). Element 95 is now known as Americium. U.S. Patent No. 3,156,523, col.1, l.11-13 (filed Nov. 10, 1964).

458. U.S. Patent No. 3,156,523 (filed Aug. 23, 1946).

459. *Seaborg*, 328 F.2d at 997.

460. *Id.*

461. *Id.* at 998–99.

Tilghman, does not address isolation and purification, and does not provide a useful alternative framework for determining inherency.[462]

Furthermore, the "point of novelty" implications of *Parker v. Flook,* where scientific principles are treated as prior art, are difficult to square with *Tilghman's* "understanding" requirement. The difficulty with interpreting *Flook* as a novelty case is that the mathematical algorithm may not have been understood by those skilled in the art.[463] Thus, the existence of the algorithm in nature should not have constituted inherent anticipation that would have barred patentability of the catalytic converter at issue.[464]

How then, should novelty and inherent anticipation be applied to the use of natural products and phenomena such as genes and medical tests? One potential solution is to apply the general rule suggested by Professors Lemley and Burk: if the public enjoys the "benefits" of a natural product, whether or not known, then no inventor can claim novelty in that product.[465] The "benefits" analysis is consistent with the "known or used" bar to novelty.[466] If the public is not obtaining a benefit, then a product is not used. This means that genes which only express proteins that already exist might not be novel, but spliced genes that express new proteins might be novel.[467] "Benefits" should be expanded to include "benefits or detriments" or perhaps even "effects." As a result, a detrimental product of nature (say, poison ivy) might still inherently anticipate an extract of poison ivy. The "known or used" consideration in novelty is concerned with whether the public has "experienced" the prior product in any way, whether positive or negative.

Another potential reconciliation is to apply the general rule that unmodified preexisting materials or knowledge are not novel, but their use in unnatural ways

---

462. *See generally Seaborg*, 328 F.2d 996. Professors Burk and Lemley attribute this ruling to a general notion that prior "inherent" creation of a product will not bar a patent where the public does not get the benefit of the inherent product. Burk & Lemley, *Inherency*, *supra* note 39, at 382–83.

463. *See* Burk & Lemley, *Inherency*, *supra* note 39, at 407–08 (stating that the better test is whether the public obtained any benefit).

464. *Id.*

465. Burk & Lemley, *Inherency*, *supra* note 39, at 407; *see, e.g.*, Merck & Co. v. Olin Mathieson Chem. Corp., 253 F.2d 156, 161 (4th Cir. 1958) ("As found in 'natural' fermentates, [vitamin B12] has no utility, therapeutically or commercially, until converted into compositions comparable to the patented products."); *see also* N. Scott Pierce, *A New Day Yesterday: Benefit as the Foundation and Limit of Exclusive Rights in Patent Law*, 6 J. MARSHALL REV. INTELL. PROP. L. 373, 416 (2007); *cf.* Foley v. United States 260 U.S. 667, 676–77 (1923) ("The assertions [that a non-infringing method inherently practices the patent] prove too much . . . . If the asserted result was inevitable in the method of the patents, it was inevitable in the method in use prior to the patents, and, we repeat, the patents are left without justification.").

466. 35 U.S.C. § 102(a) (2000) (emphasis added).

467. *See, e.g.*, Schering Corp. v. Geneva Pharms., Inc., 339 F.3d 1373, 1379 (Fed. Cir. 2003) (metabolite of loratadine that is formed in the human body inherently anticipated by loratadine patent, which "enabled" one to ingest loratadine).

may be novel.[468] For example, Edison's light bulb used a particular type of bamboo that had high electrical resistance.[469] This was an unnatural use for a natural product.[470] This rule could be considered a subset of the "benefits" test because the public might not benefit from a natural use.

This analysis confirms why *Metabolite* was a difficult case. Whether the public enjoyed the "benefit" of the natural relationship between homocysteine and vitamin deficiencies is not clear.[471] Thus, discovery and application of the relationship and disclosure of the benefit for the first time may very well have been novel.

3. The Anticipation/Infringement Dichotomy

Seaborg illustrates one additional potential problem with the novelty analysis discussed in Part III and in the previous subsections. The law of novelty follows the maxim, "that which would *literally* infringe if later in time anticipates if earlier than the date of the invention."[472] Applied to DNA, this means that if *in vivo* DNA would infringe a purified gene patent, then *in vivo* DNA would anticipate the gene patent. Those who support gene patents attempt to garner support by arguing that *in vivo* DNA would not infringe a gene patent.[473]

---

468. *See, e.g.*, Funk Bros. Seed Co. v. Kalo Inoculant Co., 333 U.S. 127, 130 (1948) ("He who discovers a hitherto unknown phenomenon of nature has no claim to a monopoly of it which the law recognizes. If there is to be invention from such a discovery, it must come from the application of the law of nature to a new and useful end."); Mackay Co. v. Radio Corp., 306 U.S. 86, 94 (1939) ("[W]hile a scientific truth, or the mathematical expression of it, is not a patentable invention, a novel and useful structure created with the aid of knowledge of scientific truth may be."); Tilghman v. Proctor, 102 U.S. 707, 729 (1880) ("The chemical principle or scientific fact upon which it is founded is, that the elements of neutral fat require to be severally united with an atomic equivalent of water in order to separate from each other and become free. This chemical fact was not discovered by Tilghman. He only claims to have invented a particular mode of bringing about the desired chemical union between the fatty elements and water.").

469. The Incandescent Lamp Patent, 159 U.S. 465, 475–76 (1895).

470. *See id.*

471. Pierce, *supra* note 465, at 450–51 (arguing that the public did not benefit from the "natural" relationship, and thus the discovery was novel); LeRoy v. Tatham, 55 U.S. 156, 175 (1853) ("A patent will be good, though the subject of the patent consists in the discovery of a great, general, and most comprehensive principle in science or law of nature, if that principle is by the specification applied to any special purpose, so as thereby to effectuate a practical result and benefit not previously attained.") (citation omitted).

472. Lewmar Marine, Inc. v. Barient, Inc., 827 F.2d 744, 747 (Fed. Cir. 1987) (emphasis in original).

473. *See, e.g.*, Kevin Noonan, *Science Fiction in the New York Times* (Feb. 13, 2007) ("[A]nyone familiar with this space or any other truthful description of DNA patenting knows that patented DNA must be 'isolated' or 'isolated and purified.' In short, no one has ownership rights over 'your' DNA . . . ."), *available at* http://patentdocs.typepad.com/patent_docs/2007/02/science_fiction.html.

The problem thus lies in the logical contrapositive of the maxim: if *in vivo* DNA cannot infringe a gene patent, then *in vivo* DNA cannot anticipate a gene patent. In *Seaborg*, for example, the contrapositive appears to hold true—the prior experiments would not have infringed the "isolated" Element 95, and thus they did not anticipate the claim.[474] Following this logic, *in vivo* DNA cannot deprive isolated gene claims of their novelty.[475]

Two potential answers exist that will allow the maxim and its contrapositive to hold true with the "benefits" view of inherency.[476] The first approach is to accept that a gene patent would be infringed by *in vivo* DNA. This is called "inherent infringement,"[477] and it is not unheard of.[478] However, this does not mean that the maxim would cause infringement of every natural extract. If the test for novelty is whether the new composition is different in kind, then there is no need to worry about *in vivo* infringement because the composition will be different in fact. Patentees would be very unlikely to argue for *in vivo* infringement, because doing so would be tantamount to admitting that the composition is not novel.[479]

The second answer is to apply an obviousness test instead of a novelty test. To the extent that an isolated composition is different from the natural prior art such that *in vivo* DNA neither anticipates nor infringes such a claim,[480] it may still be obvious and unpatentable in light of the *in vivo* DNA.[481] Each of these two answers is consistent with the application of rigorous patentability in accordance with historical analytic principles in patent law.

CONCLUSION

Abandoning judicial subject matter restrictions will not answer all of the difficult patentability questions that have arisen and may yet arise as our nation's inventors and researchers continue to discover new technologies. Those difficult questions, however, should be answered by the general criteria that Congress has established—criteria that have worked for over 150 years—to determine whether a particular patent claim should be allowed.

---

474. *In re* Seaborg, 328 F.2d 996, 997 (1964); Merges & Duffy, *supra* note 162, at 189, 374.

475. *Cf.* Schering Corp. v. Geneva Pharms., Inc., 339 F.3d 1373, 1381 (Fed. Cir. 2003) (holding that a nonisolated composition claim is anticipated by chemicals in the body, but stating in dicta that isolated composition would not be anticipated). *Schering* cites the anticipation/infringement maxim with approval. *Id.* at 1379.

476. A third, more simple answer is to accept that the contrapositive need not be true, but that solution has not been borne out over time, nor is it logically satisfying.

477. Burk & Lemley, *Inherency*, *supra* note 39, at 401 (discussing inherent infringement).

478. *Id.* (discussing Smithkline Beecham Corp. v. Apotex Corp., 365 F.3d 1306 (Fed. Cir. 2004)).

479. *See, e.g., id.* at 401 n.152 (noting that a patent that was "inherently infringed" was ultimately invalidated based on inherent anticipation).

480. Thus, it would be novel.

481. *But see* Cohesive Tech, Inc. v. Waters Corp., 543 F.3d 1351, 1364–65 (Fed. Cir. 2008) (inherent prior art need not lead to obviousness).

The exact contour of the trade-offs between innovation and patent protection are largely unknown. Therefore, the PTO and courts should focus on answering specific questions about how to best apply rigorous standards of novelty, nonobviousness, utility, and specification with a scalpel rather than simply eliminating broad swaths of innovation with a machete.

IT'S JUST NOT WORTH SEARCHING FOR WELCOME MATS WITH A KALEIDOSCOPE AND A BROKEN COMPASS

RORY RYAN*

INTRODUCTION

Each semester it seems that Justice Oliver Wendell Holmes's fan base grows as law students struggle with the elusive meaning of the words "arising under" in 28 U.S.C. § 1331.[1] Justice Holmes construed those words to mean something simple and ascertainable: "A suit arises under the law that creates the cause of action."[2] To some, the test is irresistibly appealing for its simplicity; but not to the Supreme Court. It did not take the Court long to reject the Holmes test as an exclusive test and to create a second branch of "arising under" jurisdiction.[3] For about eighty-five years, courts and commentators have struggled with defining the boundaries of the second branch,[4] defending its existence,[5] and occasionally

---

* Rory Ryan, Associate Professor of Law, Baylor Law School. Thanks to Jim Underwood, James Pfander, Lumen Mulligan, and Lonny Hoffman for reading earlier drafts. I also appreciate the contributions of my current and former students Gordon Davenport, Jonna Stallings, Jeffrey Fisher, Dan Kennedy, and Ryan Reneau.

1. *See infra* note 22.

2. Am. Well Works Co. v. Layne & Bowler Co., 241 U.S. 257, 260 (1916).

3. *See* Lonny S. Hoffman, *Intersections of State and Federal Power: State Judges, Federal Law, and the "Reliance Principle,"* 81 TUL. L. REV. 283, 294 (2006) (citing Smith v. Kansas City Title & Trust Co., 255 U.S. 180, 199 (1921)); *see also* Franchise Tax Bd. of Cal. v. Constr. Laborers Vacation Trust for S. Cal., 463 U.S. 1, 9 (1983).

4. *See generally* Patti Alleva, *Prerogative Lost: The Trouble with Statutory Federal Question Doctrine After Merrell Dow*, 52 OHIO ST. L.J. 1477 (1991) (discussing the problems associated with state law hybrid claims after the Supreme Court's decision in *Merrell Dow*); James H. Chadbourn & A. Leo Levin, *Original Jurisdiction of Federal Questions*, 90 U. PA. L. REV. 639 (1942) (discussing evolving interpretations of Article III of the Constitution and its boundaries); William Cohen, *The Broken Compass: The Requirement that a Case Arise "Directly" Under Federal Law*, 115 U. PA. L. REV. 890 (1967) (suggesting that a federal tribunal should decide whether a case arises under federal law); Donald L. Doernberg, *There's No Reason for It; It's Just Our Policy: Why the Well-Pleaded Complaint Rule Sabotages the Purposes of Federal Question Jurisdiction*, 38 HASTINGS L.J. 597 (1987) (addressing the deficiencies in the well-pleaded complaint rule); Ray Forrester, *Federal Question Jurisdiction and Section 5*, 18 TUL. L. REV. 263 (1944) (suggesting that the constitutional source of judicial power should be broad); Ray Forrester, *The Nature of a "Federal Question,"* 16 TUL. L. REV. 362 (1942) (discussing the confusion surrounding federal question jurisdiction); Richard D. Freer, *Of Rules and Standards: Reconciling Statutory Limitations on "Arising Under" Jurisdiction*, 82 IND. L.J. 309 (2007) (addressing positive and negative aspects of historical interpretations of "arising under" jurisdiction); Qian A. Gao, *"Salvage Operations Are Ordinarily Preferable to the Wrecking Ball": Barring Challenges*

to Subject Matter Jurisdiction, 105 COLUM. L. REV. 2369 (2005) (discussing belated challenges to subject matter jurisdiction); Ronald J. Greene, *Hybrid State Law in the Federal Courts*, 83 HARV. L. REV. 289 (1969) (discussing Supreme Court review of hybrid cases and problems with original, non-diversity jurisdiction in federal district courts); Linda R. Hirshman, *Whose Law Is It, Anyway? A Reconsideration of Federal Question Jurisdiction Over Cases of Mixed State and Federal Law*, 60 IND. L.J. 17 (1984) (analyzing hybrid law cases and their outcomes); Hoffman, *supra* note 3 (discussing the Supreme Court's efforts to allocate authority in cases that involve federal law or federal interest); Gregory P. Joseph, *Federal Class Action Jurisdiction After CAFA*, Exxon Mobil *and* Grable, 8 DEL. L. REV. 157 (2006) (exploring the impact of CAFA, *Exxon Mobil*, and *Grable* on federal jurisdictional principles governing class actions); Ernest J. London, *"Federal Question" Jurisdiction—A Snare and a Delusion*, 57 MICH. L. REV. 835 (1959) (asserting that the federal question test is a misleading standard); Douglas D. McFarland, *The True Compass: No Federal Question in a State Law Claim*, 55 U. KAN. L. REV. 1 (2006) (discussing federal question jurisdiction when federal laws are embedded in state law claims); Daniel J. Meltzer, Jurisdiction and Discretion *Revisited*, 79 NOTRE DAME L. REV. 1891 (2004) (analyzing David Shapiro's assertions regarding jurisdiction and discretion); Arthur R. Miller, *Artful Pleading: A Doctrine in Search of Definition*, 76 TEX. L. REV. 1781 (1998) (discussing the meaning of an artful pleading under Article III of the Constitution); Paul J. Mishkin, *The Federal "Question" in the District Courts*, 53 COLUM. L. REV. 157 (1953) (addressing considerations involved in federal question jurisdiction in district courts); John B. Oakley, *Federal Jurisdiction and the Problem of the Litigative Unit: When Does What "Arise Under" Federal Law?*, 76 TEX. L. REV. 1829 (1998) (discussing alternative interpretations of "arising under" in four Supreme Court cases); Jason Pozner, *The More Things Change, the More They Stay the Same:* Grable & Sons v. Darue Engineering *Does Not Resolve the Split over* Merrell Dow v. Thompson, 2 SETON HALL CIRCUIT REV. 533 (2006) (discussing the existing confusion over *Merrell Dow*); John F. Preis, *Jurisdiction and Discretion in Hybrid Law Cases*, 75 U. CIN. L. REV. 145 (2006) (discussing a new rule for addressing Article III problems); Robert J. Pushaw, Jr., *A Neo-Federalist Analysis of Federal Question Jurisdiction*, 95 CAL. L. REV. 1515 (2007) (analyzing Paul Mishkin's approach to federal question jurisdiction in district courts); David L. Shapiro, *Jurisdiction and Discretion*, 60 N.Y.U. L. REV. 543 (1985) (discussing the necessity of federal court discretion to hear federal question cases); Harry Shulman & Edward C. Jaegerman, *Some Jurisdictional Limitations on Federal Procedure*, 45 YALE L.J. 393 (1936) (discussing reformation of federal procedure).

5. *See, e.g.*, G. Merle Bergman, *Reappraisal of Federal Question Jurisdiction*, 46 MICH. L. REV. 17 (1948) (exploring the historical development of federal question jurisdiction and defending its current application); Cohen, *supra* note 4, at 898 (noting the "severe limitations" of the Holmes test), at 905-06 (asserting that the majority of federal civil litigation presents no jurisdictional problem); Doernberg, *supra* note 4, at 656-58 (defending the existence of the second branch of federal jurisdiction but recommending an outcome-determinative test rather than the *Merrell Dow* substantiality test) ; Freer, *supra* note 4, at 311-12 (criticizing the Holmes test and praising the broader view of federal question jurisdiction applied in *Grable* as providing "meaningful guidance" and "appropriate balance"); Mishkin, *supra* note 4, at 163 (noting the importance of retaining a broad definition of "arising under"); Pozner, *supra* note 4, at 537-38 (advocating the "substantial question" test for federal question jurisdiction without deference to Congress as required by *Grable*); Pushaw, *supra* note 4, at 1518 (suggesting federal jurisdiction "over any case that will likely depend on the resolution of a genuine dispute over the interpretation, application, or enforcement of federal law").

calling for its elimination.[6] I now join the latter group; Justice Holmes and his growing fan club are correct. Simpler may not always (or even often) be better, but it is in this context. My thesis is easy to follow: the second branch should be eliminated and Congress should do the eliminating.

This Article's title reflects the historical difficulties with defining the second branch. The second branch was born in *Smith v. Kansas City Title & Trust Co.*[7] There, just a few years after Justice Holmes had announced that a suit only "arises under the law that creates the cause of action,"[8] the *Smith* Court held that a state-created cause of action might still arise under federal law if its resolution "depends upon the construction or application" of federal law.[9] Despite *Smith*'s broad "depends upon" formulation, it soon became clear that not just any federal issue would suffice; the federal issue had to be "substantial" or "important."[10] The debate about the second branch has long centered on what types of federal issues present in a state-created cause of action create "arising under" jurisdiction.

Early on, Justice Cardozo noted the difference between the bright-line Holmes rule and the flexible second-branch standard, writing that the second branch requires a "common-sense accommodation of judgment to kaleidoscopic situations" that present a federal issue in "a selective process which picks the substantial causes out of the web and lays the other ones aside."[11] Then, intellectually armed with kaleidoscopes, the legal academy and the Court sought to determine these elusive boundaries. Unsurprisingly, the boundaries became anything but clear, and we learned from Professor Cohen's landmark article that the "arising under" compass was broken.[12] About twenty-five more years passed before the Court decided *Merrell Dow*,[13] which many thought nearly eliminated the second branch. But all *Merrell Dow* did in reality was create a three-way circuit split, as proponents of the second branch refused to read the case so restrictively.[14] Then came the Supreme Court's latest word in *Grable*, where the Court again reformulated the test and

---

6. *See, e.g.*, Hirshman, *supra* note 4, at 21-22 (suggesting a return to the Holmes test); McFarland, *supra* note 4, at 2 (advocating a return to an earlier understanding of "arising under" and asserting that a claim created by state law is not a federal question); Preis, *supra* note 4, at 149 (insisting that federal courts have jurisdiction over hybrid cases "only when the federal question embedded in the state cause of action is supported by a federal cause of action").

7. Smith v. Kansas City Title & Trust Co., 255 U.S. 180 (1921).

8. *Am. Well Works Co.*, 241 U.S. at 260.

9. *Smith*, 255 U.S. at 199.

10. *See, e.g.*, Merrell Dow Pharm., Inc., v. Thompson 478 U.S. 804, 806-07 (1986) (quoting Thompson v. Merrell Dow Pharm., Inc., 766 F.2d 1005, 1006 (1985)); Franchise Tax Bd. of Cal. v. Constr. Laborers Vacation Trust for S. Cal., 463 U.S. 1, 13 (1983); Gully v. First Nat'l Bank, 299 U.S. 109, 118 (1936).

11. *Gully*, 299 U.S. 109 at 117-18.

12. Cohen, *supra* note 4, at 890 (declaring that the "arising under" jurisdiction of lower federal courts "has been a puzzle to judge and scholar alike").

13. *Merrell Dow*, 478 U.S. 804.

14. *See* Pozner, *supra* note 4, at 576.

taught that *Merrell Dow* was actually decided under a never-before-articulated prong, which required finding "welcome mats" when exercising jurisdiction would be disruptive.[15]

Ultimately I conclude that the second branch is not worth it.[16] Throughout the second- branch evolution, we have been assured that sufficiently clear boundaries would develop, but they have not and will not. The nature of the second-branch inquiry prevents that from happening. I find particularly unpersuasive the criticism that the Holmes test is too rigid in the face of assurances that the more flexible second branch is acceptable because it will eventually develop sufficient rigidity. The costs associated with such a fuzzy jurisdictional inquiry simply outweigh the benefits. A modest sample of post-*Grable* opinion-generating second-branch removal cases shows, conservatively, that eight cases are delayed and remanded for an average of six months each for every case that satisfies *Grable*.[17] The class of delay-prone cases will remain large because most colorably removable cases are, in fact, removed and the very nature of the second-branch test casts a wide net of colorability.

First, though, before arguing that the second branch should be eliminated, I explain why Congress should effect the change by statutory amendment.[18] In *Grable*, Justice Thomas expressed his displeasure with the current construction of 28 U.S.C. § 1331 and invited original-intent arguments to return to the Holmes test in appropriate cases.[19] I agree with Justice Thomas that we should look for the Holmes test—but to today's Congress, not the Congress of 1875 when the relevant "arising under" words first appeared. Ultimately, I propose that Congress should amend 28 U.S.C. § 1331 to define "arising under" in a manner consistent with the Holmes test.

Finally, I discuss issues surrounding implementation.[20] The presence of other statutes referencing "arising under"[21] makes it preferable for Congress to define the phrase solely for purpose of § 1331 rather than either (1) remove the phrase from § 1331 or (2) define the phrase for all of the Judicial Code. Because "arising under" would be narrowly defined for only § 1331 under this proposal, the words would retain most of their current post-*Grable* meaning for the rest of the Judicial Code. Congress could therefore retain the current definition in areas of exclusive federal jurisdiction and in other selected areas where Congress decides to use more of its Article III jurisdiction-conferring power. Because the allocation would then be congressionally dependent, the *Grable* welcome-mat/disruptiveness inquiry would disappear. No longer would a federal cause of action be a welcome mat; rather, the welcome mat would be a subject-matter-specific grant of

15. Grable & Sons Metal Prods., Inc. v. Darue Eng'g & Mfg., 545 U.S. 308, 318 (2005).

16. *See infra* Part III.

17. *See infra* Part III-B.

18. *See infra* Part II.

19. *Grable*, 545 U.S. at 320-22 (Thomas, J., concurring).

20. *See infra* Part IV.

21. *E.g.*, 28 U.S.C.A. §§ 157, 613, 1295, 1334, 1337, 1338, 1339, 1340, 1400, 1407, 1409, 1411, 1441, 1445, 1491, 1505, 1658, 2410 (2008).

jurisdiction that Congress has excluded from 28 U.S.C. § 1331's newly imposed limitation.

As to counterarguments, which are addressed throughout, the normative questions here are not easy. The Holmes test is narrow and contains bright lines. No doubt, at the fringes it will exclude some cases that seem to be proper candidates for initial resolution in federal court. Bright-line rules will do that. In this context, it's worth it.

LOOKING FOR HOLMES TODAY, NOT FROM 1875

Understanding why the solution should come from Congress requires a comparison of the words and functions of both Article III, Section 2 and 28 U.S.C. § 1331. Both use the same "arising under" language but serve very different roles.[22] Article III, Section 2 does not confer jurisdiction on the lower federal courts—it is not self-executing. Instead, the lower federal courts need statutory authorization.[23] Article III, Section 2 merely defines the limits on Congress's power to give its courts jurisdiction.[24] So, the jurisdictional inquiry contains a question of statutory construction (whether Congress conferred jurisdiction) and one of constitutional power (whether Article III, Section 2 authorizes Congress to confer such jurisdiction).

Initially, the Court construed the words "arising under" in Article III, Section 2 very broadly. In *Osborn v. Bank of the United States*, Congress had authorized federal jurisdiction over all suits by or against the Bank of the United States.[25] The Court held that Congress had the authority, using its "arising under" power, to confer federal jurisdiction in even an ordinary breach-of-contract suit against the Bank.[26] This was because federal law created the Bank and its right to

---

22. *Compare* U.S. CONST. art. III, § 2, cl. 1 ("The judicial Power shall extend to all Cases . . . *arising under* this Constitution, the Laws of the United States, and Treaties made . . . under their Authority" (emphasis added)), *with* 28 U.S.C. § 1331 (2000) ("The district courts shall have original jurisdiction of all civil actions *arising under* the Constitution, laws, or treaties of the United States." (emphasis added)).

23. Although Article III, Section 2's "shall extend" language is susceptible to different interpretation, it is now well-established that statutory authorization is needed. *See, e.g.*, Bender v. Williamsport Area Sch. Dist., 475 U.S. 534, 541 (1986) ("Federal courts are not courts of general jurisdiction; they have only the power that is authorized by Article III of the Constitution and the statutes enacted by Congress pursuant thereto."); Cary v. Curtis, 44 U.S. 236, 245 (1845) ("[T]he judicial power of the United States, although it has its origin in the Constitution, is . . . dependent for its distribution and organization . . . entirely upon the action of Congress . . . .").

24. Rory Ryan, *Consistent "Deeming": A Cohesive Construction of 28 U.S.C. § 1332 in Cases Involving International Corporations and Permanent-Resident Aliens*, 3 SETON HALL CIRCUIT REV. 73, 78-79 (2006); *see* Mesa v. California, 489 U.S. 121, 136 (1989) (citing Verlinden B.V. v. Cent. Bank of Nigeria, 461 U.S. 480, 491 (1983)); Sheldon v. Sill, 49 U.S. 441, 449 (1850) ("The Constitution has defined the limits of the judicial power of the United States, but has not prescribed how much of it shall be exercised by the [lower federal courts] . . . .").

25. Osborn v. Bank of the United States, 22 U.S. 738, 824-26 (1824).

26. *Id.* at 817–19. *See generally* Anthony J. Bellia, Jr., *The Origins of Article III "Arising*

contract, and because any suit against the Bank could potentially raise a question about that authority.[27] The Supreme Court has never defined the precise boundaries of this so-called "potential ingredient" test, but the 1824 *Osborn* decision and its progeny hint at a vast universe of constitutional power.

The words "arising under" did not survive in a general grant of jurisdiction until the Judiciary Act of 1875.[28] Before that, some statutes (like the one in *Osborn*) conferred jurisdiction over particular matters or litigants, but there was no general federal-question statute. Obviously, because the statute did not exist in 1824, the *Osborn* Court did not compare the statutory and constitutional language.

What did Congress intend in 1875? It granted original federal-question jurisdiction for the first time "to prevent states from thwarting its Reconstruction-era constitutional amendments and legislation."[29] It chose the same language from Article III, Section 2 that the Court had construed so broadly in *Osborn*.[30] From the perspective of one construing the statute in 1876, it certainly seems that Congress attempted to extend jurisdiction to the constitutional limits defined in *Osborn*. That seemed to be a sensible purpose at the time, and choosing already-construed words seemed a sensible way of doing so. Even a textualist, giving the words their most reasonable meaning, has to consider the relevant context for the words: Congress was trying to expand federal power and it chose, of all available words, the two-word phrase that had already been interpreted in *Osborn*—an unlikely coincidence.[31] Additionally, the meager legislative history behind the

---

Under" Jurisdiction, 57 DUKE L.J. 263, 332–40 (2007) (discussing the Court's rationale in *Osborn*).

27. *See Osborn*, 22 U.S. at 824; James E. Pfander, *Protective Jurisdiction, Aggregate Litigation, and the Limits of Article III*, 95 CAL. L. REV. 1423, 1426-27 (2007).

28. The same "arising under" language also briefly appeared in the Judiciary Act of 1801, known as the Midnight Judges Act: "[The] circuit courts respectively shall have cognizance of . . . all cases in law or equity, *arising under the constitution and laws of the United States*, and treaties made, or which shall be made, under the authority . . . ." Midnight Judges Act, 2 Stat. 89, 92 (1801) (emphasis added). That statute was quickly repealed. *See* Act of Mar. 8, 1802, ch. 8, 2 Stat. 132 (1802). The predecessor to the modern statute, which contained the "arising under" language, first appeared in 1875. *See* Act of Mar. 3, 1875, ch. 137, 18 Stat. 470 (1875).

29. Pushaw, *supra* note 4, at 1551 n.229.

30. Act of Mar. 3, 1875, ch. 137, 18 Stat. 470 (1875) ("[T]he [lower federal courts] of the United States shall have original cognizance . . . of all suits of a civil nature . . . *arising under* the Constitution or laws of the United States, or treaties made, or which shall be made, under their authority . . . ." (emphasis added)).

31. *See* McFarland, *supra* note 4, at 23-24; Pozner, *supra* note 4, at 540-41; Pushaw, *supra* note 4, at 1522-23. *But see* Chadbourn & Levin, *supra* note 4, at 649-50 (arguing that Congress instead intended to provide relief from litigation that technically fell within the limits of the *Osborn* rationale in § 5 of the Act of Mar. 3, 1875, which required dismissal or remand if it appeared at any time "that such suit does not really and substantially involve a dispute or controversy properly within the jurisdiction of said [lower federal court]" (quoting Act of Mar. 3, 1875, ch. 137, 18 Stat. 470, 472)).

1875 Act[32] and Act's sparse contemporary commentary[33] bolster the notion that Congress used the phrase "arising under" to mean *Osborn*—that is, to extend jurisdiction to the maximum limits of Article III, Section 2.[34]

The Supreme Court quickly realized that it would not be feasible to apply the *Osborn* test to the statutory text. If the federal courts had original jurisdiction over all suits involving a potential federal ingredient, almost nothing would be outside federal reach.[35] For example, as Professor Freer has commented,

---

32. While there was very little debate on this bill, Senator Matthew H. Carpenter, the author of the bill and the spokesman for the judiciary committee when it was debated on the Senate floor, made a series of statements:

> The act of 1789 did not confer the whole power which the Constitution conferred; it did not do what the Supreme Court has said Congress ought to do; it did not perform what the Supreme Court has declared to be the duty of Congress. This bill does. . . .
>
> This bill gives precisely the power which the Constitution confers—nothing more, nothing less. The Senator from California proposes to limit the constitutional jurisdiction and restrict it because it was restricted in 1789. . . . The whole circumstances of the case are different, and the time has now arrived it seems to me when Congress ought to do what the Supreme Court said more than forty years ago it was its duty to do, vest the power which the Constitution confers in some court of original jurisdiction.

2 CONG. REC. 4986-87 (1874); *see also* Forrester, *supra* note 4, at 375 (arguing that the statute's drafters intended to give it the same scope as the judicial clause of the Constitution and to use "arising under" in the same manner as in Article III).

33. *See* Forrester, *supra* note 4, at 376 ("Thus we see, that commencing in 1864, before the close of rebellion, and culminating in March, 1875, at the very close of the last session, *Congress has exhausted its power; and has conferred upon the federal courts all the jurisdiction authorized by the [C]onstitution.*" (quoting A. I., *Our Federal Judiciary*, 2 CENT. L.J. 551, 553 (1875)); Pushaw, *supra* note 4, at 1551 n.229 ("The sole extant article on this topic in 1875 concluded that Congress, to prevent states from thwarting its Reconstruction-era constitutional amendments and legislation, gave federal courts all jurisdiction authorized by the Constitution." (citing A. I., *supra* at 553)); *see also* Chadbourn & Levin, *supra* note 4, at 644 ("[B]oth prior and subsequent to the enactment of the statute there was hardly any discussion of the problem in the legal periodicals . . . ."); Felix Frankfurter, *The Business of the Supreme Court of the United States—A Study in the Federal Judicial System: II. From the Civil War to the Circuit Courts of Appeals Act*, 39 HARV. L. REV. 35, 44 n.34 (explaining the "barren" nature of contemporary periodicals concerning the enactment of the Act of 1875); William M. Meigs, *The Relief of the Supreme Court of the United States*, 32 AM. L. REG. 360 (1884) (discussing the overload of cases in the federal judiciary but not mentioning the act's potential expansion of federal jurisdiction); Seymour D. Thompson et al., *Reorganization of the Federal Judiciary—Mr. McCrary's Bill*, 3 CENT. L.J. 68, 70 (1876) (arguing that the bill would rightly lead to the creation of intermediate courts of appeals, but that it failed to provide for a sufficient number of judges to manage the increased litigation that the jurisdictional extension would cause).

34. *See* Cohen, *supra* note 4, at 891 n.13 ("Of course, the broad language of the statute, combined with the broad statement of Senator Carpenter, who was in charge of the bill, 2 CONG. REC. 4986-87 (1874), argue for the interpretation that Congress intended to give the federal trial courts all the judicial power specified in Article 3 of the Constitution." (citing Forrester, *supra* note 4, at 374-76)).

35. *See* London, *supra* note 4, at 839 (pointing out that if the enabling statute were "given a construction coextensive with the construction . . . already given to the similar phrasing in Article III

practically every suit involving real property would invoke federal jurisdiction because one might validly question whether title was derived from the United States.[36] Redefining the words was a matter of the Court's "self-preservation."[37] And so began the divergence of the meaning of the same phrase in two different provisions.

Once the divergence began, a tension arose. Congress is charged with defining federal jurisdiction.[38] Yet, cases construing 28 U.S.C. § 1331 had little to do with statutory interpretation. Congress seemed to intend *Osborn*, yet when the Supreme Court began to construe Congress's Act, *Osborn* was not one of the choices. Of course, when I speak of divergences and departures, I do not mean to imply that the Court should have read the statutory phrase to mean *Osborn*. Narrowing the scope of "arising under" was indeed a matter of self-preservation.[39] In his classic 1953 article, Professor Mishkin forcefully argued for acceptance of this retreat from the *Osborn* interpretation.[40] Recently, while examining Mishkin's article from a neo-federalist perspective, Professor Pushaw aptly observed:

Mishkin apparently argued that Congress implicitly had given discretion to the judiciary to make determinations about federal question jurisdiction in light of practical considerations. He cited nothing in the statute's language or legislative history to support that contention, and indeed acknowledged that this evidence suggested Congress's intent to confer the full scope of Article III power over such cases.[41]

The point is that, from that moment of divergence, § 1331 questions have been far more normative than interpretive.

---

of the Constitution, access to the federal courts would then have been virtually unrestricted."). *See also* Gully v. First Nat'l Bank, 299 U.S. 109, 118 (1936) ("[C]ountless claims of right can be discovered to have their source or their operative limits in the provisions of a federal statute or in the Constitution itself with its circumambient restrictions upon legislative power."); Murdock v. City of Memphis, 87 U.S. 590, 629 (1874) ("[T]here is no conceivable case so insignificant in amount or unimportant in principle that a perverse and obstinate man may not bring it to this court by the aid of a sagacious lawyer raising a Federal question in the record . . . .").

36. Freer, *supra* note 4, at 315.

37. *Id.* at 315 ("It is not an overstatement to say that the Court's retreat from *Osborn* as the statutory standard for 'arising under' was rooted ultimately in institutional self-preservation.").

38. U.S. CONST. art. III, § 1; Kontrick v. Ryan, 540 U.S. 443, 444 (2004) ("Only Congress may determine a lower federal court's subject-matter jurisdiction."); Keene Corp. v. United States, 508 U.S. 200, 207 (1993) ("Congress has the constitutional authority to define the jurisdiction of the lower federal courts . . . ."); Scott Dodson, *Jurisdictionality and* Bowles v. Russell, 102 NW. U. L. REV. COLLOQUY 42, 43 (2007) ("Congress has the constitutional authority to regulate the courts' jurisdiction.").

39. Freer, *supra* note 4, at 315.

40. Mishkin, *supra* note 4, at 165-66 ("Just as the test should yield cases with a high probability of material federal issues, so also it should so far as possible eliminate those highly likely to go off entirely on state grounds.").

41. Pushaw, *supra* note 4, at 1539 (footnotes omitted).

The first major departure was the well-pleaded-complaint rule. While the *Osborn* rule extended constitutional power to potential ingredients, the well-pleaded complaint rule narrowed the statutory meaning considerably. In *Louisville & Nashville Railroad Co. v. Mottley*, the Court clarified that the statute requires, not just a potential federal ingredient, but a federal issue present on the face of the plaintiff's well-pleaded complaint.[42] Even though the only disputed issue in *Mottley* was a federal one, jurisdiction failed.[43] Essentially, Congress *wrote* "potential ingredient," and the Court *read* something far narrower.[44]

Then, in *American Well Works Co. v. Layne & Bowler Co.*, Justice Holmes attempted to narrow the rule again.[45] One might call the well-pleaded complaint rule a where-to-look rule. Justice Holmes took the next step by narrowing what courts should look for in a well-pleaded complaint. According to Justice Holmes, when evaluating the well-pleaded complaint, courts must find, not just a federal issue, but rather that the plaintiff asserted a federally created cause of action.[46] Once again, note the departure from *Osborn*. The ordinary breach-of-contract suit against the national bank in *Osborn* would have surely failed the Holmes test, yet Justice Holmes, based on the same words that had been construed in *Osborn*, was advocating an exclusive test that required a federal cause of action.

After *American Well Works*, congressional intent was no longer relevant. The search was for the best rule, not the enacted one. And in *Smith v. Kansas Title & Trust Co.*, the Court rejected Justice Holmes's rule, thus creating the second branch of federal-question jurisdiction.[47] In *Smith*, Missouri law created a derivative cause of action that allowed shareholders to enjoin corporations from purchasing unlawful bonds.[48] Smith sought to enjoin the corporation from purchasing bonds authorized by the Federal Farm Loan Act of 1916 ("the

---

42. *See* Louisville & Nashville R.R. Co. v. Mottley, 211 U.S. 149, 152 (1908) ("[A] suit arises under the Constitution and laws of the United States only when the plaintiff's statement of *his own cause of action* shows that it *is based upon those laws or that Constitution*." (emphasis added)).

43. *Id.*

44. *See* Cohen, *supra* note 4, at 891 (asserting that Congress had vested the broad authority of the "potential ingredient" test in 1875, but the courts chose to interpret it narrowly to avoid the impractical increase in federal litigation).

45. History has credited Justice Holmes for this rule, but Professor McFarland explained: "The rule of *American Well Works* had been emerging for at least three decades. Even to say the rule was emerging understates its clarity and force, for it sprang full-grown twenty-three years earlier . . . ." McFarland, *supra* note 4, at 6.

46. *See* Am. Well Works Co. v. Layne & Bowler Co., 241 U.S. 257, 260 (1916). Put another way, "the plaintiff [must] contend[] that a federally ordained rule specifically creates her cause of action and establishes her substantive right to a remedy for a violation of that rule." Howard M. Wasserman, *Jurisdiction and Merits*, 80 WASH. L. REV. 643, 695-96 (2005) (citing Mishkin, *supra* note 4, at 165).

47. 255 U.S. 180, 201-02 (1921).

48. *See id.* at 195.

Act").[49] He alleged that the bonds were unlawful because the Act was unconstitutional.[50] So, while Missouri state law created Smith's cause of action, his well-pleaded complaint necessarily raised a question of federal law as an element of his state-law claim. The Court rejected Justice Holmes's test as a test of *exclusion*, holding jurisdiction proper even though Smith asserted no federally created cause of action.[51] The *Smith* Court framed the test very broadly, stating that jurisdiction exists if "the right to relief depends upon the construction or application" of federal law.[52]

Justice Holmes dissented from *Smith*,[53] and the debate has not ended since. From a purely policy standpoint, should the Court have created the second branch? That is, should the "arising under" inquiry stop with the conclusion that the plaintiff has asserted no federal cause of action? Or, should Holmes's test be treated (as it currently is) as a test for inclusion but not exclusion?[54] The next section focuses on this debate, but I hope this background has shed some light on Justice Thomas's suggestion in *Grable*:

> In this case, no one has asked us to overrule [our] precedents and adopt the rule Justice Holmes set forth in *American Well Works Co.* . . . , limiting § 1331 jurisdiction to cases in which federal law creates the cause of action pleaded on the face of the plaintiff's complaint. In an appropriate case, and perhaps with the benefit of better evidence as to the original meaning of § 1331's text, I would be willing to consider that course.[55]

The academic and intra-Court debate about whether to embrace (and how to define) the second branch has not focused on an intent-based search, and for good reason. If we evaluate those "arising under" words with our congressional-intent goggles, it seems we'll find *Osborn*, not Holmes.[56] Everyone agrees that *Osborn*

---

49. *Id.*

50. *Id.*

51. *See id.* at 201-02.

52. *Id.* at 199.

53. *Id.* at 213-14 ("[T]he single ground upon which the jurisdiction of the District Court can be maintained is that the suit 'arises under the Constitution or laws of the United States' . . . . I am of opinion that this case does not arise in that way and therefore that the bill should have been dismissed. It is evident that the cause of action arises not under any law of the United States but wholly under Missouri law.") (Holmes, J., dissenting).

54. Merrell Dow Pharm., Inc., v. Thompson 478 U.S. 804, 809 n.5 (1986) ("It has come to be realized that Mr. Justice Holmes' formula is more useful for inclusion than for the exclusion for which it was intended.") (quoting T. B. Harms Co. v. Eliscu & Jungnickel, Inc., 339 F.2d 823, 827 (2d Cir. 1964)).

55. Grable & Sons Metal Prods., Inc. v. Darue Eng'g & Mfg., 545 U.S. 308, 320 (2005) (Thomas, J., concurring) (citation omitted).

56. Justice Gray observed:

> The intention of Congress is manifest, at least as to cases of which the courts of the several States have concurrent jurisdiction, and which involve a certain amount or value, to vest in the Circuit Courts of the United States full and effectual jurisdiction, as contemplated by the

cannot be the test.[57] Sure, there has been an occasional suggestion that maybe (just maybe) Congress meant something different in 1875 when it chose the same words.[58] But such suggestions seem reminiscent of those advanced in debates about jurisdiction-stripping—smart and creative people, when faced with an unwelcome result, often propose solutions that barely pass the straight-face test in the hope of giving the Court a way to reach the desired result.

The answer will not, and should not, come from the Court. After *Merrell Dow*, the Court was squarely presented with the opportunity to reject the second branch in favor of the Holmes test.[59] It declined and instead affirmed the second branch in *Grable*.[60] Aside from Justice Thomas's invitation to examine original intent,[61] there is no realistic hope that the Court will revisit that choice. No inquiry into congressional intent or the most reasonable meaning of the words in context will justify deleting the second branch.[62] The Court's credibility would be weakened by changing "arising under" to the Holmes interpretation under the guise of purpose or intent. As Professor Forrester noted long ago, "the solution is not to be found in misreading the law . . . but in passing a new statute with words that actually do limit the amount of federal question jurisdiction granted to the trial courts."[63]

WHY CONGRESS SHOULD ELIMINATE THE SECOND BRANCH

Debate about the second branch has a long and rich history.[64] It is a familiar debate about rules versus standards,[65] power allocation, and parity.[66] Justice

---

Constitution, over each of the classes of controversies above mentioned . . . .
Forrester, 16 TUL. L. REV. 362, *supra* note 4, at 377 (quoting *In re* Hohorst, 150 U.S. 653, 659 (1893)). *See also* Frankfurter, *supra* note 33, at 44 ("In the Act of March 3, 1875, Congress gave the federal courts the whole sweep of power which had lain dormant in the Constitution since 1789.").

57. *See* sources cited *supra* note 35.

58. For instance, Justice Frankfurter commented that such a dramatic change in federal jurisdiction would not have occurred with so little debate. Romero v. Int'l Terminal Operating Co., 358 U.S. 354, 366-67 (1959). *See* Alleva, *supra* note 4, at 1493 n.46.

59. Indeed, except as to a sliver of second branch cases involving federal laws that themselves create a cause of action, *Merrell Dow* is best read as returning to the Holmes test. Rory Ryan, *No Welcome Mat, No Problem?: Federal-Question Jurisdiction After* Grable, 80 ST. JOHN'S L. REV. 621, 634 (2006).

60. *Grable*, 545 U.S. at 311-12.

61. *Id.* at 320-22 (Thomas, J., concurring).

62. Of course, these same mechanisms also fail to justify the current definition, but stare decisis does justify it. *See* Merrell Dow Pharm., Inc., v. Thompson 478 U.S. 804, 820 (1986) (Brennan, J., dissenting) ("The continuing vitality of *Smith* is beyond challenge.").

63. Forrester, *supra* note 4, 18 TUL. L. REV. 263, at 288. Professor Forrester also insisted that "the solution is not to be found, after all, in arguing about the meaning of [the jurisdictional statute]." *Id.* at 287.

64. *See supra* notes 4-6.

65. Professor Alexander has articulated the distinction between rules and standards as

Holmes's test is a rigid *rule*, finding jurisdiction only when federal law creates the plaintiff's cause of action.[67] In contrast, starting with *Smith*,[68] many courts and commentators have decided that some lawsuits involving certain federal issues based on state-created causes of action deserve trial-level resolution in the federal courts.[69] The second branch exists because some believe that identifying which issues deserve federal jurisdiction is a process not susceptible to a rule, but instead requires resort to a more flexible standard. But fuzzy, case-specific standards generally are poorly suited for jurisdictional inquiries.[70] They create too much threshold litigation about where to litigate and provide excessive opportunities for delay. The question is whether, in the "arising under" arena, the benefits outweigh the costs imposed by a flexible standard. I think not.

As a preliminary matter, it is important to address a familiar refrain by proponents of the second branch: that the standard will, through judicial construction, develop boundaries that eliminate (or substantially mitigate) the costs traditionally associated with unclear jurisdictional rules. In other words, the standard will become clear enough through precedent. Professor Cohen, for example, was one such proponent, predicting in his 1967 *Broken Compass* article that clear jurisdictional standards would ultimately emerge, despite the fact that the second branch had been baffling courts and commentators for more than forty years following *Smith*.[71] This prediction proved wrong. Then in 1986, the Court

---

follows:

> A "rule" is a norm whose application turns on the presence of relatively noncontentious facts, and turns on the presence of those facts regardless whether the values that the rule is designed to serve are actually served or disserved by the particular application. Rules are often described as "bright-line" (clear and easy to follow), "formal" (to be applied without regard to substance of the results but only with regard to the rule's terms), and "opaque" (to the rules' background justifications).
>
> Standards are norms that have the opposite characteristics. A standard can be applied only by engaging in evaluation. Therefore, to the extent that evaluation is contentious and uncertain, standards will be as well. Standards are thus vague, substantive (as opposed to formal), and transparent (to background values).

Larry Alexander, *Incomplete Theorizing: A Review Essay of Cass R. Sunstein's* Legal Reasoning and Political Conflict, 72 NOTRE DAME L. REV. 531, 541 (1997). *See also* Freer, *supra* note 4, at 311 (asserting that the Court implicitly recognizes the difference between a rule and a standard); Preis, *supra* note 4, at 168-69 (noting the impact of form on the substance of a legal directive).

66. *See generally* Burt Neuborne, *The Myth of Parity*, 90 HARV. L. REV. 1105 (1977) (disputing the assumption that federal and state trial courts are equally competent forums for the enforcement of federal constitutional rights).

67. Am. Well Works Co. v. Layne & Bowler Co., 241 U.S. 257, 260 (1916).

68. Smith v. Kansas City Title & Trust Co., 255 U.S. 180 (1921).

69. *See, e.g.*, Preis, *supra* note 4, at 159-62 (discussing the context in which issues are embedded in federal law and the types of laws commonly embedded).

70. *See infra* Part III.B.

71. Professor Cohen asserted:

> It may be objected that recognition of the pragmatic nature of the decision whether a claim arises directly under federal law will lead to an ad hoc, unpredictable, case-by-case

revisited the test in *Merrell Dow*.[72] More predictions of clarity came, and the result was a three-way circuit split.[73] Realizing that the "arising under" compass was still broken, the Court decided *Grable* in 2005, and again there was[74] and is[75] optimism that clear boundaries will develop.

Eighty-five years of assurances is enough; we should learn our lesson. The nature of the second branch prevents those predictions from coming true. The Holmes rule is criticized as being too rigid in an area that requires flexibility.[76] I do not find persuasive the argument that a flexible standard, which is needed because a rigid rule will not work, will develop sufficient rigidity through litigation. Standards tend to "collapse decisionmaking back into the direct application of the background principle or policy to a fact situation."[77] As Professor Freer has suggested, it should come as no surprise when standards do not generate bright-line results.[78] Maybe the uncertainty is worth the cost, but we

---

decision of jurisdictional questions. It goes without saying that it is undesirable for jurisdictional rules to be uncertain. Particularly since objections to jurisdiction of the district court cannot be waived, and since in many cases the lack of jurisdiction can even be asserted by the party who invoked federal jurisdiction, there should not be doubt about the threshold question of jurisdiction. . . .

More important, recognition of pragmatic factors and decisions based on them will lead to predictable jurisdictional standards. Thus, no matter how close the pragmatic judgment in a particular case, once made it is bound to decide more than just the case before the court. In other words, the process is not simply case-by-case decision making, with each case standing on its own bottom, but rather a process of clarifying jurisdictional uncertainty in classes of cases before the court. It is, of course, true that a case may be so unique that a jurisdictional decision has no impact on other cases. Very often, however, an authoritative decision of a novel problem of federal question jurisdiction settles the issue for a class of cases. Shoshone Mining Co. v. Rutter relegated a large group of miners' claims to the state courts. Until the Declaratory Judgment Act, American Well Works placed suits by alleged patent infringers in the state courts. Smith v. Kansas City Title & Trust Co. established a general jurisdictional rule for constitutional challenges through the mechanism of the stockholder's derivative suit. And so on.

Cohen, *supra* note 4, at 908-09 (citations omitted).

72. Merrell Dow Pharm., Inc. v. Thompson, 478 U.S. 804 (1986).

73. McFarland, *supra* note 4, at 38 ("While *Merrell Dow* did not uphold federal jurisdiction, it did confuse this area of law for twenty years to the extent it produced a three-way circuit conflict.").

74. Indeed, my first impressions of the *Grable* test were more reserved than they are now. In my previous *Grable* article, which was entirely doctrinal, I noted that while *Grable* does not provide a bright line, it "creates a workable structure" and is a "welcome change." Ryan, *supra* note 59, at 653. I have obviously changed course as my focus turned from doctrinal to normative.

75. *See, e.g.*, Freer, *supra* note 4, at 344 ("At the end of the day, the statutory definition of 'arising under' is in better shape now than it has been in a generation, which should put to rest any latter-day calls for a return to the Holmes test for centrality.").

76. *See, e.g., id.* at 322.

77. *Id.* at 320 (quoting Kathleen M. Sullivan, *Foreword: The Justices of Rules and Standards*, 106 HARV. L. REV. 22, 58 (1992)).

78. *Id.*

should candidly acknowledge another inevitability of the second branch's continued reign—in twenty years, another wave of articles will promise that some new clarification by the Court will establish clear boundaries where today's test has failed.[79]

In a perfect world, each year there would be a handful of cases that the federal trial courts would adjudicate despite the absence of a federal cause of action. These cases would come wrapped with an easy-to-identify bow, which the jurisdiction fairy would attach to cases that satisfy the criteria of the second branch. Unfortunately, removal, remand, hearings, and delay replace bows and fairies. The real-world problem is identifying that handful of cases that ultimately satisfies the second branch.[80] Eliminating the second branch may not yield the best result for that handful. Nevertheless, of the following two options, I'll choose the first: (1) leave that handful and many others for prompt initial resolution in state courts; or (2) retain a flexible test that delays a disproportionate number of cases, for many months, just to find each case worth retaining[81] And as already described, the nature of the second-branch standard forecloses a seemingly attractive third option: better define the test in order to identify the handful without delaying the rest.

In the following subsections, I begin by briefly identifying the test set forth in *Grable*.[82] Next, I argue, pragmatically, why the *Grable* test is poorly suited for a jurisdictional inquiry. To supplement my conceptual critique of the standard, I examine two modest samples of post-*Grable* cases in the removal context to illustrate the practicality of its effects.

THE BOUNDARIES OF THE SECOND BRANCH

The second branch is amorphous by nature. Justice Holmes thought that there was only one branch of "arising under" jurisdiction—when the plaintiff asserted a cause of action created by federal law.[83] The Court has rejected this test as one of

---

79. Professor McFarland recently addressed this trend:

> This question is relatively narrow in the field of federal jurisdiction, yet the Supreme Court of the United States—not to mention lower federal courts—has returned to this issue again and again, sometimes answering yes and more often answering no. On each return, the Court's analysis grows more complex rather than more precise.

McFarland, *supra* note 4, at 1-2 (citations omitted).

80. *See id.* at 40 (evaluating the sparse number of cases in which jurisdiction was sustained under the second branch after appellate review).

81. *See infra* Part III.B (explaining how the flexible nature of the second branch standard facilitates excessive delay in the modern litigation environment).

82. *See* Grable & Sons Metal Prods., Inc. v. Darue Eng'g & Mfg., 545 U.S. 308 (2005). There is no need to linger over the scope of the second branch; plenty has been written on this topic, and the material characteristics of the test are not in dispute. *See, e.g.*, Hoffman, *supra* note 3, at 292-308; McFarland, *supra* note 4, at 3-22; Pozner, *supra* note 4, at 534-76; Preis, *supra* note 4, at 149-66; Pushaw, *supra* note 4, at 1520-27; Howard M. Wasserman, *Jurisdiction, Merits, and Substantiality*, 42 TULSA L. REV. 579, 587-91 (2007).

83. Am. Well Works Co. v. Layne & Bowler Co., 241 U.S. 257, 260 (1916).

exclusion. Sometimes resolving federal issues is necessary to adjudicate a state-created cause of action,[84] and in *Smith* the Court determined that the presence of certain types of these federal issues justified "arising under" jurisdiction.[85] It is determining which types, and in which cases, that has proven troublesome. In *Smith,* for example, the Court defined the test too broadly, writing that federal jurisdiction would be appropriate if "the right to relief depends upon the construction or application" of federal law.[86] Other buzzwords soon appeared, clarifying that the federal issue must be "important" and "substantial."[87] This type of qualitative assessment continues today.[88] It is no surprise, then, that the *Grable* Court, like those before it, crafted a flexible, fuzzy test to determine whether a case satisfied this type of qualitative assessment. So long as such a standard is the type of final determination sought, no other option is possible, and the Court has been quite forthcoming in so admitting.[89]

Grable was a second-branch case involving an embedded federal tax issue within a state claim to quiet title.[90] To satisfy a tax delinquency, the IRS seized some of Grable's real property.[91] The IRS then sold the property to Darue and gave Darue a quitclaim deed.[92] Five years later, Grable brought an action to quiet title against Darue in state court.[93] Grable conceded that it had received actual notice of the seizure but claimed that Darue's recorded title was invalid because the IRS had not strictly complied with the applicable notice provisions, which Grable contended required personal service.[94] Darue removed the case to federal court, arguing that Grable's claim, while created by state law, contained an embedded federal issue, namely the interpretation of the federal tax statute's notice provision.[95]

---

84. *See, e.g.*, Merrell Dow Pharm., Inc. v. Thompson, 478 U.S. 804, 809 n.5 (1986) ("It has come to be realized that Mr. Justice Holmes' formula is more useful for inclusion than for the exclusion for which it was intended.") (quoting T. B. Harms Co. v. Eliscu & Jungnickel, Inc., 339 F.2d 823, 827 (2d Cir. 1964)).

85. Smith v. Kansas City Title & Trust Co., 255 U.S. 180, 199-201 (1921).

86. *Id.* at 199.

87. *See Merrell Dow*, 478 U.S. at 806-07 (quoting Thompson v. Merrell Dow Pharm., Inc., 766 F.2d 1005, 1006 (6th Cir. 1985)); Franchise Tax Bd. of Cal. v. Constr. Laborers Vacation Trust for S. Cal., 463 U.S. 1, 13 (1983); Gully v. First Nat'l Bank, 299 U.S. 109, 118 (1936).

88. *See* Grable & Sons Metal Prods., Inc. v. Darue Eng'g & Mfg., 545 U.S. 308, 314 (2005) (defining the question as whether a state claim "necessarily raise[s] a stated federal issue, actually disputed and substantial").

89. *See id.* at 317; *cf.* Textile Workers Union of Am. v. Lincoln Mills of Ala., 353 U.S. 448, 470 (1957) (noting that "[t]he litigation-provoking problem has been the degree to which federal law must be in the forefront of the case and not collateral, peripheral or remote").

90. *Grable*, 545 U.S. at 310.

91. *Id.*

92. *Id.* at 310-11.

93. *Id.* at 311.

94. *Id.* at 310-11.

95. *Id.* at 311.

Although the Court's 1986 *Merrell Dow* opinion had cast doubts upon the vitality of the second branch,[96] *Grable* reaffirmed[97] and redefined it.[98] After *Grable*, the second-branch question is whether "a state-law claim necessarily raise[s] a stated federal issue, actually disputed and substantial, which a federal forum may entertain without disturbing any congressionally approved balance of federal and state judicial responsibilities."[99] The new definition can be helpfully broken into four prongs: (1) necessity, (2) actually disputed, (3) substantiality, and (4) disruptiveness. My focus is on the last two prongs.

"Substantiality" has long been a buzzword in the "arising under" realm.[100] Justice Cardozo directed us to evaluate the kaleidoscopic situations to find the substantial issues.[101] Factors are articulated for analyzing the substantiality inquiry, but the factors are as amorphous as the overall test.[102] Again, that is the nature of a standard instead of a rule.[103] For example, the Court has often evaluated the nature of the federal issue and whether that issue is "important."[104] Similarly, the Court has asked whether this federal issue requires special resort to federal expertise or uniformity of construction.[105] The standard and its factors can be tweaked indefinitely, and no more clarity will arise. As Justice Brennan noted long ago, the test seems *designed* to be broad enough to allow a court to simply step back and make the ultimate determination whether a federal court should hear the case: "if one makes the test sufficiently vague and general,

---

96. *See* ERWIN CHEMERINKSY, FEDERAL JURISDICTION 293 (5th ed. 2007).

97. *See Grable*, 545 U.S. at 319-20.

98. An extended discussion of the second branch's evolution is not needed for this article. For a more complete evolutionary analysis, see Ryan, *supra* note 59.

99. *Grable*, 545 U.S. at 314.

100. *See, e.g.*, sources cited *supra* note 87.

101. *Gully*, 299 U.S. at 117.

102. *See, e.g.*, Mikulski v. Centerior Energy Corp. 501 F.3d 555, 570 (6th Cir. 2007). The *Mikulski* court summarized:

> The Supreme Court has identified four aspects of a case or an issue that affect the substantiality of the federal interest in that case or issue: (1) whether the case includes a federal agency, and particularly, whether that agency's compliance with the federal statute is in dispute; (2) whether the federal question is important . . . ; (3) whether a decision on the federal question will resolve the case . . . ; and (4) whether a decision as to the federal question will control numerous other cases . . . .

Id. (citing Empire Healthchoice Assurance, Inc. v. McVeigh, 547 U.S. 677, 700 (2006)).

103. *See* sources cited *supra* note 65.

104. *See, e.g.*, *Grable*, 545 U.S. at 315 ("The meaning of the federal tax provision is an *important* issue of federal law that sensibly belongs in a federal court." (emphasis added)); Merrell Dow Pharm., Inc. v. Thompson, 478 U.S. 804, 815 n.12 (1986) ("In *Smith*, as the Court emphasized, the issue was the constitutionality of an *important* federal statute." (emphasis added) (citing Smith v. Kansas City Title & Trust Co., 255 U.S. 180, 201 (1921))).

105. *See Grable*, 545 U.S. at 312 ("[A] federal court ought to be able to hear claims recognized under state law that nonetheless turn on substantial questions of federal law, and thus justify resort to the experience, solicitude, and hope of uniformity that a federal forum offers on federal issues . . . .").

virtually any set of results can be 'reconciled' [with a post hoc analysis]."[106] Put another way, one might say that the words mean "everything and nothing at the same time."[107]

The Court's disruptiveness prong also deserves mention. In *Merrell Dow*, the Court wrote an opinion that suggested a federal cause of action was needed, nearly eliminating the second branch.[108] But in *Grable*, the Court told us what *Merrell Dow* really meant—that a federal cause of action was not required in every case but was merely a "welcome mat" that was needed when exercising jurisdiction would disrupt Congress's approved balance of federal-state jurisdiction.[109] Exercising jurisdiction is disruptive (and thus a welcome mat is needed) when it "herald[s] a potentially enormous shift of traditionally state cases into federal courts."[110]

This disruptiveness prong is as elusive as the substantiality prong.[111] For example, in *Grable*, the Court concluded that no welcome mat was needed because only "the rare state quiet title action [would] involve[] contested issues of federal law."[112] The Court distinguished *Merrell Dow* because allowing a garden-variety tort claim to incorporate federal law would authorize a "horde of original filings."[113] Aside from the introduction of another vague standard, an additional problem with the disruptiveness prong concerns its level of generality.[114] In *Grable*, the Court exercised jurisdiction because it viewed the class of cases narrowly—as quiet-title claims with embedded federal tax issues, not just any federal issues.[115] Of course, *Merrell Dow* is more disruptive when we generalize negligence per se as a garden-variety tort claim and cast the embedded federal drug-labeling issue as "federal law."[116] Choosing the level of generality affects the label. What if *Grable* had been framed as a quiet-title claim with an embedded question of federal law rather than tax law? On the other hand, what if *Merrell Dow* had been framed as negligence per se with embedded drug-labeling provisions rather than state tort laws with embedded federal issues? And what about the vast levels of generality in between? Furthermore, as Professor

---

106. *Merrell Dow*, 478 U.S. at 822 n.1.

107. W. Wendell Hall, *Standards of Review in Texas*, 38 ST. MARY'S L.J. 47, 59 (2006) (describing the boundaries of the abuse of discretion standard of review) (citing Landon v. Jean-Paul Budinger, Inc., 724 S.W.2d 931, 935 (Tex. App. 1987)).

108. *See* CHEMERINSKY, *supra* note 96, at 293.

109. *See Grable*, 545 U.S. at 318.

110. *Id.* at 319.

111. *See* Adam P.M. Tarleton, *In Search of the Welcome Mat: The Scope of Statutory Federal Question Jurisdiction After* Grable & Sons Metal Products, Inc. v. Darue Engineering & Manufacturing, 84 N.C. L. REV. 1394, 1399-1400 (2006).

112. *Grable*, 545 U.S. at 319.

113. *Id.* at 318.

114. *See* Carl A. Auerbach, *A Revival of Some Ancient Learning: A Critique of Eisenberg's* The Nature of the Common Law, 75 MINN. L. REV. 539, 562-63 (1991).

115. *Grable*, 545 U.S. at 315, 319.

116. *See* Merrell Dow Pharm., Inc. v. Thompson, 478 U.S. 804, 809-10 (1986).

Hoffman has noted, the substantiality prong and disruptiveness prong often "pull in opposite directions."[117] He explains:

> That is, if it has been determined that a question of federal law is substantial because its resolution will impact a wide range of persons and behavior, the consequence of shifting so much of the state caseload into federal court will often be that such a profound federalism impact will not be understood to be in accordance with the legislative judgment.[118]

The *Grable* test is narrow, and ultimately the lower courts have rejected far more *Grable* cases than they have accepted.[119] Despite meager success among those invoking jurisdiction under the second branch, the attempts continue.[120] There is enough flexibility in the test to provide nonsanctionable arguments for jurisdiction, which has unsurprisingly resulted in many *Grable* removals.[121] The test is amorphous; it must be to meet its end. But this end is not important enough to justify the delay.

FLEXIBILITY AND INEFFICIENCY

"Jurisdictional rules should be clear."[122] As Judge Posner eloquently summarized:

> Functional approaches to legal questions are often, perhaps generally, preferable to mechanical rules; but the preference is reversed when it comes to jurisdiction. When it is uncertain whether a case is within the jurisdiction of a particular court system, not only are the cost and complexity of litigation increased by the necessity of conducting an inquiry that will dispel the uncertainty but the parties will often find themselves having to start their litigation over from the beginning, perhaps after it has gone all the way through to judgment. "Jurisdictional rules ought to be simple and precise so that judges and lawyers are spared having to litigate over not the merits of a legal dispute but where and when those merits shall be litigated." . . . "The more mechanical the application of a jurisdictional rule, the better. The chief and often the only virtue of a jurisdictional rule is clarity."[123]

---

117. Hoffman, *supra* note 3, at 300.

118. *Id.*

119. *See infra* Part III.B.

120. *See infra* Part III.B.

121. *See infra* notes 128-129 and accompanying text.

122. Grable & Sons Metal Prods., Inc. v. Darue Eng'g & Mfg., 545 U.S. 308, 321 (2005) (Thomas, J., concurring). *See* McFarland, *supra* note 4, at 46 (noting that a "bright jurisdictional line allows predictability, stability, and efficiency"); *Mr.* Smith *Goes to Federal Court: Federal Question Jurisdiction over State Law Claims Post-*Merrell Dow, 115 HARV. L. REV. 2272, 2277-84 (2002) (suggesting that a clear rule is desirable to promote judicial economy and predictability).

123. Hoagland v. Sandberg, Phoenix & Von Gontard, P.C., 385 F.3d 737, 739-40 (7th Cir. 2004) (citations omitted) (citing *In re* Lopez, 116 F.3d 1191, 1194 (7th Cir. 1997); *In re* Kilgus,

Jurisdictional rules should be clear.[124] It is easy to state the principle but to then gloss over it and migrate toward functional standards whenever we encounter a situation that, when viewed in isolation, seems to warrant a different result. For example, the question arises, why, in *Grable*, should not a federal trial court be available to resolve tax issues? The proper perspective does not consider *Grable* in isolation. Nor does it stop when it uncovers a handful of Supreme Court cases where initial federal jurisdiction seems warranted. Rather, the proper perspective requires us to look systemically at the cost of allowing a jurisdictional inquiry capable of sorting out the cases.

The flexible nature of the second-branch standard facilitates excessive delay in the modern litigation environment. Typically, second-branch cases involve a plaintiff suing in state court, trying to keep the case there by asserting only state-created causes of action. Defendants typically prefer a federal forum (or, cynically, desire the opportunity for delay) and therefore remove the case, arguing that the presence of a federal issue satisfies the second branch. There is no judge or jurisdiction scholar acting as gatekeeper; rather, the defendant creates delay by simply noticing removal. Then comes the motion to remand, a hearing, and substantial delay.

Whether or not delay is the end sought, it is the result. The flexibility of the test does facilitate intentional delay, if that is the goal. Considerations of soft standards such as substantiality, importance, and disruptiveness create a significant area of nonsanctionable argument, and nonsanctionable arguments for removal are an invitation for delay. Courts may not even grant attorney fees to the plaintiff after remand unless "the removing party lack[s] an objectively reasonable basis for seeking removal,"[125] and the boundaries of substantiality have eluded the Court and legal academy for eighty-five years.[126] But even if the goal is not delay in itself, delay is rarely the enemy of the defendant, so even a slight chance of success in obtaining the federal forum comes with a delay-bonus. Importantly then, the delay-prone class of cases is broader than the class of cases that ultimately qualify for second-branch jurisdiction. The modest samples that appear on the following pages reveal that few cases will ultimately satisfy the second-branch inquiry. But the systemic efficiency concern is not limited to those few – rather, it is the delay of the larger number of colorably removable cases that are delayed in order to find the few.

To evaluate the success of second branch cases after *Grable,* we[127] considered two samples of post-*Grable* cases, focusing on the removal-remand scenario. The two samples consist of opinion-generating second-branch removal cases: the first sample comprises those cases decided during the seven months

---

811 F.2d 1112, 1117 (7th Cir. 1987)).

124. *Grable*, 545 U.S. at 321 (Thomas, J., concurring).

125. Martin v. Franklin Capital Corp., 546 U.S. 132, 141 (2005).

126. *See* sources cited *supra* note 4.

127. My research assistant, Jeff Fisher, deserves credit for his work on compiling and evaluating the samples. Therefore, in this portion of the article, "we" is the appropriate pronoun.

preceding this Article,[128] and the second sample comprises those cases decided in the seven months immediately following *Grable*.[129] Our focus was merely the success rate and delay in the common removal-remand context; it was not to quantify how many second-branch cases exist. The most important limitation of our sample is that it includes only opinion-generating cases. This limitation is important because a district court's remand to a state court for lack of subject matter jurisdiction is not reviewable on appeal[130] and because many district courts do not routinely publish remand orders.[131] So when a court remands for want of subject-matter jurisdiction, there may be no opinion accompanying remand, and there is of course no appellate opinion because the remand is unreviewable. While it would be exceedingly difficult to quantify the number, it is safe to say that most of the cases that were excluded because they generated no opinion were cases that were removed (delayed) and remanded.[132] We also excluded from our samples cases filed in federal court because such cases were rare and we wanted to evaluate the average delay in the common removal-remand scenario.[133]

The first, more recent sample generated fifty-nine opinions in removed second branch cases.[134] Jurisdiction was sustained in seven (11.9%), and fifty-two were remanded (88.1%). For those remanded cases, the average time consumed from removal until remand was 171 days, or nearly 6 months.[135] Roughly, then, to find each case in which jurisdiction was proper under the

---

128. Appendix A.

129. Appendix B.

130. 28 U.S.C. § 1447(d) (2000). *See also* Quackenbush v. Allstate Ins. Co., 517 U.S. 706, 711-12 (1996) (stating that only remands based on lack of subject matter jurisdiction or defects in removal procedure are immune from review under § 1447(d)).

131. Thus, when a court remands for lack of subject matter jurisdiction, there may be (1) no opinion accompanying remand and (2) no appellate opinion because the remand is unreviewable. *See* Hoffman, *supra* note 3, at 303 (noting that the "most likely distortion produced by reliance only on reported cases is to be underinclusive of decisions that limit the scope of the federal judicial power").

132. If jurisdiction was sustained, it would, of course, be challengeable in the appellate court. However, it is difficult to imagine many second branch cases where the issues would be so clear that neither the loser nor the court would address the second branch issue on appeal.

133. For example, in the more recent sample, our initial count revealed only six cases filed in federal court, and the court dismissed five of them.

134. This sample included four cases wherein removal was on multiple grounds and, after the court clearly declined jurisdiction based on the *Grable* argument, it retained the case on another basis. Am. Sys. Consulting, Inc. v. Devier, 514 F. Supp. 2d 1001 (S.D. Ohio 2007); Barash v. Ford Motor Credit Corp., No 06-CV-6497 (JFB) (ARL), 2007 U.S. Dist. LEXIS 44641 (E.D.N.Y. June 20, 2007); Johnson v. Precision Airmotive, L.L.C., No. 4:07CV1695 CDP, 2007 WL 4289656 (E.D. Mo. Dec. 4, 2007); Kurz v. Fidelity Mgmt. & Research Co., Case No. 07-cv-709-JPG, 2007 WL 3231423 (S.D. Ill. Oct. 30, 2007). The delay in these cases was measured from the date of removal until the date of the opinion denying jurisdiction based on the second branch theory.

135. Even ignoring those cases where the case reached the Court of Appeals before remand, the average delay was 129 days, almost 4 ½ months.

second branch, the courts delayed about eight cases for about six months each. Given the recency of the sample and the high rate of reversal in second branch cases,[136] more of the cases in the "yes" column may be remanded after appeal, thus lowering the acceptance percentage and increasing the average delay time. And again, this 8:1 ratio is likely underinclusive of remanded cases because we focused only on opinion-generating cases.[137]

Our earlier sample and Professor Preiss's findings[138] shed further light on the state of the second branch docket. Our earlier sample showed forty-six opinion-generating removal cases of the same type decided in the seven months immediately following *Grable*. Jurisdiction was sustained in six (13.0%), and forty were remanded (87.0%). The average delay for removed-then-remanded cases was 133 days.[139] While the more recent sample shows a slightly higher remand rate, it also shows an increase in attempts (59 to 46) and a slightly longer delay (171 days to 133). Professor Preiss's findings are also helpful in evaluating the workability of the second branch. He recently studied the published second branch cases that had reached the federal courts of appeal since the *Merrell Dow* decision.[140] One statistic from his article stands out: forty-four out of the sixty-seven studied cases were remanded back to state court *from the appellate court*.[141]

Zenergy, Inc. v. Palace Exploration Co.[142] is a typical case from the remand pile—a pile that makes up about 85% of the larger pile of colorably removable second-branch cases.[143] The state-court plaintiffs in *Zenergy* sought a declaration of the rights of various parties regarding 1,554 oil and gas properties.[144] The defendants removed to district court, claiming that the case necessitated resolving questions of federal tax law.[145] After 209 days,[146] the district court remanded, finding no substantial federal issue in the complaint.[147] The potential delay impact of amorphous jurisdictional standards is even more significant for cases subject to the federal multidistrict litigation ("MDL") procedures because, when a defendant removes a case subject to MDL, the district court need not resolve the motion to remand before transferring the case.[148] "The fact that there [are]

136. *See* Preis, *supra* note 4, at 165-66.

137. *See* Hoffman, *supra* note 3, at 302-03.

138. *See* Preis, *supra* note 4, at 165-66.

139. This period of delay was calculated using 34 of the 40 opinion-generating remand cases, for which the date of removal could be readily ascertained.

140. *See* Preis, *supra* note 5, at 159.

141. *See* Preis, *supra* note 4, at 166.

142. Zenergy, Inc. v. Palace Exploration Co., No. 07-CV-34-GKF-PJC, 2007 U.S. Dist. LEXIS 57288 (N.D. Okla. Aug. 7, 2007).

143. *See supra* previous paragraph.

144. *Zenergy*, 2007 U.S. Dist. LEXIS at *2, *4-5.

145. *Id.* at *2, *4.

146. Notice of removal was filed on January 10, 2007. The final order was entered August 7, 2007. *Zenergy*, 2007 U.S. Dist. LEXIS at *9.

147. *Zenergy*, 2007 U.S. Dist. LEXIS at *9.

148. *See In re* Crown Life Ins. Premium Litig., 178 F. Supp. 2d 1365, 1366 (J.P.M.L. 2001).

pending jurisdictional objections [does] not deprive the MDL panel of the ability to transfer the case."[149] For example, in *In re Pharmaceutical Industry Average Wholesale Price Litigation*,[150] the plaintiff filed in Florida state court on April 5, 2005.[151] Defendants removed on July 20, 2005, alleging jurisdiction under *Grable*.[152] Plaintiff moved to remand on August 18, 2005, but a week later, the case was transferred to the MDL court.[153] Because a similar Florida case had been removed and transferred, a decision on the motion to remand was postponed.[154] Ultimately, on September 6, 2006, almost one and a half years after the initial state court filing, the MDL transferee court remanded the case for want of jurisdiction but refused to impose sanctions because the removal attempt was not unreasonable.[155] While courts could avoid this *additional* delay by ruling on remand motions before transfer, they have the discretion to stay the remand ruling pending transfer, and many courts do so.[156]

To what end are so many cases delayed to find the few? In summary, a defendant who thinks a lawsuit contains a *Grable*-satisfying embedded federal issue is probably wrong, and will be told so almost nine times out of ten, after an average delay of 171 days.[157] As discussed in the next subsection, eliminating the second branch does not impose ominous consequences, either in principle or implementation, and it yields considerable benefits, both in terms of systemic efficiency and state interests in the federal-state balance.

FEDERAL, STATE, AND SYSTEMIC INTERESTS

The most persuasive criticism of abandoning the second branch undoubtedly will be an emphasis on the merits of its existence. Some cases involving important federal issues will not receive a trial in a federal forum, and this result is undesirable for three reasons: (1) federal judges are more experienced and better at applying federal law; (2) the state judges' inexperience will result in a

---

But see Minnesota v. Pharmacia Corp., No. 05-1394, 2005 WL 2739297, at *2 (D. Minn. Oct. 24, 2005).

149. Grispino v. New England Mut. Life Ins. Co., 358 F.3d 16, 19 n.3 (1st Cir. 2004); *see* Yvette Ostolaza & Michelle Hartmann, *Overview of Multidistrict Litigation Rules at the State and Federal Level*, 26 REV. LITIG. 47, 63-64 n.72 (2007) (citations omitted).

150. *In re* Pharm. Indus. Average Wholesale Price Litig., 457 F. Supp. 2d 65 (D. Mass. 2006).

151. *Id.* at 70.

152. *Id.*

153. *Id.*

154. *Id.* at n.3.

155. *Id.* at 65, 76.

156. *See, e.g.*, Franklin v. Merck & Co., No. 06-cv-02164-WYD-BNB, 2007 WL 188264, at *2 (D. Colo. Jan. 24, 2007); Hatch v. Merck & Co., No. 05-1252, 2005 WL 2436716, at *1 (W.D. Tenn. Oct. 3, 2005) ("Although some courts have opted to rule on pending motions to remand prior to the MDL Panel's decision on transfer, there are many more that have chosen to grant a stay, even if a motion to remand has been filed.") (citations omitted).

157. *See supra* note 140 and accompanying text.

lack of uniformity; and (3) state courts may be hostile to federal rights.[158] Further, because of the current volume of litigation, many litigants will be denied *any* federal forum because the Supreme Court cannot be expected to police the entire state docket for important federal issues. Along with the preceding discussion, the following three paragraphs contain my direct response.

My first response is a point made earlier but directly relevant here. Of course it would be desirable if we could identify the "right cases" for the second branch and bring them to federal court without disrupting the entire class of colorably right cases.[159] But that has never been possible, and promises of a clear-enough second branch have never been fulfilled.

The phrase "arising under" has stymied jurists at all levels of the federal judiciary, from district courts to several generations of Supreme Court Justices. As Charles Alan Wright and Arthur Miller noted, "[t]he meaning of this phrase has attracted the interest of such giants of the bench as Marshall, Waite, Bradley, the first Harlan, Holmes, Cardozo, and Frankfurter—to name only the dead."[160]

With the emergence of new federal rights and litigation trends, it has been impossible for precedent to draw reliable lines as to which issues are "substantial." And the nature of the second-branch standard prohibits those lines from developing.[161]

Next, arguments in favor of the second branch often overlook, or at least understate, the countervailing state interest when balancing the federalism concerns that underlie the second branch. Second-branch cases necessarily involve a state-created cause of action. Taking the case from state to federal court allows the federal forum to adjudicate the federal issue but removes a trial from the courts of the sovereign that made the issue actionable.[162] While this jurisdictional allocation provides federal expertise for the disputed federal issues in the case, federal courts assume the role of *Erie*-guessers for the rest of the case, which includes a state cause of action and state substantive issues.[163] And

---

158. *See* Neuborne, *supra* note 66, at 1120-21.

159. *See supra* Part III.B.

160. Donna C. Peavler, *Removing the Removal Mystery: When Work-Related Claims are Removable Under 28 U.S.C. § 1445(c)*, 2 FED. CTS. L. REV. 27, 35 (2007) (citing 13B CHARLES ALAN WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 3562, at 18 (2d ed. 1987)).

161. *See supra* notes 71-79 and accompanying text.

162. *See* Martin H. Redish, *Reassessing the Allocation of Judicial Business Between State and Federal Courts: Federal Jurisdiction and "The Martian Chronicles,"* 78 VA. L. REV. 1769, 1773-74 (1992) (stating that it is the primary duty of a court to develop the law of its own sovereign because a sovereign is an expert in its own laws and is more sensitive and sympathetic to its own issues and courts); *see also* Barry Friedman, *Under the Law of Federal Jurisdiction: Allocating Cases Between Federal and State Courts*, 104 COLUM. L. REV. 1211, 1237 (2004) (asserting that "only the courts of the sovereign . . . can render an authoritative interpretation of that sovereign's laws"); Ronald J. Greene, *Hybrid State Law in the Federal Courts*, 83 HARV. L. REV. 289, 293 (1969) (stating that the federal government has little interest in having its courts adjudicate second branch cases).

163. *See* Preis, supra note 4, at 199 ("Put simply, hybrid law cases almost always involve more state law than federal law.").

unlike the Supreme Court's ability to review the state court's determination of federal law,[164] the state's high court is unavailable to hear an appeal once federal proceedings end.[165]

The proper allocation must consider federal, state, and systemic interests. Given that (1) the federal court will make an essentially final *Erie* guess if the case is allocated to federal court and (2) there exists at least the potential for the Supreme Court to review if the case is allocated to state court, even if the second-branch inquiry did not result in the delay of so many cases that ultimately fail, the proper allocation of authority between the state and federal courts would be questionable. State courts will rarely create havoc by resolving federal issues.[166] While states outnumber circuits fifty to thirteen, most disputed legal questions have only a few possible answers, and difficult issues often result in splits among the circuits.[167] It seems unlikely that state courts will often diverge from either (1) following a uniform, though disputed, federal interpretation or (2) choosing among already existing disputes among the federal courts. Supreme Court review seems an appropriate safeguard against the truly rare cases that involve a state court diverging in a case that would have actually qualified for second-branch jurisdiction.[168] Ultimately, while the current test results in more frequent jurisdictional litigation, it results in few cases that actually satisfy the test and that therefore yield any benefit of enhanced uniformity. And as explained later, under my proposal, Congress will define those areas in which initial resolution should be more broadly allowed in the federal forum.[169]

Finally, the evolution of the law governing implied causes of action impacts the efficiency gained by adopting the Holmes rule. Not too many years ago, reverting to the Holmes test would have increased efficiency less because of the Supreme Court's treatment of implied causes of action. At this time, the test governing implied causes of action was amorphous, barely connected to statutory construction, and frequently satisfied.[170] Eliminating the second branch then would merely have moved jurisdictional litigation from arguments about the second branch to arguments that the Holmes test is satisfied because federal law

---

164. There is, of course, the rare occasion in which the Court of Appeals may certify a question of state law to the state's high court.

165. *See* Hirshman, *supra* note 4, at 41-42, 64.

166. *See* Preis, *supra* note 4, at 166 (noting that of the forty-four remanded second branch cases, not one resulted in a published state court opinion resolving the federal issue).

167. *See* Merrell Dow Pharm., Inc. v. Thompson, 478 U.S. 804, 826 (1986) (Brennan, J., dissenting).

168. Supreme Court review, of course, does not depend upon 28 U.S.C. § 1331, which is a grant of original jurisdiction. Because the Supreme Court's jurisdiction to review state court decisions comes from 28 U.S.C. § 1257, eliminating the second branch will not impact the issues available for its review.

169. *See infra* Part IV.

170. *See* Cort v. Ash, 422 U.S. 66, 78 (1975) (rejecting the approach of cases such as J. I. Case Co. v. Borak, 377 U.S. 426, 433 (1964)); *see also* Alexander v. Sandoval, 532 U.S. 275, 287 (2001) ("Respondents would have us revert in this case to the understanding of private causes of action that held sway 40 years ago . . . .").

(impliedly) creates the cause of action. The Court has since reined in its implied-cause-of-action jurisprudence and clarified that, "[l]ike substantive federal law itself, private rights of action to enforce federal law must be created by Congress."[171] "Raising up causes of action where a statute has not created them may be a proper function for common-law courts, but not for federal tribunals."[172] The Holmes test today thus provides a much brighter line than it did during the heyday of implied causes of action.

IMPLEMENTATION

For reasons discussed below, Congress should effect this change by adding the following definition to 28 U.S.C. § 1331: "For purposes of § 1331, a civil action arises under federal law when the plaintiff asserts a federal cause of action."

Even assuming one agrees with my conclusions that (1) the Holmes test should govern § 1331 and (2) the change must come from Congress, implementation concerns still exist because the words "arising under" appear frequently throughout the Judicial Code, not just in § 1331.[173] To be clear, my proposal is *not* to remove the words "arising under" from 28 U.S.C. § 1331 and *not* to redefine the phrase for purposes of the entire Judicial Code. It is to redefine the term solely for purposes of 28 U.S.C. § 1331. Ultimately, this redefinition will facilitate a transition with minimal problems and a return to congressional intent.

Several currently superfluous jurisdictional grants appear in the Judicial Code and use the same "arising under" phrase.[174] They are superfluous because they are now consumed by 28 U.S.C. § 1331, which provides that "[t]he district courts

---

171. *Alexander*, 532 U.S. at 286.

172. *Id.* at 287 (quoting Lampf v. Gilbertson, 501 U.S. 350, 365 (1991) (Scalia, J., concurring in part and concurring in judgment)).

173. Redefining the words "arising under" for purposes of the Judicial Code would impact many untargeted provisions in many different contexts. *See, e.g.*, 28 U.S.C. § 1295(a)(1) (2000) ("The United States Court of Appeals for the Federal Circuit shall have exclusive jurisdiction . . . except [for] a case involving a claim *arising under* any Act of Congress relating to copyrights, exclusive rights in mask works, or trademarks . . . .") (emphasis added); 28 U.S.C. § 1338(a) (2000) ("The district courts shall have original jurisdiction of any civil action *arising under* any Act of Congress relating to patents, plant variety protection, copyrights and trademarks.") (emphasis added); 28 U.S.C. § 1339 (2000) ("The district courts shall have original jurisdiction of any civil action *arising under* any Act of Congress relating to the postal service.") (emphasis added); 28 U.S.C. § 1441(b) (2000) ("Any civil action . . . *arising under* the Constitution, treaties or laws of the United States shall be removable without regard to the citizenship or residence of the parties.") (emphasis added); 28 U.S.C. § 1445(d) (2000) ("A civil action in any State court *arising under* section 40302 of the Violence Against Women Act of 1994 may not be removed to any district court of the United States.") (emphasis added); 28 U.S.C. § 1658(a) (2000 & Supp. 2005) ("a civil action *arising under* an Act of Congress . . . may not be commenced later than 4 years after the cause of action accrues") (emphasis added).

174. *E.g.*, 28 U.S.C. §§ 1337(a), 1339, 1340 (2000).

shall have original jurisdiction of all civil actions arising under the . . . laws . . . of the United States."[175] Thus, if a civil action arises under the postal laws of the United States, then § 1331 mandates federal jurisdiction; yet, § 1339 provides that "[t]he district courts shall have original jurisdiction of any civil action arising under any Act of Congress relating to the postal service."[176] The statutes have not always been superfluous. Section 1331, the general federal question statute, previously contained an amount-in-controversy requirement.[177] At one time, then, the significance of a separate "arising under" grant was to allow courts to exercise original jurisdiction in civil actions arising under particular laws without regard to the amount in controversy.[178] In other words, jurisdiction existed for cases arising under the postal laws even if they did not satisfy the amount-in-controversy requirement because that requirement was only present in the general grant of federal-question jurisdiction, not in the postal-laws grant.

Redefining "arising under" for purposes of only 28 U.S.C. § 1331 would eliminate the superfluousness of those other provisions. Under this proposal, the current second-branch inquiry would still apply to the other, non-§ 1331 "arising under" grants (with an important qualification discussed in the following paragraph). With respect to each of the currently superfluous grants, Congress would make a simple choice: either delete the grant, thereby extending the Holmes limitation to those areas, or deliberately retain the grant to use more of its *Osborn* power over particular matters that Congress deems important. Under either approach, federal-question jurisdiction would again resemble a congressionally controlled area. The Holmes test would govern federal question jurisdiction unless Congress retained or created separate "arising under" statutes. Because "arising under" would be narrowly defined only for § 1331, the words would retain their current *Grable* meaning for the rest of the Judicial Code. Congress could therefore retain the current definition in areas of exclusive federal jurisdiction and in other selected areas where Congress decides to use its *Osborn* power to extend jurisdiction.[179] For example, Congress (not the Court) could

175. 28 U.S.C. § 1331 (2000).

176. 28 U.S.C. §§ 1331, 1339 (2000).

177. *See* Pub. L. No. 96-486, 94 Stat. 2369 (1980) (eliminating the amount in controversy requirement) (codified as amended at 28 U.S.C. § 1331 (2000)); H.R. REP. NO. 96-1461, at 1-9 (1980) (proposing the change to eliminate the requirement); Exxon Mobil Corp. v. Allapattah Servs., Inc., 545 U.S. 546, 588 (2005) ("Since 1980, § 1331 has contained no amount-in-controversy requirement.").

178. *See* Lynch v. Household Fin. Corp., 405 U.S. 538, 549 & n.17 (1972) ("Congress has substantially lessened [the amount in controversy's] importance with respect to § 1331 by passing many statutes that confer federal-question jurisdiction without an amount-in-controversy requirement."); Hawaii v. Standard Oil Co. of Cal., 405 U.S. 251, 266 (1972) ("Congress has given private citizens rights of action for injunctive relief and damages for antitrust violations without regard to the amount in controversy." (citing 28 U.S.C. § 1337)).

179. Indeed, Congress is free to extend jurisdiction even further in selected areas, such as by using its Article III jurisdiction-conferring power to override the well-pleaded complaint rule. *See* Mesa v. California, 489 U.S. 121, 136 (1989) ("The removal statute itself merely serves to overcome the 'well-pleaded complaint' rule which would otherwise preclude removal even if a

retain the second branch for patent cases,[180] trademark cases,[181] and certain "Indian claims"[182] where it determines that the cost is justified.

Because the allocation would then be congressionally dependent, the *Grable* disruptiveness inquiry would disappear. Obviously, the disruptiveness inquiry would no longer have a role in the §1331 inquiry.[183] But more importantly, the disruptiveness inquiry would be eliminated even in the retained second-branch areas because Congress would have delineated the areas in which it wants the courts to hear state-created claims raising substantial federal issues. The federal cause of action would no longer be the welcome mat; the welcome mat would be a subject-matter-specific grant of jurisdiction that exists without § 1331's newly imposed limitation.[184]

Nor would this proposal affect the current removal or supplemental-jurisdiction inquiries. By leaving the terms "civil action" and "arising under" in place and by referencing "a federal cause of action," the search for a jurisdictional "hook" under *Exxon v. Allapattah*[185] is unchanged, except to the extent that a second-branch claim would no longer qualify for original jurisdiction under § 1331 and therefore could not be a jurisdictional hook. Nor would amending §

---

federal defense were alleged.").

180. 28 U.S.C. § 1338 (2000).

181. *Id.*

182. 28 U.S.C. § 1505 (2000).

183. Although the amendment would remove the second branch, it would not impact the complete-preemption doctrine. A different article might take on the Court's modern complete-preemption approach. *See* Beneficial Nat'l Bank v. Anderson, 539 U.S. 1, 11 (2003) (Scalia, J., dissenting) (disagreeing with the majority's finding that "the exclusive cause of action for claims of usury against a national bank . . . even if explicitly pleaded under state law" arise under federal law and asserting that such a view is totally unsupported by precedent or act of Congress). But it is easy enough to incorporate the complete-preemption doctrine into the Holmes test. When a state-created cause of action is completely preempted because it falls within the scope of an exclusively federal cause of action, the effectual result is that the plaintiff actually *has* asserted the federal cause of action, for no state cause of action exists. *See* Franchise Tax Bd. of Cal. v. Constr. Laborers Vacation Trust for S. Cal., 463 U.S. 1, 24 (1983) ("[I]f a federal cause of action completely pre-empts a state cause of action any complaint that comes within the scope of the federal cause of action necessarily 'arises under' federal law."); Richard E. Levy, *Federal Preemption, Removal Jurisdiction, and the Well-Pleaded Complaint Rule*, 51 U. CHI. L. REV. 634, 655-57 (1984). As Judge Easterbrook asserted, such an outcome is the federal court correcting the plaintiff's "spelling error" in order to assert the only available cause of action for the complained of conduct. Bartholet v. Reishauer A.G., 953 F.2d 1073, 1075 (7th Cir. 1992).

184. Reversion to the Holmes test will not solve all problems in the declaratory judgment context. *See, e.g.*, Donald L. Doernberg & Michael B. Mushlin, *The Trojan Horse: How the Declaratory Judgment Act Created a Cause of Action and Expanded Federal Jurisdiction While the Supreme Court Wasn't Looking*, 36 UCLA L. REV. 529, 544 (1989); Oakley, *supra* note 4, at 1835-36 (1998). However, the Holmes test should simplify the inquiry slightly in cases in which the hypothetical coercive complaint would contain a second branch claim.

185. *See* Exxon Mobil Corp. v. Allapattah Servs., Inc., 545 U.S. 546, 593 (2005) (Ginsburg, J., dissenting) (discussing the use of "another's claim as a 'hook' to add a claim that the plaintiff could not have brought in the first instance").

1331 disturb the general removal provision in § 1441(a), which requires the remover to find a grant of original jurisdiction and does not use the words "arising under."[186] Amending § 1331 would simply alter one of the original grants and would not alter the inquiry. Nor would any problems arise under § 1441(b)[187] because the only sensible interpretation of "arising under" for purposes of that subsection is to incorporate the meaning of the grant of jurisdiction that satisfied subsection (a).

One foreseeable objection to my proposal is that the words "arising under" would have different meanings in different parts of the Judicial Code. This consequence is true and perhaps a bit odd, but it is probably better than the alternatives. First, I have already discussed why the solution should not come from the Court.[188] Second, although an argument could be made either way, redefining "arising under" for purposes of § 1331 is preferable to removing the words from § 1331 altogether.[189] Removing the words from § 1331 would create no problems within § 1331 where the same meaning could simply be enacted without definition. But removing the phrase from § 1331 creates problems because of § 1331's interaction with other statutes that use "arising under" in the Judicial Code.[190] For example, the removal statute references cases "arising under" federal law and would therefore need to be amended if those words were removed from, rather than defined within, § 1331.[191] Also, because Congress would likely (and should) leave in place some of the newly broader "arising under" grants, future statutes could conveniently reference "arising under" jurisdiction and thereby incorporate both the broader grants and the narrower "arising under" definition from § 1331.[192] An undesirably third alternative would be to define "arising under" for the entire Judicial Code rather than just for

---

186. 28 U.S.C. § 1441(a) (2000).

187. 28 U.S.C. § 1441(b) (2000).

188. *See supra* Part II.

189. For example, a version of § 1331 without "arising under" language might read as follows: "The district courts shall have original jurisdiction over civil actions in which federal law creates a cause of action asserted by the Plaintiff."

190. *See, e.g.*, statutes cited *supra* note 173.

191. 28 U.S.C. § 1441(b) (2000) ("Any civil action of which the district courts have original jurisdiction founded on a claim or right arising under [federal law] shall be removable without regard to the citizenship or residence of the parties.").

192. Additionally, § 1331's interaction with the supplemental jurisdiction statute in § 1367 supports leaving the words "arising under" and "civil action" alone. *See* 28 U.S.C. § 1367(a) (2000) ("[I]n any civil action of which the district courts have original jurisdiction, the district courts shall have supplemental jurisdiction over all other claims that are so related to claims in the action within such original jurisdiction that they form part of the same case or controversy under Article III of the United States Constitution."). Changing "civil action" to "claim," or otherwise incorporating the Holmes test by removing the words "arising under," would create difficult interpretation problems given the claim-specific approach the Court has taken to § 1367's ambiguous language. *See* Exxon Mobil Corp. v. Allapattah Servs., Inc. 545 U.S. 546, 554 (2005) ("In contrast to the diversity requirement, most of the other statutory prerequisites for federal jurisdiction, including the federal-question and amount-in-controversy requirements, can be analyzed claim by claim.").

§ 1331. But this alternative would reach untargeted areas[193] and cause unnecessary confusion.

Under my proposal, no difficult interpretation issues should arise, as congressional intent to restrict "arising under" only for purposes of § 1331 would be made clear by a definition that is limited to only that provision. Such an approach is not novel, for legislatures often define terms for purposes of one statutory section only.[194] And in any event, it is not as though this definition confuses a clear area. We already have redundant jurisdictional grants and words with different meanings depending on the context, all of which are inconsistent with what the words mean under any approach to statutory construction. For now, I prefer the redefining approach; however, I would not-so-grudgingly accept Congress's decision to remove "arising under" from § 1331, so long as proper study is given to the resulting impact on other statutes intended to reference § 1331 with the phrase "arising under."

CONCLUSION

The second branch should be eliminated for purposes of the general federal question statute. The clamor for bright-line rules is often misplaced. In many contexts, legal rules ought to be fluid, flexible, and adaptable. However, such flexible standards are generally poorly suited for jurisdictional inquiries where flexible standards create too much litigation about where to litigate. Justice Holmes adopted a bright-line rule governing when a case "arises under" federal law. Unsurprisingly, cases arose that seemed to fall on the wrong side of the Holmes test; bright-line rules have that effect. But before abandoning the bright-line approach, consideration must be given to the benefits of its existence and, relatedly, the costs of its removal. By eliminating that bright line in favor of flexible standards such as substantiality, importance, and now disruptiveness, the Court has created a class of delay-prone cases that is much broader than the class of cases ultimately benefitting from the use of those standards. Colorably removable cases will be removed, and the second branch, though difficult to satisfy, creates a broad class of colorably removable cases. Even without considerations of delay, policy does not overwhelmingly favor federal adjudication of cases that satisfy the current second-branch test. Second-branch cases are necessarily hybrid cases, containing issues of both state and federal law, and are made actionable only by a state creating a cause of action. If the case proceeds in state court, review (though limited) is available in the Supreme Court. But if the case is removed, federal courts decide the federal issue and are forced

193. For example, the phrase is used in provisions pertaining to habeas corpus, 28 U.S.C. § 2261(a) (2000), the federal circuit courts' appellate jurisdiction, 28 U.S.C. § 1295(a) (2000), and disbursement and certification of officers, 28 U.S.C. § 613(c)(2) (2000).

194. *See, e.g.*, 26 U.S.C.A § 45N (Supp. 2008) (defining "qualified mine rescue team employee," "eligible employer," and "wages"); TEX. PROB. CODE ANN. § 450(c)(1) (Vernon 2003) (defining "employees' trust").

to become essentially final *Erie* guessers as to the statue issues. The limited benefits for the few do not justify the costs on the many.

Congress can effect this change with minimal impact on the rest of the Judicial Code, and the 28 U.S.C. § 1331 inquiry will begin to resemble an exercise in statutory construction. The best approach is probably to define "arising under" solely for purposes of § 1331, the general federal question statute. The Holmes test would thus govern only § 1331 and not other "arising under" provisions. As to the other "arising under" provisions, Congress would choose whether to remove them in order to bring those matters within the Holmes test or to keep them, therefore keeping those matters within the current "arising under" realm (including the Second Branch). Future statutes could conveniently reference "arising under" and incorporate both types. Because the allocation would now be congressionally dependent, the *Grable* disruptiveness inquiry would no longer be needed. Obviously it would not govern cases under the amended § 1331. In the broader "arising under" areas, there would be no need to search for a federal-cause-of-action welcome mat—the welcome mat would be a subject-matter-specific grant of jurisdiction that exists without § 1331's limitation.

We have tried looking through kaleidoscopes, and we have tried fixing our compass. At each step, assurances are given that this time the boundaries will be acceptably clear (though not too clear because that is why the Holmes test fails). Eighty-five years is enough.

APPENDIX A

Case	Name	Cite	Citation	Court	Remanded?	Removal	Remand	Time Lapsed
1	A.O. Sherman, LLC v. Bokina	2007 U.S. Dist. LEXIS 61848	3:07-cv-0240 (CFD)	D. Conn.	1	1/21/2007	8/22/2007	213
2	Adventure Outdoors, Inc. v. Mayor of New York	519 F. Supp. 2d 1258		N.D. Ga.	0	8/18/2006	9/20/2007	
3	Air Measurement Techns., Inc. v. Akin Gump Strauss Hauer & Feld, L.L.P.	504 F.3d 1262		Fed. Cir.	0	6/27/2003	10/15/2007	
4	Altman v. Flue-Cured Tobacco Coop. Stabilization Corp.	2007 WL 3286413	Civil Action No. 7:07-CV-83-HL	M.D. Ga.	1	7/3/2007	11/6/2007	126
5	Am. Sys. Consulting, Inc. v. Devier	514 F. Supp. 2d 1001		S.D. Ohio	1**	8/20/2007	8/28/2007	8
6	Barash v. Ford Motor Credit Corp.	2007 U.S. Dist. LEXIS 44641	No 06-CV-6497 (JFB) (ARL)	E.D.N.Y.	1**	12/6/2006	6/20/2007	196
7	Beechwood Dev. Group, Inc. v. Konersman	2007 U.S. Dist. LEXIS 79657	C.A. No.: 2:07-1394-PMD	D.S.C.	1	5/16/2007	7/23/2007	68
8	Blue Cross Blue Shield of Minn. v. GlaxoSmithKline, P.L.C.	2008 WL 304888	Civil No. 07-4072.	D. Minn.	1	9/24/2007	1/31/2008	129
9	Bockrath v. General Motors	2007 U.S. Dist. LEXIS 52612	Case No. 3:07CV00784	N.D. Ohio	0	3/16/2007	7/20/2007	
10	C&H Contracting of Miss., L.L.C. v. Lakeshore Eng'g Servs., Inc.	2007 U.S. Dist. LEXIS 62627	CAUSE NO. 1:07CV700	S.D. Miss.	1	6/19/2007	8/24/2007	66
11	Campbell v. Sutter Health	2007 U.S. Dist. LEXIS 66057	No. C-07-3406 MMC	N.D. Cal.	1	6/28/2007	8/28/2007	61
12	Chase Manhattan Mortgage Corp. v. Smith	507 F.3d 910		6th Cir.	1	12/5/2005	11/14/2007	709
13	Chen v. United Way of the Bay Area	2007 U.S. Dist. LEXIS 61367	No. C 07-02785 WHA	N.D. Cal.	1	5/29/2007	8/13/2007	76
14	Chirik v. TD BankNorth, N.A.	2008 WL 186213	Civil Action No. 06-04866.	E.D. Pa.	0	11/1/2006	1/15/2008	
15	City of Greenwood v. Martin Marietta Materials, Inc.	2007 U.S. Dist. LEXIS 46183	No. 07-157-CV-W-DW	W.D. Mo.	1	2/28/2007	6/26/2007	118
16	DeAngelo-Shuayto v. Organon USA Inc.	2007 WL 4365311	Civil Action No. 07-2923 (SRC).	D.N.J.	1	6/22/2007	12/12/2007	173
17	Ekas v. Burris	2007 WL 4055630	Nos. 07-61156-CIV, 07-61157-CIV.	S.D. Fla.	1	8/14/2007	11/14/2007	92
18	Fields v. Organon USA Inc.	2007 WL 4365312	Civil Action No. 07-2922 (SRC).	D.N.J.	1	6/22/2007	12/12/2007	173
19	Finisar Corp. v. U.S. Bank Trust Nat. Assoc.	2007 WL 4326765	No. C 07-4052 JF PVT.	N.D. Cal.	0	8/7/2007	12/7/2007	
20	Gloucester County Improvement Auth. v. Gallenthin Realty Dev., Inc.	2008 WL 336784	Civil Action No. 07-5328.	D.N.J.	1	11/6/2007	2/5/2008	91
21	Greene v. Novartis Pharms. Corp.	2007 WL 3407429	Civil Action No. 7:07-CV-00091-HL.	M.D. Ga.	1	8/12/2007	11/14/2007	94
22	Griffis v. Flue-Cured Tobacco Coop. Stabilization Corp.	2007 WL 3285749	Civil Action No. 7:07-CV-84-HL	M.D. Ga.	1	7/3/2007	11/6/2007	126
23	Hemberger v. Mansfield Twp. Bd. of Educ.	2007 U.S. Dist. LEXIS 81512	CIVIL ACTION NO. 07-3805 (MLC)	D.N.J.	1	6/29/2007	11/1/2007	125
24	Hirschbach v. NVE Bank	496 F. Supp. 2d 451		D.N.J.	1	12/22/2006	7/24/2007	214
25	Ho v. Colony Ins. Co.	2008 WL 145023	Civil Action No. 07-6990.	E.D. La.	1	10/17/2007	1/14/2008	89
26	Huntington Nat'l Bank v. JP Morgan Chase Bank	2007 U.S. Dist. LEXIS 52790	Case No. 2:07-cv-378	S.D. Ohio	1	4/27/2007	7/20/2007	84
27	Jarmuth v. Cox	2007 U.S. Dist. LEXIS 75732	Civil Action No. 1:07CV33	N.D. W. Va.	0***	3/12/2007	9/28/2007	
28	Jericho Sys. Corp. v. Booz Allen Hamilton Inc.	2007 U.S. Dist. LEXIS 81495	Civil Action No. 3:07-CV-2009-L	N.D. Tex.	1	6/20/2007	10/31/2007	133

29	Johnson v. Precision Airmotive, L.L.C.	2007 WL 4289656	No. 4:07CV1695 CDP.	E.D. Mo.	1**	10/2/2007	12/4/2007	63
30	Knaack v. Warren Indus., Inc.	2007 WL 4201121	No. 07-13787.	E.D. Mich.	1	9/10/2007	11/28/2007	79
31	Kostmayer Constr., L.L.C. v. M.R. Pittman Group, L.L.C.	2007 WL 4553991	Civil Action No. 07-7814.	E.D. La.	1	11/1/2007	12/19/2007	48
32	Kurz v. Fidelity Mgmt. & Research Co.	2007 WL 3231423	Case No. 07-cv-709-JPG	S.D. Ill.	1**	10/10/2007	10/30/2007	20
33	Lee v. Flue-Cured Tobacco Coop. Stabilization Corp.	2007 WL 3286397	Civil Action No.: 7:07-CV-82-HL.	M.D. Ga.	1	7/3/2007	11/6/2007	126
34	Maciolek v. Maciolek	2008 WL 150947	Civil Action No. 07-2510.	E.D. Pa.	1	6/20/2007	1/15/2008	209
35	Massachusetts v. Fremont Inv. & Loan	2007 WL 4571162	Civil Action No. 07-11965-GAO.	D. Mass.	1	10/16/2007	12/26/2007	71
36	McGovern v. CBT Direct, L.L.C.	2007 WL 3407392	No. 8:07-CV-1245-T-30TBM.	M.D. Fla.	1	7/16/2007	11/13/2007	120
37	Mikulski v. Centerior Energy Corp.	501 F.3d 555		6th Cir.	1	12/13/2002	8/21/2007	1712
38	Miss. Veterans Home Purchase Bd. v. State Farm Fire & Cas. Co.	492 F. Supp. 2d 579		S.D. Miss.	1	3/2/2007	6/21/2007	111
39	N. Cent. Pa. Dialysis Clinics, L.L.C. v. Toda	2007 U.S. Dist. LEXIS 59053	CASE NO. 4:07 CV 1968	N.D. Ohio	1	7/2/2007	8/13/2007	42
40	N.J. Dep't of Envtl. Prot. v. Minn. Mining & Mfg. Co.	2007 U.S. Dist. LEXIS 49613	Civil Action No. 06-2612(NLH)	D.N.J.	1	5/7/2007	7/5/2007	59
41	Neighborhood Mortgage, Inc. v. Fegans	2007 U.S. Dist. LEXIS 63241	CIVIL ACTION NO. 1:06-CV-1984-JOF	N.D. Ga.	1	8/23/2006	8/28/2007	370
42	New York v. Dell, Inc.	514 F. Supp. 2d 397		N.D.N.Y.	1	5/30/2007	10/16/2007	139
43	Nicholson v. Countrywide Home Loans	2007 WL 4460984	No. 1:07 CV 3288.	N.D. Ohio	0	10/24/2007	12/13/2007	
44	Pa. Employees Benefit Trust Fund v. Eli Lilly & Co.	2007 U.S. Dist. LEXIS 74579	CIVIL ACTION No. 07-2057	E.D. Pa.	1	5/21/2007	10/5/2007	137
45	Pennsylvania v. Ely Lilly & Co.	511 F. Supp. 2d 576	Civil Action No. 07-1083	E.D. Pa.	1	3/19/2007	6/27/2007	100
46	Red Lion Area Sch. Dist. v. Bradbury	2007 WL 2850584	Civil Action Nos. 07-cv-00563-JF, 07-cv-00720-JF.	E.D. Pa.	1	2/9/2007	10/1/2007	234
47	Rigby v. Flue-Cured Tobacco Coop. Stabilization Corp.	2007 WL 3285599	Civil Action No. 7:07-CV-85-HL.	M.D. Ga.	1	7/3/2007	11/6/2007	126
48	Robichaux v. Nationwide Mutual Ins. Co.	2007 U.S. Dist. LEXIS 70673	CIVIL ACTION NO.1:06CV1165 LTS-RHW	S.D. Miss.	1	11/22/2006	9/21/2007	303
49	Schmidt v. St. Clair	2007 WL 4463933	No. 5:07-cv-271-Oc-10GRJ.	M.D. Fla.	1	7/5/2007	12/17/2007	165
50	Scott v. Sysco Food Serv. of Metro New York, L.L.C.	2007 U.S. Dist. LEXIS 79519	Civ. A. No. 07-3656 (SRC)	D.N.J.	1	8/1/2007	10/26/2007	86
51	South Carolina v. Eli Lilly & Co.	2007 U.S. Dist. LEXIS 56847	C.A. No. 7:07-1875-HMH	D.S.C.	1	7/6/2007	8/3/2007	28
52	South Carolina v. Janssen Pharmaceutica, Inc.	2007 U.S. Dist. LEXIS 49904	C.A. No. 7:07-1452-HMH	D.S.C.	1	5/22/2007	7/10/2007	49
53	Swain v. Flue-Cured Tobacco Coop. Stabilization Corp.	2007 WL 3285707	Civil Action No. 7:07-CV-86-HL.	M.D. Ga.	1	7/3/2007	11/6/2007	126
54	Utah v. Eli Lilly & Co.	509 F. Supp. 2d 1016		D. Utah	1	6/11/2007	9/4/2007	85
55	Von Essen v. C.R. Bard, Inc.	2007 WL 2086483	No. Civ.A. 06-4786(SDW).	D.N.J.	1	10/4/2006	6/18/2007	257
56	Wexler v. United Air Lines, Inc.	496 F. Supp. 2d 150		D.D.C.	1	11/9/2006	7/31/2007	264
57	Williams v. Viva Health, Inc.	2008 WL 220799	No. 2:07-cv-321-WKW.	M.D. Ala.	1	4/12/2007	1/25/2008	288
58	Wolf v. Bankers Life & Cas. Co.	519 F. Supp. 2d 674		W.D. Mich.	1	6/29/2007	9/25/2007	88

59	Zenergy, Inc. v. Palace Exploration Co. 2007 U.S. Dist. LEXIS 57288	Case No. 07-CV-34-GKF-PJC	N.D. Okla.	1	1/10/2007	8/7/2007	209
				48	Average Delay		184.9583333
					Average Delay w/o appeal		129.14

Remanded? 1 = remanded (back to state court)
0 = retained (stayed in fed court)

Total	59
% Remanded	81.3559322
% Retained	18.6440678

** Not technicallily remanded. *See infra* n.135.

*** The court finds jurisdiction under Grable, but dismisses the claims pursuant to a 12(b)(6) motion. The remanded claim is the sole remaining state-law claim, of which there was no allegation of jurisdiction under the Grable analysis.

APPENDIX B

Case	Name	Cite	Citation	Court	Remanded?	Removal	Remand	Time Lapsed
1	Snook v. Deutsche Bank AG	410 F.Supp.2d 519		S.D. Tex.	1	8/4/2005	1/17/2006	166
2	Stechler v. Sidley Austin Brown & Wood, L.L.P.	2006 WL 90916	No. Civ.A. 05-3485(HAA).	D.N.J.	1	7/12/2005	1/13/2006	185
3	In re FedEx Ground Package Sys., Inc.	2006 WL 148945	No. 3:05MD527 RM, MDL-1700	N.D. Ind.	0	1/21/2005	1/13/2006	
4	Casale v. Metro. Transp. Auth.	2005 WL 3466405	No. 05 Civ. 4232(MBM).	S.D.N.Y.	1	4/28/2005	12/19/2005	235
5	RA Investment I, L.L.C. v. Smith & Frank Group Sers.	2005 WL 3299710	No. 4:05CV363.	E.D. Tex.	1	9/14/2005	12/5/2005	82
6	Texas *ex rel.* Ven-A-Care of Fla. Keys, Inc. v. Abbott Labs. Inc.	2005 WL 5430194	No. A-05-CA-897-LY.	W.D. Tex.	1	10/21/2005	12/5/2005	45
7	County of Santa Clara v. Astra USA, Inc.	401 F.Supp.2d 1022		N.D. Cal.	0	*	12/2/2005	
8	Ling v. Deutsche Bank AG	2005 WL 3158040	No. 4:05CV345.	E.D. Tex.	1	8/26/2005	11/28/2005	94
9	City Nat'l Bank of W. Va. V. Mountaineer Capital, L.P.	2005 WL 3098923	No. Civ.A. 2:05-CV-00842.	S.D. W. Va.	1	10/20/2005	11/18/2005	29
10	Buis v. Wells Fargo Bank, N.A.	401 F.Supp.2d 612		N.D. Tex.	1	3/21/2005	11/18/2005	242
11	Virtue v. Bethel Group Home, Inc.	2005 WL 3069139	No. F05-0006 CV (RRB).	D. Ala.	1	*	11/14/2005	
12	Armitage v. Deutsche Bank AG	2005 WL 3095909	No. C 05-3998 PJH.	N.D. Cal.	1	10/3/2005	11/14/2005	42
13	Acker v. AIG Intern., Inc.	398 F.Supp.2d 1239		S.D. Fla.	1	7/27/2005	11/8/2005	104
14	Burney v. 4373 Houston, LLC	2005 WL 2736515	No. Civ.A. 5:05-CV-255(C.	M.D. Ga.	1	7/29/2005	10/24/2005	87
15	State of Minnesota v. Pharmacia Corp.	2005 WL 2739297	No. 05-1394 (PAM/JSM).	D. Minn.	1	7/13/2005	10/24/2005	103
16	Leggette v. Washington Mut. Bank, FA	2005 WL 2679699	No. 3:03-CV-2909-D.	N.D. Tex.	1	12/2/2003	10/19/2005	687
17	Wandel v. American Airlines, Inc.	2005 WL 2406017	No. 4:05MD1702 JCH.	E.D. Mo.	1	3/3/2005	9/28/2005	209
18	Cantwell v. Deutsche Bank Securities, Inc.	2005 WL 2296049	No. Civ.A. 305CV1378-D.	N.D. Tex.	1	*	9/21/2005	
19	Maletis v. Perkins & Co., P.C.	2005 WL 3021254	No. CV-05-820-ST.	D. Or.	1	6/6/2005	9/13/2005	99
20	Pennsylvania v. Tap Pharmaceutical Products, Inc.	415 F.Supp.2d 516	No. Civ.A. 2:05-CV-03604.	E.D. Pa.	1	7/13/2005	9/9/2005	58
21	Harmon v. Maury County, TN	2005 WL 2133697	No. 1:05 CV 0026.	M.D. Tenn.	1	4/27/2005	8/31/2005	126
22	Sheridan v. New Vista, L.L.C.	406 F.Supp.2d 789		W.D. Mich.	1	6/20/2005	8/30/2005	71
23	Caggiano v. Pfizer, Inc.	384 F.Supp.2d 689		S.D.N.Y.	1	4/26/2005	8/26/2005	122
24	*In re* Circular Thermostat	2005 WL 2043022	No. MDL C05-01673WHA.	N.D. Cal.	1	*	8/24/2005	
25	Cafe Rio, Inc. v. Costa Azul Holdings, L.L.C.	2005 WL 2264135	No. 2:05-CV-0470J.	D. Utah	1	6/1/2005	8/17/2005	77
26	Thomas v. Friends Rehab. Program, Inc.	2005 WL 1625054	No. Civ.A.04-4288.	E.D. Pa.	1	9/10/2004	7/11/2005	304
27	Employers-Shopmens Local 516 Pension Trust v. Travelers Cas. and Sur. Co. of Am.	2005 WL 1653629	No. 05-444-KI.	D. Or.	1	3/29/2005	7/6/2005	99
28	*In re* Zyprexa Prods. Liab. Litig.	375 F.Supp.2d 170		E.D.N.Y.	0	*	7/1/2005	
29	Becnel v. KPMG L.L.P.	387 F.Supp.2d 984		W.D. Ark.	0	3/3/2005	6/21/2005	
30	Baker v. BDO Seidman, L.L.P.	390 F.Supp.2d 919		N.D. Cal.	1	6/20/2005	9/29/2005	101
31	Wisconsin v. Abbott Labs.	390 F.Supp.2d 815		W.D. Wis.	1	7/13/2005	9/29/2005	78
32	Kentucky v. China Tobacco Anyang Cigarette Factory	2005 WL 2234638	No. Civ.A.3:05 38 JMH.	E.D. Ky.	1	*	9/14/2005	
33	Brill v. Countrywide Home Loans, Inc.	427 F.3d 446		7th Cir.	1	5/6/2005	10/20/2005	167
34	Comes v. Microsoft Corp.	403 F.Supp.2d 897		S.D. Iowa	1	10/13/2005	11/22/2005	40
35	Kaye v. Southwest Airlines Co.	2005 WL 2074327	No. Civ.A.3:05CV0450-D.	N.D. Tex.	1	3/4/2005	8/29/2005	178

36	Kentucky v. China Tobacco Anyang Cigarette Factory	383 F.Supp.2d 917		E.D. Ky.	1	6/9/2005	8/4/2005	56
37	Kentucky v. Cibahia Tabacos Especiais Limitada	2005 WL 1868808	No. Civ.A. 3:05-39-JMH.	E.D. Ky.	1	*	8/4/2005	
38	Kentucky v. Claymore Group of Am. Corp.	2005 WL 1868810	No. Civ.A. 3:05-40-JMH.	E.D. Ky.	1	*	8/4/2005	
39	City of Livingston v. Dow Chem. Co.	2005 WL 2463916	No. C 05-03262 JSW.	N.D. Cal.	1	8/10/2005	10/5/2005	56
40	City of Oceanside v. Dow Chem. Co.	2005 WL 2463917	No. C 05-02482 JSW.	N.D. Cal.	1	6/17/2005	10/5/2005	110
41	Montara Water and Sanitary Dist.	2005 WL 2463918	No. C 05-02480 JSW.	N.D. Cal.	1	6/17/2005	10/5/2005	110
42	City of Alhambra v. Dow Chem. Co.	2005 WL 2463952	No. C 05-02595 JSW.	N.D. Cal.	1	6/24/2005	10/5/2005	103
43	Totherow v. Cent. Transp. Int'l, Inc.	2005 WL 4755219	No. 1:05-cv-126.	E.D. Tenn.	0	5/3/2005	10/3/2005	
44	Bar J Sand & Gravel, Inc. v. W. Mobile N.M., Inc.	2005 WL 3663689	No. Civ. 05-800JBWPL.	D.N.M.	1	7/25/2005	9/29/2005	66
45	Sarantino v. Am. Airlines, Inc.	2005 WL 2406024	No. 4:05MD1702 JCH.	E.D. Mo.	1	3/17/2005	9/29/2005	196
46	Broder v. Cablevision Sys. Corp.	418 F.3d 187		2nd Cir.	0	2/23/2004	8/11/2005	
					40		**Average Delay**	133.1470588

Remanded? 0 = jurisdiction sustained (stayed in federal court)
1 = remanded (back to state court)

Total	46
% Remand	86.95652174
% Retained	13.04347826

* Removal date not published. *See* nn.132 & 135-36 and accompanying text.

ABILITY TO PAY AND THE TAXATION OF VIRTUAL INCOME

ADAM S. CHODOROW*

Most people's intuition suggests that virtual income should be exempt from tax so long as it remains in-world. However, given the real-world economic value inherent in such income, it is difficult to square that intuition with current tax doctrine and policy. Those who have considered the question to date have attempted to classify virtual income as either falling outside the theoretical definition of income or into an existing tax category that is excluded from the real-world tax base.

In this Article, I argue that the existing analyses and the proposals it produces are not fully satisfactory. Instead, I offer a new approach that focuses on virtual income's impact on a taxpayer's ability to pay real-world taxes. Ability to pay is a core tax concept, and using it in this context yields results consistent with both intuition and existing tax policy and doctrine. In particular, I argue that the taxation of virtual income should be a function of a taxpayer's ability to cash out. Virtual income from worlds that permit participants to cash out should be taxed because receipt of such income increases a taxpayer's ability to pay real-world taxes. Virtual income from worlds that preclude participants from cashing out should be excluded from the tax base.

However, practical considerations, including concerns regarding tax evasion, administrative costs relative to expected revenues, and government intrusion into people's lives, may warrant overriding initial conclusions based on ability to pay. To account for these concerns, I propose that the IRS designate worlds as either open or closed, based on the ability to cash out and the considerations described above. Virtual income earned in closed worlds would be exempt from tax, so long as it is left in-world. Virtual income earned in open worlds would be subject to tax. However, in a nod to administrative concerns, I propose adopting a $600 *de minimis* threshold, under which no tax would be owed. Taxpayers whose virtual income in a given world exceeds the threshold would owe tax on all such income.

---

* Professor of Law, Sandra Day O'Connor College of Law at Arizona State University, Tempe, Arizona. I would like to thank Marjorie Kornhauser, Doug Sylvester, Mary Sigler, Carissa Hessick, F. Andrew Hessick, Ruth Mason, Kerry Ryan and the participants in the St. Louis University and Thomas Jefferson Faculty Workshops for comments on earlier drafts. While I do not always agree with the analysis and proposals put forth by Leandra Lederman and Bryan Camp, I nonetheless owe them a great debt, as they have set the stage for this article and provided insightful comments on an earlier draft of this article. Finally, I would like to thank Ted Seto, whose initial thoughts on this topic helped me formulate my own.

I. INTRODUCTION

Virtual worlds provide a forum for people to engage in a wide variety of activities, ranging from role playing in medieval fantasy worlds to setting up online branches of real-world businesses. As more people take part in these online communities, questions have arisen whether real-world laws do—or should—apply to activities occurring solely in-world.[1] These questions are not simply academic, as several lawsuits allege that real-world laws do apply.[2]

Given that virtual goods have inherent real-world economic value, it is not surprising that Congress and tax scholars have begun to consider whether real-world tax laws do or should apply to virtual activities.[3] A consensus exists for the

---

1. Commentators have focused on areas as diverse as property law, F. Gregory Lastowka & Dan Hunter, *The Laws of Virtual Worlds*, 92 CAL. L. REV. 1, 29-51 (2004); criminal law, Orin S. Kerr, *Criminal Law in Virtual Worlds* (GWU Pub. Law & Legal Theory Paper No. 391), *available at* http://ssrn.com/abstract=1097392; and intellectual property, Tom W. Bell, *Virtual Trade Dress: A Very Real Problem*, 56 MARYLAND L. REV. 384, 385 (1997). *See also* Allen Sipress, *Does Virtual Reality Need a Sheriff?*, WASH. POST, June 2, 2007, at A1 (discussing criminal investigations into virtual acts); Benjamin Duranski, *Rampant Trademark Infringement in Second Life Costs Millions, Undermines Future Enforcement*, VIRTUALLY BLIND, May 4, 2007, *available at* http://virtuallyblind.com/2007/05/04/trademark-infringement-virtual-worlds (discussing the misappropriation of intellectual property rights by Second Life participants).

2. *See, e.g.*, Bragg v. Linden Research, Inc., 487 F. Supp. 2d 593 (E.D. Pa. 2007) (concerning property rights in Second Life). This case has settled.

3. The Joint Economic Committee (JEC) has taken up this question but has yet to issue a report. However, Rep. James Saxton, the chairman and ranking Republican member of that committee, has stated that it would be a "mistake" for the IRS to tax income earned solely within the confines of virtual space. Robert Janelle, *Taxing Virtual Income*, SUITE101.COM, Oct. 24, 2006, http://videogames.suite101.com/article.cfm/taxing_virtual_income. In addition to the income tax consequences, the JEC is concerned about potential money laundering and terrorist threats. Virtual World News, *Congress Joint Economic Committee Talking to Virtual Worlds Operators*, Aug. 22, 2007, http://www.virtualworldsnews.com/2007/08/congress-joint-.html. Academic consideration of this issue includes Bryan T. Camp, *The Play's the Thing: A Theory of Taxing Virtual Worlds*, 59 HASTINGS L.J. 1 (2007) (arguing that activity within a virtual world does not and should not constitute taxable income); Edward Castronova, *The Right to Play*, 49 N.Y.L. SCH. L. REV. 185 (2004) (considering taxation as part of the broader question of whether real-world laws should apply in virtual worlds); Steven Chung, *Real Taxation of Virtual Commerce*, 29 VA. TAX REV. (forthcoming Winter 2009) (arguing that, to the extent one taxes virtual income, one should use the foreign currency rules found in Subchapter J of the Internal Revenue Code to determine gains and losses); Leandra Lederman, *Stranger than Fiction: Taxing Virtual Worlds*, 82 N.Y.U. L. REV. 1620 (2007) (arguing all transactions in "intentionally commodified worlds" should be taxed, while only "real-market sales or exchanges" should be taxed in game worlds); Dustin Stamper, *Taxing Ones and Zeros: Can the IRS Ignore Virtual Economies?*, 114 TAX NOTES 149 (Jan 15, 2007) (examining impending problems of taxing virtual income for the IRS); Timothy J. Miano, *Virtual World Taxation: Theories of Income Taxation Applied to the Second Life Virtual Economy* (2007) (unpublished article, http://works.bepress.com/timothy_miano/1/) (examining the federal tax implications of Second Life); *see also* BRYAN T. CAMP & THEODORE P. SETO, MAKING MONEY WHILE HAVING MORE FUN: TAXATION OF ONLINE GAMING, (American Bar Association–

proposition that anyone who "cashes out," or converts virtual wealth to real-world wealth, should be taxed on their gains.[4] The question of whether, and to what extent, the IRS can or should seek to tax activity that occurs entirely in-world is more difficult.

Those who have considered this question to date generally have concluded that most or all virtual income should not be taxed. One scholar takes, what I call, a transaction-based approach, where the tax result depends on the type of transaction that generates the income.[5] However, to avoid double taxation of consumption activities, she would exclude from the tax base virtual income generated as a result of a "consumption-oriented" activity.[6] Under the resulting proposal, all virtual income earned in "game worlds"[7] would be exempt from taxation, while most virtual income earned in "unscripted worlds"[8] would be exempt.[9] Another scholar would classify all virtual income as imputed and therefore outside the tax base, regardless of the type of virtual world or transaction involved.[10]

In this Article, I argue that none of the analyses to date—nor the proposals they produce—are fully satisfactory. Instead, I offer a new approach that focuses on virtual income's impact on a taxpayer's ability to pay real-world taxes. This approach better reflects core tax principles, is consistent with existing tax doctrine, implements tax policy objectives, and takes into account administrative concerns.[11]

I propose a two-step process for assessing whether, and to what extent, virtual income should be subject to real-world taxation. One must first determine whether the receipt of such income increases a taxpayer's ability to pay real-world taxes. In the context of virtual worlds, ability to pay depends on the

---

Section of Taxation 2008), *available at* http://www.abanet.org/tax/taxiq/mid08.html (follow "5470200801025" hyperlink). Professor Seto has since posted his comments in article form. Theodore P. Seto, *When is a Game Only a Game?: The Taxation of Virtual Worlds*, 77 U. CIN. L. REV. ___ (2009) (forthcoming).

4. *See, e.g.*, Camp, *supra* note 3, at 2, 45; Lederman, *supra* note 3, at 1623; Miano *supra* note 3, at 2.

5. Lederman, *supra* note 3, at 1625.

6. *Id.* at 1663.

7. An example of a "game world" is World of Warcraft.

8. An example of an "unscripted world" is Second Life.

9. Lederman, *supra* note 3, at 1625.

10. Camp, *supra* note 3, at 60.

11. I focus in this Article on how best to handle virtual income within the context of our existing income tax. However, other mechanisms exist to get at this income. For instance, one could attempt to create tax regimes that operate within the context of virtual worlds, such as a head tax, a sales tax, or a full-fledged income tax. Alternately, one could impose a tax on virtual world developers, the cost of which they could pass on in some form or another. Such efforts would not perfectly match the results one would achieve using the income tax system. However, to the extent that people really do think of their virtual income as pretend, such an approach might meet with less resistance than including virtual income in the real-world tax base. Consideration of such efforts is beyond the scope of this Article.

taxpayer's ability to cash out, i.e., convert his virtual wealth into real-world wealth—a determination that must be made on a world-by-world basis. Worlds that prohibit cashing out should be considered "closed," and virtual income acquired in such worlds would be excluded from the tax base.[12] If the taxpayer is able to cash out, the virtual world should be considered "open," and one must then proceed to the second step of the analysis.[13] In this step, one must determine whether the virtual income should be included in the tax base under existing doctrine or policy. This requires a transactional approach, and I argue that the virtual nature of the transactions at issue does not warrant different tax treatment from that which would apply to real-world transactions.

Taxing virtual income is not without its difficulties. For instance, markets may be thin, making valuation difficult. Administrative costs may well swamp the revenue obtained. Compliance issues may arise, as people generally resist the notion that we should tax people who are simply playing and making no effort to use their income outside the context of a particular virtual world. In a bow to these concerns, I would only impose real-world taxes on virtual income in years where a taxpayer earned virtual income (on a gross, not net, basis) in excess of a $600 threshold.[14]

In the past year, Chinese, Swedish, and South Korean tax authorities have all indicated an intent to tax virtual income left in-world.[15] Given the current budget crisis, the U.S. may not be far behind. Should the U.S. follow suit, this Article provides a theoretical and practical guide to the question of whether and how best to do so.

This Article proceeds as follows. Part II provides a brief description of virtual worlds and the potential income events with which the tax system must grapple. Part III considers the existing analyses and proposals. I begin with the transactional/consumption approach and demonstrate that it is inconsistent with existing doctrine and tax policy and raises a host of difficult administrative problems. I then turn to the imputed-income approach and illustrate how it stretches the notion of imputed income past its breaking point and that it, too, creates administrative problems.

---

12. I borrow the terms "closed" and "open" from Edward Castronova. Castronova, *supra* note 3, at 201–02. However, as will be evident from the discussion below in Part IV.A, I define those terms somewhat differently from Castronova. I focus on the impact virtual worlds have on the real world, while Castronova focuses on protecting virtual worlds from real-world influences. *Id.* at 209.

13. *Id.* at 201-02.

14. As described below in Part IV.C, this threshold is based on the most common reporting requirement threshold found in the Code, which arguably reflects a judgment regarding the administrative costs of reporting and the potential revenue losses associated with noncompliance absent reporting. Unlike with a floor, such as that found in I.R.C. § 67, not just the income in excess of the floor would be taxed under this proposal. I.R.C. § 67 (2000). Instead, all virtual income would be taxable in any year the threshold is reached.

15 Flora Graham, *Slapping a Tax on Playtime*, http://news.bbc.co.uk/2hi/technology/7746094.sun (last visited Dec. 13, 2008).

Part IV sets forth my policy and doctrinal analysis, as well as the specific proposal that flows from such analysis. In addition, I lay out a method for determining when worlds should be considered open and closed, a mechanism for making such determinations, and a justification for the $600 threshold I propose. Part V concludes.

II. BRIEF DESCRIPTION OF VIRTUAL WORLDS

In this Part, I give a brief description of virtual worlds, as a basic understanding of such worlds is necessary to understand the tax issues they engender. Each world is different, and, therefore, this description is necessarily general.[16]

Virtual worlds are online spaces that permit people to interact with one another through characters they create, often called avatars.[17] Virtual worlds differ from traditional video games in a number of important respects. First, these worlds do not pause or end when a user exits.[18] Rather, virtual life continues, and a returning player may well discover that things have changed significantly since she last visited the world.[19] Second, most worlds are designed to permit players to create their own, unique experience within the world, and much of what happens in a given world is as much a function of the participants' initiative as it is the developer's. Finally, most worlds have a virtual economy, where players can make, find, win, buy, sell, and exchange virtual goods.[20] To gain access to these worlds, users sign End User License Agreements (EULAs) or Terms of Service (TOS) agreements, which purport to establish participants' rights with respect to

16. For a description of the historical developments leading to the creation of virtual worlds, as well as the variety of virtual worlds, see Camp, *supra* note 3, at 3–8; Lastowka & Hunter, *supra* note 1, at 4–7; Lederman, *supra* note 3, at 1625–30.

17. For purposes of this article, I assume that avatars are puppets or agents of their owners and that their actions are to be attributed to the people who operate them. Thus, any activities in which avatars engage or income they generate are to be attributed to their owners/operators. However, in some cases the relationship between avatar and human may be more difficult than described above. Some avatars may be operated by several people who trade off over time. Such joint control potentially raises questions of partnership and could require complex attribution rules that look to formal ownership or who was "in control" when the income was earned. Intelligent and autonomous avatars are also on the horizon. Woodrow Barfield, *Intellectual Property Rights In Virtual Environments: Considering the Rights of Owners, Programmers, and Virtual Avatars*, 39 AKRON L. REV. 649, 653-54 (2006). I leave those considerations for another day.

18. *See* Lastowka & Hunter, *supra* note 1, at 5–6 (describing virtual worlds as "persistent and dynamic," in contrast to traditional video games, which are user-centric and dormant when the user is not present).

19. *Id.* at 6.

20. *See* Edward Castronova, *Virtual Worlds: A First-Hand Account of Market and Society on the Cyberian Frontier* (CESifo, Working Paper No. 618, 2001) (2001), *available at* http://ssrn.com/abstract=294828; Bryn Davies, *World of Warcraft Economics*, http://www.progsoc.org/~curious/economics.html (last visited Aug. 27, 2008).

the virtual goods created and obtained in those worlds.[21] The specific terms differ from world to world.

Virtual worlds are often divided into two categories: structured and unstructured.[22] Structured worlds are highly scripted environments where the game's developer provides a rich environment with scenery, preset roles, plot lines, and rules for interaction, among other things.[23] Players complete quests and search for treasure, often joining together with other players to do so.[24] As players acquire more virtual currency, property, and experience within a world, they are able to do more within the context of the game, often rising to a new "level" and undertaking more difficult quests.[25] Examples of structured worlds include World of Warcraft, EverQuest, Lineage, City of Heroes, Dark Age of Camelot, Entropia Universe, and Ultima Online.

In contrast, unstructured worlds create a space for people to interact without providing any set storyline or activities.[26] As a result, these worlds grow according to the tastes and inclinations of those who participate.[27] Examples of unstructured worlds include Second Life, The Sims Online, and There.[28]

The main point of most structured worlds is for players to enjoy the world the developers have constructed.[29] Accordingly, developers typically oversee player activity to ensure people do not ruin the environment for others or otherwise

---

21. See Andrew Jankowich, *EULAw: The Complex Web of Corporate Rule-Making in Virtual Worlds*, 8 TUL. J. TECH. & INTELL. PROP. 1, 7-11 (2006), for a discussion of these provisions and the weight that should be given to them.

22. *See* Camp, *supra* note 3, at 4. Camp and Lederman use different terminology to describe the different types of worlds. Camp identifies all virtual worlds as Massively Multiplayer Online Role Playing Games or MMORPGs and differentiates between "structured" and "unstructured" games. *Id.* at 4–8. Structured games correspond to what Lederman calls "game worlds," while unstructured games correspond to what she calls "unscripted worlds." *See* Lederman, *supra* note 3, at 1628–31. While many structured worlds today are geared towards entertainment, it need not be the case, and it is likely to change as these worlds evolve. Moreover, the term "game world" is somewhat loaded, in that it characterizes the activities that go on in such worlds and may affect one's thinking about the appropriate tax treatment of such activities. Indeed, as I argue below, the relevant quality for tax purposes is not whether the world is entertainment-oriented, but rather whether the world can be considered "open." Accordingly, I use Camp's terminology except when discussing Lederman's proposal.

23. *See, e.g.*, Camp, *supra* note 3, at 4–6 (describing the details of the structured game World of Warcraft).

24. *Id.* at 4–5.

25. *Id.* at 6.

26. *See id.* at 7.

27. *Id.* at 7–8.

28. Google is reportedly working on another unscripted virtual world in connection with Arizona State University. Google Operating System, *A Social Network for Google Earth*, (Sept. 23, 2007), http://googlesystem.blogspot.com/2007/09/social-network-for-google-earth.html; Chris Taylor, *Google Moves into Virtual Worlds*, BUSINESS 2.0, Dec. 14, 2006, http://money.cnn.com/2006/05/11/technology/business2_futureboy_0511/index.htm?postversion=2006051215.

29. Camp, *supra* note 3, at 4–6.

damage the game's integrity.[30] Developers generally reserve the right to kick out disruptive players, purporting to retain all property rights in the virtual goods created. To prevent players from using their real-world wealth to affect their in-world status or play, developers often restrict players' abilities to exchange virtual items outside the context of the world.[31] For instance, Blizzard Entertainment, which operates World of Warcraft, strictly prohibits the sale of virtual goods or currency for real money and reserves the right to take legal action against those who violate its rules.[32] As described more fully below, such worlds can be described as "closed."

---

30. This behavior is not strictly limited to structured worlds. Some developers of unstructured worlds also attempt to control players' behavior. *See, e.g.*, Welcome to There, What is There?: Info for Parents, http://www.there.com/parentInfo.html#5 (last visited Oct. 13, 2008) ("In addition to a customizable profanity filter which screens inappropriate language from text chat communications, There enforces strict "PG-13" content standards and takes extra measures to review all items to make sure they meet our content standards before they appear in There."). The blending of the lines between structured/game worlds and unstructured/unscripted worlds helps demonstrate why such categorization is not useful from a tax perspective.

31. *But see* Entropia Universe, http://www.entropiauniverse.com/en/rich/5676.html (last visited Aug. 27, 2008) (describing how to transfer PED (Entropia's virtual currency) to a debit card that can be used in the real world to purchase goods or to obtain cash); Michael Zenke, *SOE's Station Exchange: The Results of a Year of Trading*, GAMASUTRA, Feb. 7, 2007, http://www.gamasutra.com/features/20070207/zenke_01.shtml.

32. In 2004, Blizzard Entertainment released the following statement:

> It has come to our attention that certain individuals are selling Blizzard's in-game property for cash on auction sites such as eBay and on personal websites. The World of Warcraft Terms of Use clearly state that all of the content in World of Warcraft is the property of Blizzard, and Blizzard does not allow "in game" items to be sold for real money. Accordingly, Blizzard Entertainment will take any and all actions necessary to stop this behavior. Not only do we believe that it is illegal, but it also has the potential to damage the game economy and overall experience for the many thousands of others who play World of Warcraft for fun. In order to promote a fun and fair environment for all our customers, we are actively investigating those individuals who engage in this inappropriate activity and reserve the right to take legal action against these individuals to protect World of Warcraft for all those who "play by the rules." If you are found to be selling in-game property (such as coins, items, or characters), for real money, you will lose your characters and accounts, and Blizzard Entertainment reserves its right to pursue legal action against you as well.

Edward Castronova, *Blizzard Goes to War*, TERRA NOVA, Dec. 2004, http://terranova.blogs.com/terra_nova/2004/12/blizzard_goes_t.html; *see also* Greg Sandoval, *Sony to Ban Sale of Online Characters from its Popular Gaming Sites*, CNET NEWS, Apr. 10, 2000, http://www.news.com/2100-1017_3-239052.html (discussing Sony's ban of real-world auctions of virtual commodities from its game EverQuest); PlayNoEvil Game Security News & Analysis, *Blizzard Bans 76,000 Accounts and Removes 11 Million Gold From World of Warcraft*, Oct. 16, 2006, http://playnoevil.com/serendipity/index.php?/archives/883-Blizzard-bans-76,000-accounts-and-removes-11-Million-Gold-from-World-of-Warcraft.html (announcing Blizzard's ban of World of Warcraft members who were cheating by selling their virtual commodities for real-world currency).

Despite these types of restrictions, people can and do buy and sell items associated with these worlds on eBay, Itembay.com, Internet Gaming Entertainment (IGE.com), and similar sites.[33] In a practice known as "gold farming," participants play solely to acquire virtual goods and currencies that they then sell for real-world currency to other players who are too impatient to work through the game to garner the virtual wealth and experience necessary to play at an advanced level.[34]

The main point of most unstructured worlds is for the participants to create their own experiences. Almost all of the goods in unstructured worlds are made by those who participate from constituent building blocks made available in the given world.[35] Because unstructured worlds are designed to allow participants to act freely, developers take a far less active role in managing what goes on in-world.[36] In addition, they tend to grant users significantly greater rights in their creations than is typical in structured worlds and are generally more open to exchanges between virtual and real worlds.[37] For example, Second Life operates the LindeX, an official exchange that facilitates people buying and selling virtual currency.[38]

Although many people spend time in unstructured worlds as an escape from reality, at least some of these worlds are beginning to look more and more like an alternate forum in which to conduct real-world activities. For instance, a host of

---

33. *See, e.g.*, Internet Gaming Entertainment, http://www.ige.com (last visited Aug. 27, 2008) (specializing in the sale of virtual items via the internet). Writer Julian Dibbell reported that he was able to amass and then convert $11,000 of virtual goods in U.S. currency by "playing" Ultima Online. JULIAN DIBBELL, PLAY MONEY: OR HOW I QUIT MY DAY JOB AND MADE MILLIONS TRADING VIRTUAL LOOT 284 (2006); Stamper, *supra* note 3, at 150. As of this writing, eBay has banned most or all sales of virtual goods from World of Warcraft. Softpedia.com, *eBay Banned World of Warcraft Virtual Goods Auctions* (hereinafter *eBay Banned World of Warcraft*), Jan. 30, 2007, http://news.softpedia.com/news/eBay-Banned-World-of-Warcraft-Virtual-Goods-Auctions-45809.shtml (suggesting that the ban was implemented to stop people from using Trojan software to steal participants' passwords); HowStuffWorks.com, *Why is eBay Banning the Sale of Online-game Virtual Assets?*, Feb. 3, 2007, http://money.howstuffworks.com/ebay-ban.htm (suggesting that eBay was attempting to avoid possible lawsuits by developers).

34. *See* Julian Dibbell, *The Life of the Chinese Gold Farmer*, N.Y. TIMES MAGAZINE, Jun. 17, 2007, at 38.

35. Cory Ondrejka, *Escaping the Gilded Cage: User-Created Content and Building the Metaverse*, 49 N.Y.L. SCH. L. REV. 81, 87-88 (2004).

36. *See, e.g.*, Second Life Community Standards, http://www.secondlife.com/corporate/cs.php (last visited Oct. 13, 2008) (listing six rules for community conduct and stating that "Second Life is an adult community, but Mature material is not necessarily appropriate in all areas" and that "[e]very Resident has a right to live their [sic] Second Life"). *But see supra* note 29.

37. It might be more accurate to say that they demand fewer waivers, as common law and statutes create the rights, and the EULAs or TOSs readjust them.

38. In addition to the LindeX, other unofficial exchanges also exist. Linden Labs monitors the exchange rate and maintains a fairly constant rate of around L$ 270 to $1 U.S. dollar. See Camp, *supra* note 3, at 13, and Miano, *supra* note 3, at 10, for a discussion of the LindeX and other exchanges.

businesses, including Nissan, IBM, and Nike, have established a presence in Second Life to market their real-world goods.[39] Numerous universities have set up virtual sites to promote themselves, to allow students to interact, and even to hold classes and other meetings.[40] Even politicians have begun to establish a virtual presence.[41]

Virtual world participants acquire their virtual wealth in a number of different ways, many of which are analogous to real-world wealth acquisition. In both structured and unstructured worlds, people create their own goods, buy, sell, and exchange virtual goods, earn salaries for working, and receive gifts.[42] Game worlds further allow participants to acquire goods as "drops" or quest rewards..[43]

Over the past several years, people have begun to amass significant virtual wealth, which has significant real-world value. This has occurred primarily in open worlds, both structured and unstructured, though this need not be the case. At least two Second Life participants have been reported to have virtual wealth valued at over $1 million in U.S. currency.[44] Several Entropia Universe participants have also been reported to have earned significant sums.[45] The *Wall Street Journal* recently reported that a growing number of college students are

---

39. *See* Allison Enright, *How the Second Half Lives*, MARKETING NEWS, Feb. 15 2007, *available at* http://us.i1.yimg.com/us.yimg.com/i/adv2/pdf/brainfood/second_life.pdf; Games Brands Play, http://gamesbrandsplay.com (last visited Aug. 27, 2008) (describing Nissan's marketing of the Sentra in Second Life); Rachel Konrad, USA TODAY, *IBM To Open Islands in Virtual World,* Dec. 13, 2006, *available at* http://usatoday.com/tech/news/2006-12-13-ibm-second-life_x.htm.

40. *See, e.g.,* Andrew Johnson, *Business and Education Test Online Virtual World's Real-World Potential,* WIS. RAPIDS TRIB., Nov. 16, 2007, at A11 (predicting that the University of Arizona would have an operating Second Life site by the end of 2007).

41. See CBSNews.com, *Running For President In A Virtual World*, Apr. 3, 2007, http://www.cbsnews.com/stories/2007/04/02/politics/main2639476.shtml, and the accompanying photo essay of the virtual headquarters of the various candidates.

42. *See, e.g.,* World of Warcraft, Beginners, http://www.worldofwarcraft.com/info/beginners/index.html (last visited Oct. 13, 2008); Second Life, What is Second Life? http://secondlife.com/whatis (last visited Oct. 13, 2008).

43. Drops derive their name because defeated foes "drop" items that players can then pick up.

44. In 2006, *BusinessWeek* reported on Second Life's first millionaire (referring to the value in U.S. currency of her virtual wealth). *See* Rob Hof, *Second Life's First Millionaire*, BUSINESSWEEK, Nov. 26, 2006, http://www.businessweek.com/the_thread/techbeat/archives/2006/11/second_lifes_fi.html; *see also* David Naylor, *Technology, Media & Telecoms: A Virtual Reality*, LEGALWEEK, July 6, 2007, http://www.legalweek.com/Articles/1030336/A+virtual+reality.html (reporting that another Second Life participant claimed to have made over $1 million from an initial $10 investment).

45. Neha Tiwari, *Teen Pays Siblings' College Fees by Selling Virtual Weapons*, CNET'S NEWS, Oct. 10, 2006, http://news.com.com/8301-10784_3-6124572-7.html (describing a teenager who made $35,000 playing in Entropia Universe); Marketwire.com, *Champion Gamer NEVERDIE Rakes in $100,000 USD in Virtual Reality*, Aug. 21, 2006, http://www.marketwire.com/2.0/release.do?id=696685&sourceType=1 (describing someone who earned $100,000 in Entropia Universe).

foregoing traditional summer jobs to work online in virtual words; one student reportedly earned approximately $35,000 over the past four years.[46]

The extent and real-world value of some of these virtual economies is surprisingly large. For instance, a 2004 study concluded that the Gross National Product for EverQuest was $135 million, or $2,226 per capita.[47] It has also been estimated that annual real-world sales of virtual goods exceeded $880 million in 2004.[48] As of mid-2007, the number of participants in virtual worlds was estimated to exceed 35 million,[49] with some predicting that the figure will rise to 50 million by the year 2011.[50] Given the increasing popularity and growth of virtual worlds, the amount and real-world value of virtual income is certain to grow.[51]

The IRS has not sought to tax the acquisition of virtual income from purely in-world activities, though anyone who cashes out is clearly subject to tax on his or her gains.[52] The question at issue here is whether and to what extent existing tax doctrine and policy support this practice. Given the large (and growing) amounts at issue, this inquiry has morphed from a purely academic question to one with potentially significant real-world consequences for those who participate in virtual worlds. Indeed, depending on the conclusion one reaches, it could change the very nature of virtual reality.

III. EXISTING PROPOSALS FOR TAXING VIRTUAL INCOME

Those who have considered whether virtual income should be subject to real-world taxation generally oppose the idea. While some mention taxation in passing,[53] two published articles, one by Leandra Lederman and another by Bryan Camp, focus exclusively on this question.[54] Lederman combines a

---

46. Alexandra Alter, *My Virtual Summer Job*, WALL ST. J., May 16, 2008, at W1.

47. Castronova, *supra* note 20, at 33.

48. *See* Dibbell, *supra* note 33.

49. *See* Blake Snow, *GigaOM Top 10 Most Popular MMOs*, GIGAOM, June 13, 2007, http://gigaom.com/2007/06/13/top-ten-most-popular-mmos/.

50. Wagner James Au, *Virtual World Population 50 Million by 2011*, GIGAGAMEZ, May 24, 2007, http://gigagamez.com/2007/05/24/virtual-world-population-50-million-by-2011/.

51. Some recent actions may act to limit this growth, but they will not likely stop it. *See, e.g.*, *eBay's Ban on the Sale of Some Virtual Goods*, supra note 33; Posting of Robin Linden to Second Life Grid, http://blog.secondlife.com/2007/07/25/wagering-in-second-life-new-policy (July 25, 2007, 16:47 PDT).

52. I.R.C. § 61 (2000).

53. *See, e.g.*, Castronova, *supra* note 3, at 204. Castronova's objections relate to the nature of virtual worlds and the impact taxation would have on such worlds. However, if a virtual world is merely an extension of the real world, he sees no reason why tax laws should not apply.

54. Camp, *supra* note 3, at 60; Lederman, *supra* note 3, at 1625. Other articles in various prepublication stages include Miano, Seto, and Chung, *supra* note 3, and Terando, William D. Mennecke, Brian, Janvrin, Diane Joyce and Dilla, William N., *Taxation Policy in Virtual Worlds: Issues Raised by Second Life and Other Unstructured Games* (Oct. 1, 2008) J. LEGAL TAX RESEARCH (forthcoming, *available at* http:/ssrn.com/abstract=1287232). Miano, Chung, and

transaction-based, doctrinal analysis with a consumption-oriented policy analysis.[55] While these analyses often lead to the same conclusions, where they differ, Lederman contends that policy considerations should inform doctrine.[56] She concludes that in most cases virtual income should be excluded from tax until cashed out.[57] In contrast, Camp asserts that all virtual income is a form of imputed income and excludible from the tax base on that ground.[58] In this Part, I briefly summarize and then assess their reasoning and proposals. I find that neither approach comports well with existing doctrine or policy and that both raise serious administrative difficulties.

A. The Transactional/Consumption Approach

Lederman proposes a bifurcated approach, based on the type of virtual world and transaction involved.[59] For game worlds, she would adopt a pure cash out rule.[60] For unscripted worlds, such as Second Life, she would allow in-world barter transactions to go untaxed.[61] Thus, as in game worlds, barter in these worlds would be taxed on a cash out basis.[62] However, she would tax sales of virtual goods or services for virtual currencies.[63] Some of these conclusions are supported by her doctrinal analysis.[64] However, in some cases, she contends that policy considerations should override what might otherwise be the correct doctrinal result.[65] Her proposal is based on the conviction that "pure consumption" should not be taxed (other than by disallowing a deduction for expenses to participate), while profit-seeking activity should be taxed, and on the contention that taxing virtual income would lead to a double tax on consumption.[66]

This approach to the question of taxing in-world activities raises two questions. The first involves the propriety of Lederman's doctrinal analysis. The second involves the propriety of her claim that taxing virtual income would amount to a double tax on consumption in most cases. I address each in turn.

Terando all focus on commodified worlds, and on Second Life in particular. They conclude that taxation is warranted. Seto takes a broader look at virtual activities and concludes that taxation should depend on whether income is convertible or redeemable. I focus here on the two published articles in part because of timing (the other articles were made public after this article was nearing completion) and in part because the two published articles argue for non-taxation in most cases.

55. Lederman, *supra* note 3, at 1650, 1658.
56. *Id.* at 1625.
57. *Id.*
58. Camp, *supra* note 3, at 60.
59. Lederman, *supra* note 3, at 1625.
60. *Id.*
61. *Id.*
62. *Id.*
63. *Id.* at 1658.
64. *Id.* at 1641–57.
65. *Id.* at 1658.
66. *Id.* at 1659.

1. Doctrinal Analysis

Lederman's doctrinal approach is transaction-based in that the appropriate tax treatment of virtual income depends on the type of transaction that generates it.[67] She distinguishes between drops, sales and barter exchanges, and attempts to place each of these types of transactions within existing tax categories.[68]

a. Drops

Drops are usually associated with game worlds.[69] They are generally provided by developers as a reward for vanquishing some foe and are referred to as drops because defeated foes may drop items, which players are able to pick up and use. Lederman identifies four possible real-world income categories under which drops might fall, including (1) imputed income, (2) found objects (also called windfalls), (3) taken objects (i.e., objects obtained as the result of significant effort on the taxpayer's part), and (4) prizes and awards.[70] Imputed income is generally excluded from the tax base.[71] Found items are currently included in income.[72] Taken items (such as harvests) are generally excluded from income.[73] Prizes and awards are expressly included in the tax base under I.R.C. § 74.[74] Lederman concludes that, as drops are typically obtained after significant effort on the part of a participant, they should be considered a species of taken object and excluded from income on that ground.[75] I believe this conclusion to be

---

67. *See id.* at 1625.

68. *Id.* at 1643–57.

69. Lederman focuses solely on drops. *Id.* at 1643–50. Quest rewards—rewards provided by developers for accomplishing a task—are similar to drops. Thus, the tax consequences of obtaining such awards should be the same. In fact, this analysis extends beyond traditional structured worlds and drops/quest rewards. It applies to any situation where participants win items or points for the completion of some task, such as winning an online game. See notes 95–99 and accompanying text for a discussion of Worldwinner, an online gaming site that awards participants points for winning online contests. For the sake of clarity, I refer in this Part only to drops.

70. Lederman, *supra* note 3, at 1643–50.

71. *See* Joseph M. Dodge, *Accessions to Wealth, Realization of Gross Income, and Dominion and Control: Applying the "Claim of Right Doctrine" to Found Objects, Including Record-Setting Baseballs*, 4 FLA. TAX. REV. 685, 705 (2000).

72. Treas. Reg. § 1.61-14 (1993). *But see* Lawrence A. Zelenak & Martin J. McMahon, *Taxing Baseballs and Other Found Property*, 84 TAX NOTES 1299-1301 (1999) (arguing that the taxation of true windfalls in inappropriate and urging the Treasury Department to rescind this regulation).

73. *See, e.g.*, Treas. Reg. § 1.61-4 (1997) (regarding self-grown crops, fish caught by commercial fishermen, hunters' trophies, and minerals); Zelenak & McMahon, *supra* note 72, at 1302–04 (discussing commercial fisherman, big game hunters, prospectors and miners, and treasure hunters).

74. I.R.C. § 74 (2000).

75. Lederman, *supra* note 3, at 1646.

in error. Instead, as demonstrated below, drops should properly be considered prizes and subject to tax as such.

First, I concur with Lederman that drops cannot be considered imputed income.[76] Economists have long considered the question of imputed income and whether it should be subject to taxation.[77] Joseph Dodge has identified two definitions typically used for imputed income:

> (1) the flow of satisfactions obtained by a taxpayer (which would include not only the value of satisfactions derived from owning and spending but also the value of leisure, sleep, a happy marriage, etc.), and (2) the market-price equivalents of non-market *economic* activity (such as the value of self-grown crops and the rental value of self-owned assets, and possibly the value of self-performed services).[78]

Dodge criticizes the first definition as being too broad to be useful, finding the second better because it focuses on the measurable benefit that would accrue from market activity.[79]

Regardless of the definition one chooses, drops fit poorly. Imputed income is the benefit received from owning property or performing services for oneself. The value of such income is established by determining the value one could obtain by entering the market and renting the good or performing the service for someone else.[80] Receiving a virtual sword as a drop after battle with a dragon simply does not fit this definition. The value of the sword received is different from the imputed value of owning it, i.e., the value one might obtain by renting the sword out.[81] Moreover, the value of the sword is different from the value of the self-performed services undertaken to obtain the sword.[82] In fact, the income

---

76. *Id.* at 1644–46. Thus, both Lederman and I disagree with Camp, who contends that drops, like all other forms of virtual income, should be considered imputed income. *See* Camp, *supra* note 3, at 65.

77. *See* Dodge, supra note 71, *passim.*

78. *Id.* at 691–92 (internal citations omitted).

79. *Id.* at 692.

80. *Id.*

81. The difference between the value of a good and the imputed value of owning a good can be seen in the following simple example. Assume someone owns a home. If the value of the home and imputed value of owning the home were the same, then one would tax homeowners on the full value of their homes year after year, assuming imputed income were included in the tax base. In fact, if imputed income were included in the tax base, homeowners would be taxed only on the rental value of their homes, i.e., the money they could have earned from renting their homes that year, as that reflects the value they received during the tax year.

82. Arguably, the value of the good obtained could reflect the value of the self-performed services consumed in obtaining the good. However, if the values are equivalent, it is only fortuitous. For instance, assume a fisherman goes out and comes back empty handed. He has performed services for himself, but reaped no reward. This does not mean that the value of those services is $0. Indeed, *ex ante*, he willingly may have paid someone to fish for him. Thus, the services have value apart from the good ultimately obtained. No one has suggested taxing the value of self-

associated with obtaining the sword need not be imputed at all, as the player has actually received it for accomplishing a task.

Second, I reject Lederman's argument that drops should be considered a form of "taken" item that is traditionally excluded from income.[83] Shortly after Mark McGuire hit his record-breaking home run, a controversy erupted regarding the proper tax treatment of a fan lucky enough to catch such a valuable ball.[84] Larry Zelenak and Marty McMahon argued that the Treasury regulation including windfalls in income was wrong-headed, and they urged the Treasury Department to rescind it.[85] In particular, they noted that other types of found objects, such as fish, hunting trophies, and minerals were generally excluded from the tax base until sold or exchanged, and they argued that record setting baseballs should be treated no differently.[86] Doctrinally, they argued that all types of found items were properly excluded from income as a form of imputed income.[87]

In response, Joseph Dodge rejected the notion that such items could properly be classified as imputed income.[88] Instead, he distinguished between two types of found objects -- those that represented a true windfall (or were not the result of taxpayer efforts to obtain the item) and those that the taxpayer sought out and expended energy to obtain.[89] He termed these "found" and "taken" objects, respectively.[90] He argued that "taken" objects are properly excluded from income because they represent an investment, and the value in excess of such investment represents appreciation, which is typically not subject to tax absent a realization event.[91] Accordingly, he argued that "taken" objects, but not "found" objects, should be excluded from income until sold.[92] Lederman argues that Dodge's theory of "taken" objects applies to drops because players exert significant and directed efforts to obtain them, either by slaying dragons or completing whatever task the developers have set for them.[93]

While Dodge's theory might explain why we tax found, but not taken, objects, it does not necessarily follow that drops are "taken" objects. Indeed, drops differ from "taken" objects in one significant way. Each example of a "taken" object derives from nature. No third party is involved. In contrast, drops are provided by a world's developers as a reward for accomplishing a specific task. Although developers exercise god-like powers within the context of their

---

performed services, which properly constitute imputed income.

83. Lederman, *supra* note 3, at 1644.

84. *See* Zelenak & McMahon, *supra* note 72, at 1299.

85. *Id.* at 1301–02.

86. *Id.* at 1302–04.

87. *Id.*

88. *See* Dodge, *supra* note 71, at 694.

89. *Id.* at 694–704.

90. *Id.* at 704.

91. *Id.*

92. *Id.*

93. Lederman, *supra* note 3, at 1646–48. Lederman also concludes that players would likely have no basis in their drops. *Id.* at 1648–50.

worlds and are often referred to as "game gods,"[94] receiving a drop from such a god is fundamentally different from taking something from nature, i.e., receiving "God's bounty." In this regard, drops are like compensation or game show prizes, where participants earn the prizes for completing tasks set by others.

The prize-like nature of drops may become clearer if one considers the issue in the context of a different type of online activity. The website Worldwinner hosts competitions in a number of different real-world games, such as Scrabble, Boggle and Monopoly.[95] Players purchase points, which they use to enter specific competitions.[96] The winners of these competitions have their accounts credited with a number of points, which they can then either cash out or use to keep playing.[97] It seems self-evident that the award of points in this context is a prize or award for winning a Scrabble match. Certainly, if a real-world game show gave people money or other property for winning a real-world Scrabble match, the receipt would be considered a prize.[98] Obtaining a sword dropped by a slain dragon or receiving a quest reward is no different.[99] Accordingly, drops fall squarely into the prize and award category, which the Internal Revenue Code explicitly taxes.[100]

One could certainly argue that prizes and awards should be excluded from income because they represent the fruits of a recipient's effort and should be considered a form of "taken" object. Participants may practice for years, first to

---

94. *See* Media Coverage, *Where Have All the Game Gods Gone?*, GAMEDAILY, June 8, 2006, http://www.gamedaily.com/articles/features/where-have-all-the-game-gods-gone/69038/?biz=1.

95. See Camp & Seto, *supra* note 3, for a discussion of Worldwinner and whether these prizes are taxable before they are cashed out.

96. *Id.*

97. *Id.*

98. I make no predictions regarding the network ratings such a game show might garner.

99. Camp takes what he calls an "external" view of virtual worlds. Camp, *supra* note 3, at 44. In this view, one is not really winning a sword, but rather participating in a continually evolving play. Under this view, what occurs when one "wins a sword" is an alteration of the player's account that increases its value. If the property at issue is the account, then the change to the account is akin to a tree producing a fruit. Were one to adopt such a view, then taxation would not be appropriate until the fruit was separated from the tree, i.e., until the player cashed out. The difficulty with this view is that the account has not changed or increased in value *sua sponte*. Rather, someone—in this case the world's developer—has deposited something into the account. If my employer or a game show deposits my paycheck or a prize into my bank account, I cannot avoid taxation by arguing that my account has simply grown in wealth, and therefore taxation must await a realization event.

100 *See* Camp, *supra* note 3, at 47–48 & n.179 (discussing I.R.C. § 74). One complication is that people pay to participate in virtual worlds, and an argument could be made that drops are something they have purchased and therefore are not prizes. Setting aside the question whether one purchases the opportunity to play and the possibility of obtaining virtual income, as opposed to the income itself, this might be an issue where the value of the drops obtained was less than the fees paid to participate. However, participants are not guaranteed to receive drops, and there is no way to know whether the value of the drops a participant receives over a year will exceed the value paid to obtain them, or if so when. See discussion *infra* Part III.A.2.

get onto a television game show and then to win it. However, the same can be said for compensation. Indeed, prizes are more analogous to compensation than fishing in that one performs a task someone else sets forth and then receives from that person a reward.[101] Regardless, current doctrine clearly includes prizes and awards, and if one is to classify drops into an existing tax category, they are most properly defined as prizes and are includible in the tax base absent some overriding policy or administrative concern.

b. Sales and Exchanges

Lederman next considers sales and exchanges of virtual goods in game worlds.[102] She takes a rights-based approach and concludes that the answer to the question of whether sales and exchanges can be considered taxable events under current law depends on the rights people have in their virtual currency and goods.[103] If virtual items are considered property, she concedes that sales and barter exchanges are taxable events because they likely will be considered realization events under the standard established in *Cottage Savings Ass'n v. Commissioner*.[104] In contrast, if virtual goods are not considered property, but rather are treated under a license theory, she contends that the argument against taxation is stronger, noting that "[t]ransactions amounting to mere reallocations of possession of items in which all participants have use rights do not constitute realization events, and thus are not taxable."[105]

This approach to sales and exchanges presents two difficulties. The first is administrative, in that it requires a case-by-case analysis that could lead to different results for similar transactions in different worlds.[106] The second is substantive, in that the case for non-taxation, if people are deemed to have use rights, is weaker than Lederman suggests.[107]

The rights-based approach raises three administrative difficulties. First, determining whether to tax in-world transactions requires a difficult and potentially costly world-by-world analysis of the rights involved. Each world has its own rights regime, consisting of both state law and a EULA or TOS, which

---

101. That the drops are provided by virtual world developers also helps explain why drops cannot be considered imputed income. The presence of this third party reveals that the income is not imputed but the result of a market transaction that involves two parties. The recipient provides services or performs tasks in return for the world's developer providing the drop. That no money changes hands does not make the value received imputed.

102. Lederman, *supra* note 3, at 1650–55.

103. *Id.* at 1652–55.

104. *Id.* at 1653. *See also* Cottage Sav. Ass'n v. Comm'r, 499 U.S. 554, 556 (1991) ("[a]n exchange of property gives rise to a realization event so long as the exchanged properties . . . embody legally distinct entitlements").

105. Lederman, *supra* note 3, at 1658. Even if licenses are considered property, Lederman contends that their exchange might not constitute a realization event, even under the lax *Cottage Savings* standard. *Id.* at 1657–58.

106. *See supra* text accompanying note 21.

107. *See* Lederman, *supra* note 3, at 1653–58.

may purport to waive or otherwise alter the underlying allocation of rights. Thus, this approach requires taxing authorities to determine on a state-by-state basis the initial allocation of rights, the purported effect of the EULA or TOS, and the state law concerning the efficacy of any EULAs or TOS before determining the appropriate doctrinal answer. Even if the EULAs and TOSs were the last word on participants' rights, disputes are likely to erupt, and it seems inevitable that determining participant rights will require costly and time-consuming litigation, as was the case in *Bragg v. Linden Research, Inc.*[108]

Second, because these rights rest in part on contract, they are subject to change.[109] Thus, if a court issues a ruling adverse to the interests of a game's developer or its users, it seems likely that the developer will revise the EULA or TOS, setting off another round of litigation to determine what rights the new documents create and the proper tax ramifications of in-world transactions. The result could be a costly and time-consuming game of cat and mouse between the IRS and virtual world participants.

Third, basing taxation on people's underlying rights in virtual goods gives world developers significant discretion to determine tax consequences by changing rights to shield transactions from tax. Tax law routinely gives taxpayers discretion to structure their transactions to avoid adverse tax consequences.[110] However, at some point, taxpayers elevate form so far above substance that the IRS may challenge the transaction using the economic substance and sham transaction doctrines to look past the formal structure or apparent rights to the underlying economic reality of the transaction.[111] Under these principles, it is not clear whether labeling rights in virtual goods as property or mere use rights should or will govern the tax treatment of in-world transactions.

---

108. Bragg v. Linden Research, Inc., 487 F. Supp. 2d 593, 595 (E.D. Pa. 2007) (considering the parties' respective rights to virtual property in Second Life). This case settled, depriving the court of an opportunity to define the legal status of virtual property. *See* Posting of Greg L to Terra Nova, http://terranova.blogs.com/terra_nova/2007/10/virtual-law-upd.html #more (Oct. 16, 2007).

109. *See supra* text accompanying note 22.

110. The most obvious example can be seen in corporate reorganization provisions, which allow tax-free reorganizations, provided one follows the strict requirements of the statute. *See* I.R.C. § 368 (2000). Other examples include leasing in lieu of buying to take advantage of the deduction rules. *See, e.g., id.* §§ 162(a)(3), 263. As Judge Learned Hand noted, "[a]ny one may so arrange his affairs that his taxes shall be as low as possible; he is not bound to choose that pattern which best pays the Treasury; there is not even a patriotic duty to increase one's taxes." Helvering v. Gregory, 69 F.2d 809, 810 (2d Cir. 1934), *aff'd*, 293 U.S. 465 (1935).

111. Situations where the IRS has challenged the form of a transaction include sales disguised as leases and dividends disguised as salary. *See* Exacto Spring Corp. v. Comm'r, 196 F.3d 833 (7th Cir. 1999), *rev'g* Heitz v. Comm'r, T.C. Memo.1998-220; Estate of Starr v. Comm'r, 274 F.2d 294 (9th Cir. 1959). *See generally* Joseph Bankman, *The Economic Substance Doctrine*, 74 S. CAL. L. REV. 5 (2000) (describing the resurgence in common law responses to the problem of tax shelters); Allen D. Madison, *The Tension Between Textualism and Substance-Over-Form Doctrines in Tax Law*, 43 SANTA CLARA L. REV. 699 (2003) (reviewing the doctrines courts have used).

Finally, the possibility of tax rules applying differently to identical transactions raises administrative and compliance problems for both taxpayers and the government. Federal tax laws are supposed to be uniform, and a rights-based approach almost guarantees that results will differ from state to state.[112] Of course, the possibility of different tax results for identical transactions is inherent in the federal tax law, as federal income tax consequences routinely flow from state property law.[113] Where state laws vary, federal income tax consequences can lead to different tax results for seemingly identical behavior.[114] Nonetheless, in the interests of uniformity, Congress, the IRS, and the courts have routinely taken steps to ensure that uniform tax laws apply, regardless of underlying state law and taxpayers' choices in structuring their transactions.[115] While not every situation calls for uniform federal tax treatment, the taxation of in-world transactions seems to raise the issues and complexities that would warrant such treatment.

Lederman avoids these administrative problems by arguing that a policy against double taxing pure consumption warrants overriding the doctrinally correct answer.[116] However, as described below, it is not at all clear that those who generate virtual income while participating in virtual worlds are engaged in "pure consumption" or that it would amount to a double tax were one to tax such income. If one rejects either of these claims, one is left with the arduous task of parsing the myriad virtual worlds and their rights regimes, as well as any legal precedent that might bear on the question of how the tax law currently treats such transactions.

---

112. *See infra* note 115.

113. *Id.*

114. For instance, whether payments set forth in a divorce decree are considered alimony depends upon whether state law determines that payments end with the death of the recipient. *See* PUB. L. NO. 99-514, § 1843(b), 100 Stat. 2085, 2853 (1986) (retroactively removing a requirement that divorce decrees explicitly provide that payments not continue post-death, thereby permitting state law to control where such decrees were silent).

115. For example, in *Poe v. Seaborn*, the Supreme Court held that spouses in states with community property could split their income because both spouses had equal rights to the income, regardless of who earned it. Poe v. Seaborn, 282 U.S. 101, 109 (1930). This option was not available to spouses living in non-community property states, creating a patchwork of different rules for people in otherwise identical situations. *See* Carolyn C. Jones, *Split Income and Separate Spheres: Tax Law and Gender Roles*, 6 LAW & HIST. REV. 259, 266–67 (1988). A number of states considered and adopted community property laws to take advantage of the ruling in *Seaborn*; some did not. *Id.* at 267–73. In 1948, Congress permitted income splitting between spouses by adopting rate schedules for the joint filing category. *Id.* at 259–60. In essence, Congress created a uniform rule for all married people, regardless of the underlying state law, thereby avoiding the patchwork system that slowly developed as a result of *Seaborn*. *Id.* at 267–68; *see also* Lyeth v. Hoey, 305 U.S. 188, 193 (1938) (definition of inheritance in Internal Revenue Code held not to depend on state law); Drye v. United States, 528 U.S. 49, 52 (1999) (federal tax liens attach to taxpayer property even where state law determines taxpayer has no property interest).

116. Lederman, *supra* note 3, at 1659.

Another difficulty with a property rights-based approach is substantive. It is unclear whether the tax treatment of exchanges and sales should or will differ depending on whether people are deemed to have property rights or use rights in their virtual goods. Lederman concedes that if one has property rights in virtual goods, an exchange will likely count as a realization event under the *Cottage Savings* standard.[117] Accordingly, unless they qualify as exchanges under I.R.C. § 1031 or some other non-recognition provision, such exchanges create taxable income.[118] This result makes sense, as the exchange of items would surely be subject to tax if one party sold his virtual asset to the other for U.S. currency, who then used the cash to purchase a virtual asset from the first party. The form of barter cannot obscure the underlying economic substance of a mutual sale and purchase.

In contrast, Lederman suggests that if a group of people have use rights in a class of goods, as is the case in some worlds, those who exchange those goods among themselves may have a "[s]trong argument that . . . in-game trades are not taxable."[119] First, she notes that such trades do not constitute realization events because no property has actually been exchanged.[120] Accordingly, no tax is due. Second, she cites two real-world examples where people trade items in which they have only use rights without incurring taxable income.[121] Neither of these arguments is compelling.

First, property is often described as a "bundle of rights," distinct from any underlying asset.[122] Thus, even if people are deemed only to have use rights in their virtual goods, these rights can nonetheless be considered a form of property. Indeed, as Bryan Camp explains, they are considered a type of property called a "chose in action."[123] The exchange of such property should lead to realization under the lax *Cottage Savings* standard and therefore to taxation.[124]

Assuming that use rights are not construed as property under state law, the IRS could seek to apply some variation of the substance-over-form doctrine to

---

117. *See supra* note 104.

118. I.R.C. § 1031 (2000). By its terms, I.R.C. § 1031 applies to "[p]roperty held for productive use in a trade or business or for investment. . . ." *Id. See also* Lederman, *supra* note 3, at 1650–51 & n.161 (citing I.R.C. § 1031(a)(1), (b)). Thus, any claim that virtual property qualifies for § 1031 treatment would be inconsistent with the notion that it should escape taxation because those exchanging property are engaged in consumption activities. And while § 1031 certainly has supporters, the theoretical justification for such a rule seems weak. *See generally* Marjorie Kornhauser, *Section 1031: We Don't Need Another Hero*, 60 S. CAL. L. REV. 397 (1987) (discussing the history of I.R.C. § 1031). Moreover, issues regarding what constitutes like-kind in a virtual context could be troubling. Accordingly, I would be hesitant to extend it to cover virtual goods or beyond the trade and business or investment limitation.

119. Lederman, *supra* note 3, at 1653–55.

120. *Id.* at 1654.

121. *Id.* at 1653–54; *infra* text accompanying notes 128–130.

122. *See* Herwig v. United States, 105 F. Supp. 384, 389 (Cl. Ct. 1952) (describing a "bundle of rights" inherent in a copyright as "[a] 'sale' of personal property rather than a mere 'license'").

123. Camp, *supra* note 3, at 54–55.

124. *See supra* note 104.

find that the states' rights labels have no bearing on the federal tax consequences of these types of trades or exchanges.[125] While the efficacy of such a strategy would be in doubt, the IRS and courts have been sufficiently successful in their efforts to bring about uniform tax treatment and prevent tax avoidance that it cannot be completely discounted.[126]

Accepting, *arguendo*, that (1) use rights are not a form of property; (2) even if they are, the exchange of use rights in the context Lederman describes does not constitute a realization event; and (3) the IRS would not prevail on a substance-over-form argument, the argument for non-taxation may nonetheless falter.

Clearly, not all exchanges of use rights escape taxation. For example, assume Andy has a license to use TurboTax, while Betsy has a license to use Quicken, both of which are owned by Intuit.[127] If Andy charges Betsy $10 to use his copy of TurboTax, he surely would have to pay tax on his gain. Similarly if Betsy were to charge Andy $10 to use Quicken, she would be taxed on her gain. If, instead of engaging in a cash transaction, Andy and Betsy agree to exchange use rights, they should not be allowed to avoid taxation, even in the absence of any realization event. It is not clear from this example why online exchanges of use rights should escape taxation.

Lederman cites to two real-world examples where people trade possession of items to which they have use rights but are not subject to tax.[128] The first involves co-workers who trade office equipment owned by their employer.[129] The second involves passengers on a cruise who trade deck chairs owned by the cruise line.[130] Non-taxation in these examples seems correct and raises the question of whether exchanges of use rights in the virtual world setting should similarly escape tax.

In determining whether these examples present a model to follow in virtual worlds, one must consider why such exchanges are not taxed. If it is for theoretical reasons, then non-taxation of in-world transactions may also be appropriate. However, if it is for fact-based or practical reasons, these examples may be distinguishable.

As demonstrated above, if one considers the cash flow implicit in the exchange of items, those who trade items for which they have use rights have engaged in a mutual sale or rental of such rights. Collapsing the different steps of the transaction should not avoid taxation.[131] Accordingly, the conclusion that we should not tax the exchange of office furniture or deck chairs appears to stem

125. Madison, *supra* note 111, at 716–38 (discussing the various forms of the doctrine).

126. *See supra* note 115.

127. Intuit Corporate Profile (2008), http://about.intuit.com/about_intuit/profile.

128. Lederman, *supra* note 3, at 1653–54.

129. *Id.* at 1653.

130. *Id.* at 1654.

131. If the exchange is deemed a sale, the question of basis arises, but as Lederman notes in her consideration of whether people have basis in virtual goods by virtue of their monthly fee, it seems inappropriate to allocate any of the cost of the cruise to the use right. Lederman, *supra* note 3, at 1649–50. If the exchange is seen as a rental, the issue of basis does not arise.

from different legal rights or practical limitations and circumstances that may or may not exist in the exchange of virtual goods.

As a legal matter, the rights people have to use office furniture and deck chairs may differ in nature from their rights to virtual goods, rendering the analogy inapt. For instance, in Lederman's examples, it may be that each person has use rights in all the property at issue, whereas in the Intuit example above each person only has use rights in their respective program. However, this distinction does not seem to matter, as more than one person usually cannot use property at the same time. Instead, possession matters. One person cannot insist on using a specific deck chair occupied by another because they have a use right in all deck chairs. Where relinquishing possession is compensated by the receipt or use of other property, as a theoretical matter, the recipient has taxable income.[132]

Even if the rights are nominally identical, factual differences may lead to different tax results. For instance, there is no evidence that a market for office furniture use or deck chairs exists, creating significant valuation issues. Non-taxation in such cases makes practical and administrative sense. In contrast, a real, functioning market for most virtual goods exists, whether sanctioned or unofficial.[133] Thus, it is relatively easy to determine the value of use rights for such goods. Moreover, because the market exists, it is possible to convert one's use rights in virtual goods into real-world property or currency.[134] The inability to do so in the deck chair context may influence our thinking on how best to treat the exchange.

Other factors may affect our thinking, as well. For instance, one generally possesses a deck chair for a short period, relinquishing it at the end of each day, if not sooner. Even if one were to hold a chair for the cruise's duration, cruises generally last a short time, and the chair must be given up when the cruise ends. In contrast, virtual worlds have no set end. While people may technically have only use rights, they are not expected to give them up on a set schedule.

The universe of people with whom one may trade on a cruise is also limited, as people can only trade with others on the same cruise. They cannot hold seats for people getting on the next cruise. These restrictions significantly limit the market for the deck chairs. In contrast, virtual worlds are open. Thus, even though one would only trade with a person in-world, new people are constantly joining.

Finally, the level of control the users exercise over the property in question may differ. It seems likely, or at the very least possible, that employers and cruise companies exercise significantly greater control over their property than do the developers of virtual worlds. In other words, if the developers of virtual worlds cede control of virtual goods to the players, the use rights begin to function as property rights, and this difference may affect our sentiments regarding taxation.

In sum, the theoretical arguments against taxing the exchange of items to which participants have use rights fail, although for practical reasons some such

132. *See, e.g.*, I.R.C. § 1031 (2000).

133. *See* discussion *supra* Part II.

134. *Id.*

exchanges are and should be non-taxable. The decision of whether to tax the exchange of use rights in the virtual world context seems to be a judgment call rather than something that should be theoretically determined.

2. Policy Analysis

In the final section of her article, Lederman explores the implications of her doctrinal analysis using the traditional measures of administrability, efficiency, and equity.[135] She also offers an overarching tax policy objective that she contends should inform and, if necessary, override the doctrinal analysis.[136] Lederman distinguishes between activities that can be considered consumption and those that are profit-seeking.[137] She states: "[I]n general, online activity generating a profit should bear taxation, while that which generates mere entertainment value [i.e., consumption] should not [have federal income tax imposed on it]."[138] She contends that "it is inappropriate under an income tax to impose a tax directly on consumption because, given that the funds spent on that consumption were not deductible, taxing the consumption value would effectively tax the activity twice."[139]

This policy leads her to propose different tax treatment for different types of virtual income depending on the world and the transaction in which it is earned.[140] Based on the assumption that those who earn virtual income in game worlds but do not cash it out are engaged in consumption, she argues that such income should be excluded from the tax base as "mere entertainment value."[141] However, it should be subjected to tax if it is cashed out.[142] She contends that it

---

135. Lederman, *supra* note 3, at 1658–70. As part of this analysis, she addresses issues such as difficulty of valuation, compliance concerns, and administrative costs. She also asserts that any tax on virtual income would likely be regressive. In the interest of space, I do not address these arguments here. Instead, I focus on the policy argument she raises regarding the possible double taxation of consumption.

136. *Id.* at 1665.

137. *Id.* at 1659.

138. *Id.*; *see also id.* at 1663 (stating "[f]rom a policy perspective, the right result is not to tax mere entertainment but to tax profit").

139. Lederman, *supra* note 3, at 1659 (citing Joseph M. Dodge, Zarin v. Commissioner*: Musings about Debt Cancellation and "Consumption" in an Income Tax Base*, 45 TAX L. REV. 677, 680–81 n.18 (1991)). Under the Haig-Simons formulation of income, income = consumption + Δ wealth. HENRY SIMONS, PERSONAL INCOME TAXATION 50, 61, 206 (1938). In mapping this formula to the tax system, one focuses both on income inclusion and deductions. Consumption is taxed by disallowing deductions for consumption expenses. Lederman's contention is that disallowing deductions for consumption expenses *and* taxing virtual income produced by consumption activities amounts to the double taxation of consumption.

140. Lederman, *supra* note 3, at 1658–70.

141. *See, e.g., id.* at 1648, 1663. Recall that Lederman uses the term "game world" for unstructured worlds, thereby reinforcing the notion that what goes on in such worlds is consumption. *See id.* at 1624, 1628.

142. *Id.* at 1663.

is the act of cashing out that distinguishes a profit-seeking from an entertainment-seeking motive, and thus, is the distinguishing factor between consumption and profit-seeking activities.[143] For virtual income earned in unscripted worlds, she would exclude gains associated with barter transactions from the tax base because she presumes that barter is "consumption-oriented."[144] In contrast, she concludes that sales in such worlds are likely "profit-oriented" and, therefore, should be taxed regardless of whether the seller cashes out his virtual take.[145]

Lederman's policy analysis suffers from four main problems. First, it overlooks the differences between virtual income and the value obtained in other consumption-oriented activities. Second, it approaches the double taxation issue from the wrong direction by trying to alleviate the purported problem by excluding certain types of income from the tax base instead of allowing offsetting deductions. Third, it requires one to base income inclusion on the nature of the activity that generated it. This approach is tantamount to an intent-based test for income inclusion, contrary to existing tax policy and practice.[146] Finally, assuming that it is appropriate to exclude from the tax base income generated during consumption-oriented activities, the proposal seems poorly suited to accomplish that goal.

Lederman correctly notes that taxing the value of consumption without allowing for a deduction for the expenses associated with it leads to the double taxation of consumption.[147] For instance, if one were to buy a family car, it would be wrong to both deny a deduction for the cost of the car and include in income the value of the car purchased. Most virtual world participants pay for the privilege of participating in their chosen worlds, and therefore, one could argue, as Lederman does, that any virtual receipt is nothing more than what has been purchased.[148] Accordingly, its value, just like the value of the car in the example above, should be excluded from income.

Another way to approach this problem, and one that highlights the difficulty of this argument in the context of virtual income, is to determine whether there has been an accession to wealth. Income tax should be owed only if there is a demonstrable accession to wealth.[149] Under normal circumstances, the benefits received from consumption expenditures will be presumed to equal the expenditure, and there will be no question of gain. To return to the car example above, the value of the car purchased is presumed to be equal to its cost, and therefore buying the car does not lead to an accession to wealth. Accordingly, it

143. *Id.*

144. *Id.*

145. Lederman, *supra* note 3, at 1666. She would permit people to offset the income with the expenses incurred in acquiring the income. *Id.*

146. *See infra* notes 162-166 and accompanying text.

147. Lederman, *supra* note 3, at 1659.

148. *Id.* at 1622, 1658.

149. Comm'r v. Glenshaw Glass Co., 348 U.S 426, 431 (1955).

would be wrong to include the value of the car in income without allowing for an offsetting deduction.[150]

One who pays to participate in a virtual world receives the right to participate and the opportunity to earn virtual income. How to value this right and opportunity is difficult. Arguably, someone who pays to participate but never logs on has received less value than someone who spends eighty hours per week online, and one could judge value received by their usage. Alternatively, one could look to subjective enjoyment regardless of time spent online. Neither of these measures is administrable. Instead, as with the car example above, the only administrable way to determine value is to presume that the right to participate and opportunity to earn virtual income is equal in value to the amount paid. As with the car, the right to play cannot and should not be included in the tax base. It is what was purchased and is not, in and of itself, income.[151]

This leaves open the question of what to do with the income that one actually earns while participating in virtual worlds. If the right to participate and the opportunity to earn income are considered the value purchased with the access fees, then any income earned reflects an accession to wealth, and it should be fully taxable.[152] One could argue that the opportunity to earn income has an expected value and that the receipt of virtual income up to that amount should be considered part of what was purchased. However, it would be hard to determine what that expected value was or when it had been exceeded. Regardless, skilled players can earn virtual income far in excess of the access fees paid such that there can be no question of an accession to wealth.[153]

The argument for taxing virtual income is analogous to that for taxing lottery winnings. Arguably, the person who buys the lottery ticket has "purchased" the

---

150. In fact, the value received may differ from the price paid, depending on a person's utility curve. For instance, if I pay $15,000 for the car, but do not value it at $15,000, I arguably have not received $15,000 of value. Conversely, if I love the car and would have paid more, I may have received far more value than I paid. Under normal circumstances, we presume for tax purposes that the market establishes the value paid and received, thus avoiding the need for subjective or difficult evaluations. *See, e.g.*, Treas. Reg. § 20.2031-6 (as amended in 1960) (articulating the willing buyer/willing seller standard to be used in valuing property for estate tax purposes).

151. Alternatively, one could include the value in income and allow a deduction for the expense. As these amounts would offset, the result of no tax would be the same.

152. Assuming no profit motive, I.R.C. § 183(b)(2) (2000) permits taxpayers to deduct the cost of earning income in an activity to the extent of income earned in that activity. Thus, the access fees may be deductible against the value of virtual income earned. Where those fees equal or exceed the income, no tax would be due. The theory underlying this rule is that taxing income without allowing a deduction for expenses would lead to double taxation. One could argue that allowing a deduction for access fees under § 183 is improper here because the expenses purchased the right to play, the value of which was excluded from income. Allowing a deduction would amount to a double benefit. A deduction would be appropriate only if the value of the right to play were included in income. In contrast, virtual income earned in-world reflects value in excess of the fees paid and is therefore previously untaxed income. Accordingly, it should be taxed in full, without allowing a deduction for participation fees. I am not advocating this.

153. *See supra* note 33 and accompanying text.

outcome of the lottery. Therefore, they can argue if they win that the winnings are merely what was purchased and should not be included in the tax base, just as the value of the car one purchases should not be included. However, in this case, the consumption has lead to an accession to wealth, and tax should be paid on that accession. While the person who plays the lottery has paid to participate, and the pleasure of participating is properly excluded from the tax base, if he or she wins, the winnings should and will be taxable.[154]

The question is how best to tax accessions to wealth while avoiding double taxation of consumption. Options include allowing a deduction for the expense that generates the income or excluding the value received from the tax base. Lederman takes the latter approach and proposes exempting from the tax base certain types of virtual income depending on whether the activities that generated the income in question are deemed "consumption-oriented" or "profit-oriented."[155] This approach is inconsistent with existing law and policy. Tax law avoids imposing a double tax on consumption by including *all* income in the tax base, but allowing deductions for expenses to the extent the expenses are associated with such income.[156] This can readily be seen in the rules regarding gambling and not-for-profit activities.[157] For example, if someone earns $100 in a hobby but incurs expenses to do so, he must include the $100 in income but is allowed to net out expenses up to $100 such that he will only pay tax on profits. In the context of virtual worlds, this means that participants should include all of their virtual income in income but be allowed to deduct their subscription fees costs to avoid a double tax.[158]

Lederman's proposal that virtual income be excluded from the tax base to avoid double taxation of consumption treats virtual income inconsistently, terming it "mere entertainment value" when left in-world but an accession to

---

154. *See* Lange v. Comm'r, 90 T.C.M. (CCH) 69 (2005).

155. Lederman, *supra* note 3, at 1663. While Lederman couches the argument in the language of consumption, on closer examination it becomes clear that she equates consumption activities with leisure activities. *See id.* at 1662. Indeed, the link between consumption and leisure can be seen in the terminology she chooses. In referring to "game worlds," she rhetorically attributes a non-profit-seeking motive to those who play. *Id.* It is precisely because they are playing that participation in these worlds can and likely should be seen as consumption.

156. Lederman herself alludes to this by noting that a taxpayer in Second Life would not owe any taxes on sales in Second Life if his expenses exceeded income in a given year. Lederman, *supra* note 3, at 1666. However, she takes this approach only *after* she has previously determined that the income from the sales should be included in the tax base. *Id.* In my opinion, all income should be included in the tax base, regardless of the presumed intent, and the taxation of such income should depend on the existence and quantity of relevant expenses.

157. *See* I.R.C. §§ 166(d), 183 (2000).

158. As noted previously, if all virtual income is deemed an accession to wealth, the argument for deducting these costs is significantly weakened. Allowing a deduction of those fees would lead to under-taxation because the value of the right to play is not included in the tax base. In the analogous situation of a lottery, we allow taxpayers to deduct the cost of the ticket and pay tax on the net winnings. I would follow the lottery example and allow costs to be deducted, despite this argument.

wealth if cashed out.[159] The value to the taxpayer remains the same regardless of whether he or she converts it to a real-world currency and therefore virtual income should be considered an accession to wealth even if it remains in-world. This approach effectively imposes a form of realization requirement on virtual income that allows for tax deferral.

Moreover, Lederman's non-taxation of barter in unscripted worlds, on the theory that to do so would lead to a double tax on consumption, is problematic because taxing barter does not represent a tax on consumption. Instead, it is a tax on an accession to wealth that was delayed by virtue of the realization requirement. Thus, taxing barter transactions does not impose a separate, double tax on consumption. If it did, this logic would apply equally to barter exchanges of real-world property, where the goal was consumption, rather than profit. For instance, if Catherine trades her appreciated surfboard for Darcy's appreciated hang glider, both are pursuing consumption. If Lederman's analysis were extended to this real-world barter, no tax should be due. This result, for which Lederman does not argue, is clearly contrary to existing law, as Lederman herself acknowledges.[160]

The third problem with Lederman's approach is that it requires one to distinguish between consumption-oriented and profit-seeking activities. Activities standing alone cannot be classified as consumption or profit-seeking. For instance, I may participate in a virtual world solely for entertainment, or in other words, as a form of consumption. In the course of my activities, I may trade with others, buy and sell goods, or go on quests. Someone else might engage in the exact same activities but with the aim of earning money. Indeed, as described above, a number of college students are forgoing traditional summer jobs to earn virtual income, which they can later convert to U.S. currency.[161]

The only way to determine the appropriate tax treatment under this approach is to ascertain or to make assumptions about intent.[162] Inquiries into intent are notoriously difficult and are generally eschewed in the tax law, when possible.[163] Moreover, under existing law, while intent and the type of activity matter for purposes of determining whether and to what extent to allow deductions,[164] they

159. *Id.* at 1662.

160. *Id.* at 1624.

161. *See* Atler, *supra* note 46, at W1.

162. Rather than look at individual intent, Lederman suggests creating presumptions regarding the nature of online activities based on the players' actions. Lederman, *supra* note 3, at 1663. For instance, she presumes that any participant in a "game" world who leaves her income in-world is engaged in consumption. *Id.* As a result, she would classify the virtual income received as "entertainment" and outside the tax base. *Id.* However, if that same person were to cash out his or her virtual income, Lederman presumes that he or she has manifested a profit-seeking intent and should be taxed on the value of his or her virtual income. *Id.*

163. When tax law includes intent-based tests, Congress, the IRS, and the courts generally look to objective indicia of intent. *See, e.g.*, I.R.C. § 183 (2000) (listing factors to be considered in presuming profit-seeking intent).

164. *See, e.g., id.*. § 162 (allowing trade or business deductions); *id.* § 163 (allowing interest deductions); *id.* § 183 (allowing hobby deductions); *id.* § 212 (allowing for-profit deductions); *id.* §

are completely irrelevant when determining whether something should be included in income.[165] For instance, a child playing with trading cards would have income on the sale or exchange of a card for some other type of property, regardless of whether he or she intended to make a profit or simply complete his or her collection.[166] If virtual income represents an accession to wealth, it should not matter that it was generated during a consumption activity.

Finally, assuming, *arguendo*, that virtual income earned in consumption-oriented activities should be excluded from the tax base, Lederman's approach is poorly suited to achieve her policy goal. The propriety of Lederman's proposal depends on its ability to differentiate between those engaged in consumption and those seeking to make a profit.[167] Intent is quite difficult to discern, and any bright-line rule will likely be both over- and under-inclusive, taxing some who should go untaxed, and failing to tax some who should be taxed. Lederman's proposal infers intent from behavior, thus avoiding the need for a case-by-case analysis of a participant's state of mind.[168] However, the use of broad categories is too inaccurate and susceptible to significant abuse, at least with regard to unscripted worlds.

At the moment, Lederman's proposal to adopt a cash-out rule for game worlds seems appropriate given her policy analysis.[169] Most game world developers take significant steps to ensure the integrity of their games. For example, Blizzard Entertainment, the developer of World of Warcraft, and Sony, the developer of EverQuest, have banned sales outside the context of the worlds to protect the integrity of the games, threatening to punish those who violate the rules, and actually doing so in many cases.[170] Moreover, given the potential volatility of virtual economies, hoarding large quantities of gold in game worlds is risky, and those seeking profits will likely cash out their earnings.[171] Thus, it seems appropriate to infer that those who participate in game worlds and leave their income in-world are engaged in leisure activities, and the possibility and consequences of misclassification of taxpayers, by allowing them to defer taxation until they cash out, seem small. If the goal is not to tax pure consumption, a cash-out rule for game worlds seems reasonable.

---

262 (disallowing personal deductions).

165. Camp, *supra* note 3, at 47, 61.

166. *Id.* at 61–62. However, I.R.C. § 1031 would not apply to a child trading baseball cards, as we assume he or she has no intent to make a profit. I.R.C. § 1031.

167. *Id.*

168. *Id.* at 1663.

169. *Id.* at 1648.

170. *See supra* note 32; *see also* Brian O'Halloran, *Blizzard Files Lawsuit against Gold Spammer*, WOW INSIDER, May 26, 2007, http://www.wowinsider.com/2007/05/26/blizzard-files-lawsuit-against-gold-spammer (announcing Blizzard's filing of a lawsuit against a problematic player in World of Warcraft).

171. Indeed, Lederman notes that most who participate in game worlds to earn money typically cash out quickly so that they can purchase real world items. Lederman, *supra* note 3, at 1648. This triggers tax on their gains and eliminates any significant deferral advantages. As virtual worlds evolve, this may not always be the case.

Nonetheless, some subset of people may participate in game worlds to hoard virtual gold in the hopes of getting rich.[172] While this number is likely to be quite small, that could change. According to Edward Castronova, so long as it is possible to cash out, whether legitimately or in violation of the world's rules, the number of people engaged in more than play is likely to increase over time, as the profit-seeking meme displaces the leisure meme.[173] Assuming he is correct, at some point the underlying reality may change sufficiently that a cash-out rule for game worlds will no longer be appropriate.

Lederman's proposed rules for unscripted worlds are far more troubling. She presumes that those who barter are likely to be engaged in consumption, and therefore should not be taxed until they cash out, while those who sell are profit-seeking and therefore should be taxed on their profits, regardless of whether they cash out.[174] These claims are difficult to defend. First, as Lederman acknowledges, it is likely that some people buying and selling items are engaged solely in consumption.[175] This is precisely what one would expect: currencies exist to facilitate trade, because it may not be possible to find someone who has what you desire and who also desires what you have. Even if one finds such a person, trading may not be possible if the goods each person has are of different value. Thus, the likelihood that those who buy and sell goods are engaged in play is quite high. As a result, this proposal could subject to tax a large number of people who purportedly should not be taxed.[176]

Second, it is not clear that the majority of those engaged in barter are engaged in consumption. Lederman uses an example of someone exchanging a copy of a T-shirt for a copy of some virtual jeans.[177] It is not clear that this example is emblematic of most barter exchanges, thus warranting an exclusion from income of all such exchanges. For instance, assume that the real Levi Company (Levi) opened a virtual store, selling virtual jeans. Further assume that Levi has two goals: to earn Lindens, which it can then exchange for dollars, and to advertise its real-world goods by encouraging people to clothe their avatars in virtual jeans.

Next, assume that Levi opens a "second-hand" store and offers to exchange virtual jeans for other virtual clothing that it could sell at this store. Finally,

---

172. *See* Dibbell, *supra* note 33, at 284 (reporting that during 2003 he was able to amass and then convert $11,000 of virtual goods by "playing" Ultima Online); Dibbell, *supra* note 34 (describing the practice of gold-farming). For wealth accumulation in unscripted worlds, *see generally* ROBERT FREEDMAN, HOW TO MAKE REAL MONEY IN SECOND LIFE: BOOST YOUR BUSINESS, MARKET YOUR SERVICES AND SELL YOUR PRODUCTS IN THE WORLD'S HOTTEST VIRTUAL COMMUNITY (2008) (guiding readers on how to make a profit using Second Life); DANIEL TERDIMAN, THE ENTREPRENEUR'S GUIDE TO SECOND LIFE: MAKING MONEY IN THE METAVERSE (Candace English et al., eds., 2008) (discussing how to successfully start and run a business in Second Life);

173. Castronova, *supra* note 3, at 200.

174. Lederman, *supra* note 3, at 1663.

175. *Id.* at 1666.

176. As Lederman notes, under existing law, such people would only be taxed on their profits, as they would be allowed to deduct their expenses under I.R.C. § 183 as hobbies. *Id.* at 1666.

177. *Id.*

assume that Chloë, who has made a virtual sweater for her avatar and participates only for fun, decides to trade it for a pair of virtual 501s. This barter exchange would entail one party seeking to make a profit and the other engaging in consumption.[178] It is not hard to imagine barter transactions where both parties seek to make a profit. For instance, Chloë may well be dressing not for a virtual party, but rather in an effort to earn currency that she can later cash out and use in the real world. Absent some empirical study to support the division between those engaged in leisure activities and those seeking profits, it is not possible to evaluate whether Lederman's proposal accomplishes the stated policy goal.

Even if we could conclusively demonstrate that most barter transactions currently fall on the leisure or consumption side of the line, while most sales fell on the profit side, adopting a rule that would tax one but not the other would almost certainly lead to a dynamic response. Those seeking to earn a profit would structure their transactions to receive goods instead of currency to avoid taxes. Indeed, one can imagine participants in virtual worlds designating some good to serve as an alternate virtual currency to avoid taxation. Such a rule would impose significant complexity and inefficiency on the virtual world. And, it would lead to a (virtual) reality completely inconsistent with the factual assumption underlying the non-taxation of barter transactions. As a result, Lederman's proposal for unscripted worlds likely will fail to accomplish its underlying policy goals.[179]

3. Conclusion

In sum, both Lederman's doctrinal and policy approaches are problematic. She places drops in the wrong conceptual category for tax purposes.[180] Her rights-based approach to sales and exchanges requires difficult and time-consuming analysis, leading to the possibility that identical transactions will yield different tax results.[181] This approach also leads to the possibility that the government and virtual world developers will embark on a game of cat and mouse, as developers seek to modify the rights they confer to affect their desired tax results. Substantively, the distinction she attempts to draw between the exchanging of property and use rights finds less support than she suggests.[182]

Lederman's policy of not taxing the fruits of consumption for fear of double taxation is correct, but only to the extent that the value created does not lead to an

---

178. In fact, Levi could be seeking to make a virtual profit *and* to increase its real-world profits. For instance, one goal may be to gather goods in trade, sell them, and then cash out. Another may be to generate real-world sales of Levis by "advertising" them on willing avatars. Regardless, Levi would be seeking a profit and should be subject to tax under the theory Lederman advances. *Id.* at 1659.

179. Lederman notes that barter transactions present valuation problems that sales avoid. Lederman, *supra* note 3, at 1663. However, that issue is distinct from the question of whether virtual income should be considered income in a theoretical sense.

180. *Id.* at 1659.

181. *See id.* at 1658.

182. *Id.*

accession to wealth.[183] Unlike other types of consumption, participation in virtual worlds may well lead to an accession to wealth, which should be taxed.[184] The better approach to the double taxation concern is to include all income in the tax base and allow for deductions where appropriate. Approaching the question from the opposite direction yields an intent-based test, which is inconsistent with current law, and raises a host of practical problems, namely the difficulty of determining which online activities are consumption oriented and which are not. Even accepting this policy goal, taxing sales in unscripted worlds, while allowing exchanges to go untaxed, seems ill-suited to the task of separating those seeking profit from those engaged in consumption. Moreover, it is susceptible of manipulation, such that, even if it were appropriate *ab initio*, the dynamic response to such a rule would soon render it inadequate.

B. The Imputed-Income Approach

Bryan Camp proposes a different approach to the question whether the government should tax virtual income: he contends that income generated through in-world activity constitutes imputed income and therefore falls outside of the current doctrinal definition of gross income.[185] He also contends that virtual income is analogous to casino chips, for which the IRS appears to have adopted a cash-out rule.[186] Finally, Camp makes a policy-based argument that virtual income should be excluded because a "magic circle" or "fourth wall" separates virtual reality from our own.[187] He recognizes that these boundaries might one day crumble and concedes that virtual income left in-world should be subject to tax when in-world activities displace real-world economic activity, as evidenced, for instance, by real-world businesses accepting virtual currency.[188]

This approach has many virtues. Tax results do not depend on the type of world, the type of transaction involved, or the kind of legal rights one has in one's virtual property and currency. Rather, one rule applies to all virtual income, and tax is imposed only when a person cashes out, thus establishing the real-world market value of the virtual goods. It also has the virtue of keeping the IRS out of virtual worlds, a result most virtual-world developers and participants would cheer. Nonetheless, this approach also suffers from a number of problems. First

183. *See id.* at 1659.

184. *See* Comm'r v. Glenshaw Glass Co., 348 U.S 426, 431 (1955).

185. Camp, *supra* note 3, at 60.

186. *Id.* at 64.

187. *Id.* at 60.

188. *Id.* at 66. Camp expressly sets aside a number of questions that might arise if virtual activity were considered taxable, including whether someone might be allowed deductions for online activities because they qualify as a trade or business under I.R.C. § 162 and how the hobby loss rules of I.R.C. § 183 might apply. *Id.* at 14 n.34. He also sets aside questions of how one should report income earned from cashing out and the consequences of factual distinctions or pending legal issues, including questions of whether one is selling property, a copyrighted work, a license, etc. *Id.* at 70–71.

and foremost, virtual income cannot properly be classified as imputed income. Second, the analogy to casino chips is strained and assumes that the accounting convention used for chips reflects theoretical, as opposed to practical, considerations. Finally, determining when the boundaries surrounding virtual worlds should be deemed porous enough to warrant taxation of in-world transactions creates a host of administrative problems.

1. Doctrinal Analysis

Camp begins his doctrinal analysis by conceding that, absent some intervening doctrine, virtual income should be included in the tax base because it represents a clearly realized accession to wealth.[189] He offers two justifications for the non-taxation of virtual income. First, he asserts that virtual income should be considered imputed income and thus excluded from the tax base.[190] He contends that trading of virtual goods in-world are "not normal market transactions but represent self-provided services or, at most, enjoyment of self-owned property."[191] Second, Camp argues virtual income is analogous to the casino chips in *Zarin v. Commissioner*,[192] which the Third Circuit held represented an opportunity to gamble and not, in and of themselves, money.[193] Accordingly, no tax should be due until someone cashes out either his chips or his virtual income.[194] While these arguments may have some initial appeal, on closer examination they fail.

Camp defines imputed income as "a flow of satisfactions from . . . goods and services arising out of the personal exertions of the taxpayer on his own behalf," and describes the policy reasons for excluding such income.[195] Camp observes

---

189. Camp, *supra* note 3, at 47.

190. *Id.* at 44 ("The concept of imputed income doctrine provides the appropriate protections for taxpayers in all virtual worlds and, more importantly, tells us at what point in-world transactions ought to be taxable."). Camp identifies three possible limitations that can affect whether an item should be considered income for federal tax purposes. They include: (1) difficulty of valuation, which he calls "Priceless"; (2) realization, which implicates timing and is often associated with liquidity concerns; and (3) imputed income, which implicates both difficulty of measurement and a level of government intrusiveness that would be unacceptable. *Id.* at 49–61. Camp concludes that neither the priceless nor realization limitations would bar the taxation of virtual income. Instead, he concludes that "[i]t is the concept of imputed income that draws the proper line." *Id.* at 61.

191. *Id.* at 60.

192. Zarin v. Comm'r, 916 F.2d 110, 114 (3d Cir. 1990).

193. *Id.* The reasoning in *Zarin* has been highly criticized. *See, e.g.*, Dodge, *supra* note 139, at 678; Calvin H. Johnson, Zarin *and the Tax Benefit Rule: Tax Models for Gambling Losses and the Forgiveness of Gambling Debts*, 45 TAX L. REV. 697, 698, 706 (1990) (concurring with the result, but disagreeing with the reasons set forth in the opinion). In the interest of space, I do not reargue here the wisdom of that case, but rather focus on whether it could properly be applied to virtual income.

194. Camp, *supra* note 3, at 64.

195. Camp, *supra* note 3, at 37 (quoting Donald B. Marsh, *The Taxation of Imputed Income*, 58 POL. SCI. Q. 514, 514 (1943)).

that taxing self-performed services would create a strong incentive for people not to perform them.[196] At its most extreme, taxing self-performed services could lead to health issues, as people might refrain from cleaning their bathrooms or refrigerators for fear of being taxed.[197] Despite these concerns, Camp argues the strongest policy reasons for not taxing imputed income from services are practical.[198] The services at issue occur outside the marketplace, and the difficulties in determining their value, combined with their sheer volume "would create an administrative nightmare for taxpayers and the IRS alike,"[199] requiring "intrusive oversight" that simply would be unacceptable to most people.[200] These same problems exist with respect to imputed income from property.[201]

Camp would classify all virtual wealth as imputed income because it results from self-performed activities and because it is self-benefiting in that it is used to enhance one's play.[202] I take a somewhat different view of the issue. As noted above in connection with Lederman's discussion of drops, most virtual wealth is received from third parties, whether developers or other virtual world participants, thus precluding its classification as imputed income.[203] As with real-world income, the nature of virtual income depends on how it is acquired. I thus consider self-created virtual income, drops, sales and exchanges in turn.

First, the question of self-created assets. Camp correctly notes that in the real world self-created items are generally not included in the tax base.[204] For instance, artists are not taxed when they create a new work of art. Instead, such items are taxed upon disposition. While self-created assets are often talked of as a form of imputed income,[205] as explained above,[206] an asset's value is different from the imputed value of the self-performed services employed to create or

196. *Id.* at 37–38.

197. *See id.*

198. *Id.* at 38.

199. *Id.*

200. *Id.* at 41.

201. Camp notes that valuation problems could possibly lead to arbitrary enforcement and create enforcement problems. Camp, *supra* note 3, at 43. Moreover, the imputed value of the property is likely to vary over time. For instance, the rental market for property often fluctuates with the value of housing and home loan interest rates.

202. *See id.* at 60.

203. As further evidence that virtual income is not imputed income, I would note that imputed income cannot be transferred to another individual. If I shave myself, I have received the imputed value of the self-performed service. Similarly, if I own and live in a home, I have received a benefit valued as the rental value of the house. I cannot transfer the imputed value received to another. Most virtual goods can readily be transferred and therefore cannot be imputed. This is not to suggest that transferability is an element of the definition of imputed income. Rather, the very nature of imputed income precludes its transfer. Nor can restrictions on transfer transmute regular income into imputed income.

204. *Id.* at 37.

205. *See, e.g.*, Dodge, *supra* note 71, at 705.

206. *See supra* Part III.A.1.a.

obtain the asset. Thus, categorizing self-created assets as imputed income seems wrong.

Dodge offers a different explanation for excluding self-created items from income, contending that such property reflects an investment, or a means to making profit, but not the profit itself.[207] Such assets are simply fruit on the metaphorical tree, and until the fruit is separated from the tree when a market transaction occurs, no realization occurs and taxation is inappropriate. Whichever theory one adopts, there appears to be no reason to treat self-created virtual items differently from real-world items. Both are properly excluded from income.

Second are situations where people simply hold their virtual goods, an issue Camp does not address. In such cases, people clearly derive some benefit from holding those goods. The value of that benefit can be imputed by considering what the person would have received if he had entered the marketplace and rented the good to someone else. This benefit clearly fits the definition of imputed income. Consistent with the treatment of all other types of imputed income, the benefits that flow from the ownership or possession of virtual goods or the performance of services for oneself in-world should not be subject to taxation, and no one has suggested taxing this type of income.

Third, consider drops. As described above in Part III.A.1.a, drops fit poorly into the traditional definition of imputed income.[208] Imputed income is the value someone could obtain if he entered the market and rented a good he owns or performed a service that he had performed for himself.[209] One attributes to the taxpayer rental or service income he could have received, but actually did not. The value of a drop is not imputed. Instead, the recipient actually acquires something of value from the world's developer—a third party—for performing a task. Thus, the need for imputation does not exist.[210]

This leaves open the question of what drops are and whether they should be taxed. As previously discussed, Lederman argues that drops fall under Dodge's theory of "taken" items because players work hard to acquire them, just as fishermen work hard to acquire their catch.[211] Just as fishermen are only taxed on the disposition of their catch, which is considered inventory, she argues drops should be taxed only on their disposition.[212] In contrast, I argue that drops are properly considered prizes and therefore subject to tax under I.R.C. § 74.[213] Whatever one's conclusion regarding the true nature of drops, it seems that Camp's contention that drops are a form of imputed income is surely wrong.

---

207. Dodge, *supra* note 71, at 694.

208. *See* Lederman, *supra* note 3, at 1644–45.

209. *See* Camp, *supra* note 3, at 37–44.

210. While one could argue that the value of the good received reflects the value of self-performed services necessary to obtain the drop, the value of the good received may well be different from the imputed value of the services. *See supra* Part III.A.1.a.

211. Lederman, *supra* note 3, at 1647.

212. *Id.*

213. *See supra* Part III.A.1.a.

Finally, consider in-world sales and exchanges of virtual goods. Accepting, that the acquisition or holding of virtual goods could be classified as imputed income, it does not follow that the sale or exchange of virtual property yields imputed income. Rather, in such cases, people have entered a market and have received something of real-world value from someone else in return. In other words, they have actual income as a result of the trade. Camp's assertion that exchanges of goods are "not normal market transactions but represent self-provided services or, at most, enjoyment of self-owned property"[214] flies in the face of what actually happens.

This assertion can be demonstrated with a simple example. Imagine two neighbors, each of whom has a garden in the backyard. One grows tomatoes; the other grows squash. For purposes of this example, I will accept that such crops are imputed income and should not be taxed. As soon as the growers sell the produce, however, they have entered the market and have received value for their services, goods, or both. The policy reasons relating to valuation and government intrusiveness for excluding imputed income no longer apply. Indeed, as they have received real income, the need to impute disappears. Accordingly, those who sell self-grown crops must pay tax on the proceeds of any sale.

Similarly, if the neighbors trade tomatoes for squash, they will be subject to tax, based on the fair market value of the produce each receives. Simply exchanging imputed income in one's hands for imputed income in another's hands does not convert the gain into imputed income. Rather, by engaging in barter, people have entered the market, received real income, and under current tax law must be subject to taxation.[215] That it is a virtual market or that the people engaged in the barter are simply playing is irrelevant. The theory of imputed income cannot exclude exchanges of virtual goods from the income tax base.[216]

Camp's second doctrinal argument is equally difficult. Based on the reasoning set forth in *Zarin*,[217] Camp contends that virtual wealth should be analogized to casino chips, which he argues represent an "opportunity to play," as

214. Camp, *supra* note 3, at 60.

215. *See* Rev. Rul. 79-24, 1979-1 C.B. 60 (concluding that barter creates income). Note that the justification for taxing the barter transaction here is distinct from the justification that taxing barter is necessary to protect the tax base, as people might trade goods to escape a tax on cash receipts. Here, the exchange is taxed because something of value is received from a third party. Certainly, valuation problems will arise in any barter transaction, but Camp has expressly rejected such limitations in a non-imputed income context. Camp, *supra* note 3, at 25.

216. Miano also addresses the question of whether virtual income can be considered imputed income and rejects the claim. Miano, *supra* note 3, at 38–44. However, he does so using an example of a virtual gaming company that sells stock and compares it to a real-world company engaged in the same business. *Id.* As explained above in the text, it is not the similarity between real-world and virtual activities that precludes a determination that virtual wealth creation is a form of imputed income. Nor is it the build-up of value as a result of personal effort or the market. Rather, it is the receipt of something of value from a third party that takes such wealth out of the imputed income category.

217. Zarin v. Comm'r, 916 F.2d 110, 114 (3d Cir. 1990).

opposed to income.[218] Moreover, he notes that taxes generally are not determined on one's gambling winnings until one cashes out.[219] He argues that *Zarin* and the underlying policies apply equally to virtual income.[220] This analogy fails for a number of reasons.

First, discussions of casino chips in the case law generally focus on the receipt of chips from the casino or by the casino from the customers. By and large, they involve chips that were advanced to customers in return for a marker, which is a promise to pay.[221] Translated into the language of virtual reality, this might arguably equate to game world drops. Developers make drops available to players, possibly in an effort to keep players engaged so that they will continue to play and pay their monthly fees. However, this analogy is strained, as players must complete tasks to obtain their drops. Players are not expected to repay the developers the value of the drops at some later point. It would be a mistake to generalize court rulings about the tax attributes of chips advanced customers and apply them to virtual income.

Second, the legal issue in *Zarin* and its underlying facts are sufficiently different from those at issue in the virtual world context that it is inappropriate to apply the reasoning and holding of that case here.[222] The legal issue in *Zarin* was whether the casino's decision not to require Mr. Zarin to repay the chips provided to him would subject Mr. Zarin to cancellation of indebtedness income.[223] Notably, because of his debt to the casino, Mr. Zarin was not able to cash out his chips, but was required to gamble them at the casino.[224] Under those facts, the claim that chips merely represent an opportunity to gamble makes some sense.

Depending upon the type of world, people participating in virtual reality can and routinely do cash out by converting their virtual wealth into real-world wealth. In this context, the "units of play" argument makes less sense.[225] While virtual income certainly makes it possible for participants to continue playing, where players have the opportunity to cash out, receiving income that can be

---

218. Camp, *supra* note 3, at 64. In some respects, this argument is analogous to Joseph Dodge's argument that "taken" items represent an investment that may lead to income but should not be considered income in-and-of-themselves. Dodge, *supra* note 77, at 694.

219. Camp, *supra* note 3, at 64.

220. *Id.* at 65.

221. For instance, *Zarin* focused on the tax consequences of Zarin's receipt of chips from a casino in return for his marker and his subsequent release from any obligation to repay his debt to the casino. *Zarin*, 916 F.2d at 114. Another case, *Flamingo Resort v. United States*, 664 F.2d 1387, 1388 (9th Cir. 1982), involved the receipt by a casino of chips it had lent to a patron in return for a marker. The legal issue was whether the receipt of the chips qualified as income under the accrual method of accounting, or more specifically, whether the "all events" test had been met. *Id.*

222. For a discussion of the *Zarin* case, see Theodore P. Seto, *Inside* Zarin, 59 SMU L. REV. 1761, 1763 n.23 (2006) (discussing the law surrounding taxable discharge of debt).

223. *Zarin*, 916 F.2d at 112.

224. *Id.* at 114.

225. Some worlds allow people to cash out, while others prohibit cashing out. Even so, for those worlds that prohibit cashing out, gray markets exist making it possible—though contrary to the rules—to cash out. *See infra* Part IV.

cashed out is a measurable accession to wealth that increases the recipient's ability to pay real-world taxes.[226]

Finally, the cash out rule associated with casino chips can be seen as nothing more than an accounting method. Gambling losses are deductible to the extent of gains.[227] Thus, it makes little sense to calculate gains and losses separately for each bet. Rather, it is more efficient to tally the wins and losses when the gambler cashes out. Thus, the treatment of casino chips does not depend on the theory that they are merely "units of play,"[228] and that such units are excludible from tax until they are converted to currency. Instead, it reflects administrative convenience by treating gambling as an open transaction until a player cashes out. In an online setting, where wins and losses are carefully tracked, the need for this practice seems less clear.

The practice of assessing winnings only when a player cashes out could lead to abuse where a taxpayer gambles across two accounting periods. By leaving his winnings in the form of casino chips over New Year's Eve, a calendar-year taxpayer may effectively delay the tax owed to the next year. While this certainly must occur, with the IRS allowing it to happen,[229] as a matter of tax law it seems highly unlikely that a taxpayer challenged on such a ploy could successfully argue that he did not have income until he cashed out. The receipt of chips should be taxed as income in the year they are won and cannot be excluded on the theory that they are merely "units of play," exempt from income tax.[230]

2. Policy Analysis

Camp bases his policy argument for non-taxation of virtual wealth on the claim that people engaged in virtual worlds are "playing" because they are engaged in an activity that is make-believe and separate from real-world activity.[231] Using the metaphors of the "magic circle" and the "fourth wall" that separates the actors from the audience, he argues that what happens in virtual worlds is pretend, and one should not be taxed if he pretends to have earned $1 million.[232] However, he concedes that the circle may break and the wall may crumble, so that taxation would be appropriate.[233] Camp identifies this point as occurring where "economic activity in Second Life begins to displace economic activity outside Second Life."[234] He notes that "[t]he most likely evidence of that

---

226. *Id.*

227. I.R.C. § 165(d) (2000).

228. Camp, *supra* note 3, at 2.

229. *See generally* Treas. Reg. § 31.3402(q)-1 (as amended in 2000) (providing for tax withholding from gambling winnings).

230. See Camp and Seto, *supra* note 3, for a debate regarding the constructive receipt doctrine in this context.

231. Camp, *supra* note 3, at 60.

232. *See id.* at 60–61.

233. *Id.* at 69.

234. *Id.*

shift will be when account owners gain the ability to trade Linden Dollars for real goods and services that are useful outside of Second Life, beyond the fourth wall."[235] At that point, the virtual currency will have morphed into a real currency, and "you will not be able to tell the players from the audience."[236] Until then, "concerns about government overreaching and intrusion implied by the imputed income doctrines should outweigh the attraction of taxing trackable transactions of objects that have a readily ascertainable fair market value."[237]

Camp's policy argument raises two concerns. The first relates to the notion that we should not tax transactions that are considered to be pretend. The second relates to the difficulty of deciding when the boundaries between virtual worlds and our own are sufficiently porous so as to warrant taxation.

In support of the notion that we should not tax income earned in pretend environments, Camp uses the example of Zelda, who designs and sells dresses in virtual reality. He states: "Zelda is not a wedding dress designer; she simply plays one in Second Life."[238] Her skill allows her to play more within the Second Life world and should be considered play or pretend.[239] Accordingly, Zelda should not be taxed on her virtual income.[240] This example raises the question of just what Camp means by pretend. Several possibilities exist. For instance, pretend could refer to the fact that Zelda simply pretends to be a dress maker when she is not. However, whether someone pretends to be something they are not is not the relevant question for tax purposes. Each year, people pretend to be lawyers, duping their clients into paying for their legal services. Such would-be lawyers cannot escape taxation by claiming they were just pretending.[241]

Pretend could also relate to a participant's state of mind or purpose in engaging in virtual activities, where those who are playing are deemed to be pretending, while those actively seeking profit are not. If this is the meaning of pretend, determining whether Zelda is pretending assumes tremendous significance and presents a host of difficulties. In particular, there is no reliable way to determine intent, making such a standard difficult to police. What if Zelda really were a wedding dress designer in real life, and she were using her online persona to promote her real-world business? What if she decided to spend more time designing dresses online than in the real world, precisely because she sought to take advantage of the tax deferral that would arise under the cash-out rule? A number of professions, such as graphic design, can be done equally well in virtual worlds and the real one, making the distinction between purportedly pretend activities in virtual worlds and "real" activities even more difficult to maintain. The graphic artist who does his or her work in Second Life is no more pretending

235. *Id.*
236. *Id.*
237. *Id.* at 60.
238. *Id.* at 66.
239. *Id.* at 60.
240. *Id.*
241. I.R.C. § 61 (2000).

than the one who does so at a bricks and mortar workplace. This meaning of pretend would require a costly and likely inaccurate case-by-case analysis.

More important, accepting the possibility that one could distinguish between those playing and those seeking profit, then this meaning of pretend would excuse those engaged in leisure from taxation, a position Camp rejects.[242] As Camp aptly notes, the question for tax purposes is not whether one seeks to make a profit.[243] The question is whether one actually does so.

Camp's metaphors of the "magic circle" and a play with its "fourth wall" suggest that he uses the term "pretend" to refer to relation of virtual worlds to reality.[244] Under this approach, the determination of "pretend" occurs not on the individual level, but rather at the world level, and it depends upon the separation between any given virtual world and reality.[245] So long as the boundaries between a virtual world and reality remained sufficiently strong, income generated in that world would be deemed "pretend," regardless of whether an individual is simply looking for entertainment or is deadly serious about making money he could use in the real world.

While this approach solves some of the problems identified above, it still falters because what matters for tax purposes is whether someone receives and can exercise control over something of value.[246] I should not be taxed because I imagine I have just won $1 million because the million dollars is pretend and has no value. In contrast, virtual income may well have real world value. As noted above, a number of students are forgoing traditional summer jobs and instead seeking to make money in virtual worlds, with one student apparently earning over $35,000 over the past four years.[247] Such activities and the wealth they generate are hardly pretend, even if they occur in the context of a game or virtual world. Thus, while Zelda may be pretending to be a dress maker, if she receives something that has a real world value, the fact that she obtains it in a virtual world should have no relevance for tax purposes.

I turn next to the question of boundaries, permeability, and the effect permeability has on the taxability of virtual income. For the sake of argument, I accept Camp's contention that the existence of strong boundaries warrants non-taxation of virtual income because such income is "pretend."[248] That raises the question of how to determine when such boundaries have sufficiently deteriorated so as to warrant taxation.[249] As demonstrated below, this inquiry presents a host of difficulties.

---

242. Camp, *supra* note 3, at 61.

243. *Id.*

244. *Id.*

245. *Id.*

246. *See* Helvering v. Horst, 311 U.S. 112, 117–18 (1940).

247. *See* Alter, *supra* note 46, at W1.

248. Camp, *supra* note 3, at 61.

249. As set forth below in Part IV, I concur that the permeability of the boundaries between a virtual world and reality should matter for tax purposes. However, I base this conclusion on the effect such permeability has on a taxpayer's ability to pay his taxes, as reflected by a participant's

In an essay entitled *The Right to Play*, Edward Castronova laments reality's encroachment into virtual worlds, noting that "[v]irtual worlds represent a new technology that allows deeper and richer access to the mental states evoked by play, fantasy, myth and saga."[250] Such access is limited to the extent the real world intrudes upon such worlds, and Castronova proposes special legislation to protect play spaces.[251] He posits two types of worlds: closed worlds, in which game developers retain complete control and in which the outside world, including the tax collector, is kept at bay, and open worlds, where the borders are "considered completely porous."[252] In open worlds, he contends that the outside laws, including the tax laws, would fully apply.[253]

Unfortunately, virtual worlds do not fall neatly into these closed and open categories. Rather, the boundaries between reality and the variety of virtual realities are all permeable to one degree or another. Some worlds, such as Second Life, encourage and facilitate the exchange between worlds, thus approaching completely porous boundaries. Others, such as World of Warcraft, attempt to regulate or limit such exchanges. Nonetheless, even for those worlds that explicitly prohibit exchanges and actively try to enforce their rules, it seems reasonable to assume that the boundaries nonetheless remain porous because of gray markets. Even where the rules are strictly enforced and no grey markets exist, virtual worlds can have real-world impacts. For instance, a Japanese woman was recently arrested for breaking into another's virtual account and killing his avatar after his avatar divorced hers.[254] A woman in Delaware was

---

ability to cash out, and not on the "pretend" status of what happens online.

250. Castronova, *supra* note 3, at 185.

251. *Id.* at 200–05.

252. *Id.* at 201–02.

253. *Id.* at 202. Furthermore, Castronova states:

> To be preserved as play space under the law, the synthetic world would have to conform to standards of construction and policy, just as corporations must conform to such standards in order to retain their special status. For example, an interration would have to maintain strict separation of the synthetic economy from the economy of the Earth. If players can regularly buy and sell assets from the synthetic world for real dollars, the synthetic world is no longer clearly distinct from the outside world's economy; it is no longer a play space, it is a tax haven. A tax haven has no right to special privileges under the real world's law, and its case for interration is weak. In general, interrations would be subject to scrutiny on real world matters. As a result, the lack of good faith efforts to maintain the space as a *play* space could lead to the revocation of the charter.

Id. at 204 (footnote omitted).

In the omitted footnote, Castronova notes that external sales may be difficult to police. *Id.* at 204 n.25. However, such sales result in what appears to be a gift, in-world, as one avatar gives an item to another for compensation received outside of the world's boundaries. *Id.* Castronova suggests prohibiting gratuitous transfers as a means of enforcing the no external sales rules. *Id.*

254. Chicago Tribune Wire Report, *Arrested for Virtual Murder*, Oct. 24, 2008, http://www.chicagotribune.com/news/nationworld/world/chi-arrested-for-virtual-murder-081024-ht,0,6551360.story.

charged with plotting to abduct in real life her virtual boyfriend.[255] With apologies to Las Vegas, what happens online does not necessarily stay online.

Camp's proposal for determining when virtual income should be considered pretend, and therefore excluded from the tax base, maps loosely onto Castronova's open or closed taxonomy. Camp would pick displacement of real-world economic activity as the marker for when a world should be considered sufficiently open as to warrant taxation. He argues that the best evidence of this occurring will be when virtual currencies morph into real ones, i.e., when they are accepted like real currencies.[256] This marker and the evidence necessary to establish that it has been met present a number of difficulties. First, as a matter of theory, one must choose an indicator for determining when real-world economic activities have been displaced. Second, one must choose a threshold above which in-world activities can no longer be considered pretend. There seems no principled way to pick either. Finally, whatever indicator or threshold one chooses, it is unclear how one would gather the information necessary to determine when the line is crossed.

Camp argues that the best evidence that virtual activities are displacing real-world economic activities will be when real-world businesses begin to accept virtual currencies.[257] However, this has already occurred in a few limited circumstances for at least two virtual worlds.[258] If one were to accept this standard, one must decide whether to look at the number of people accepting payment in virtual currencies, the identity of those accepting virtual currency (e.g., first adopters or the common man), or the value of the virtual currency being accepted. It is not at all clear the basis on which one would choose among these indicators.

Each of these indicators requires a threshold above which taxation of in-world activities would be appropriate. Thus, it would be necessary to decide how many people could accept virtual currency or how much currency had been accepted and by whom before taxation would be appropriate. Again, there is no principled way to set these thresholds. Finally, taxing authorities must be able to verify that the threshold has been reached. It is not at all clear how one would go about gathering the data necessary to make an informed decision. No system exists to track who receives virtual currencies for real-world work or the value of such currency.

I would argue that the best indicator that real-world economic activity is being displaced is the value of displaced activities. The question then becomes how one determines that activities have been displaced, and assuming one can

---

255. *Id.*

256. Camp, *supra* note 3, at 69.

257. *Id.*

258. An advertising company, online auction store, and chain of pizza parlors have all indicated that they will accept Lindens, the currency used in Second Life. *See* Lederman, *supra* note 3, at 1667 n. 235. Similarly, in the town of New Oxford in Entropria Universe, one can purchase real-world objects with PEDs, the currency used in that world. *See* Chung, *supra* note 3, at 35 n.198.

identify such activities, to measure their value. In some cases, the determination that real-world activities are being displaced seems easy. For instance, companies as diverse as Nissan, IBM, and Nike have all established a presence in virtual reality.[259] These companies are clearly not pretending, and it is almost certain that their expenditures on in-world activities are displacing other real-world economic activities the companies could engage in.

Other cases are not as easy. Take Zelda, the dress maker.[260] While Camp argues that she is just pretending, she may well be a dress maker in real life who has decided to work online to take advantage of the tax deferral afforded by a cash-out rule. Even if Zelda is simply pretending to be a dress maker and has no intent to make a profit, her virtual activities may well displace real-world economic activities. First, Zelda may simply decide to spend more time playing in virtual reality and less time doing whatever she does to make a living. Second, Zelda's virtual activities may well affect the real world, regardless of her intentions. For instance, Zelda's customers could take her virtual designs and use them in the real world to make real dresses, bypassing real-world dress designers.

Assuming one could determine that real-world economic activity was being displaced and that we could readily value it, we would still have to decide how much economic displacement would have to occur before the boundaries were deemed sufficiently porous. There is no principled way to pick the threshold, nor any way to determine that it has been met.

Finally, regardless of the metric chosen, given the value inherent in virtual goods, the online activities of real-world companies, universities, politicians, and individuals seeking real-world profit, not to mention the fact that some real-world companies actually do accept Lindens, a strong argument can be made that, at least in Second Life, the fourth wall has already crumbled, and the concept that this is all pretend can no longer be maintained.[261]

3. Conclusion

In sum, Camp's doctrinal arguments founder on the claim that virtual income is imputed income and on his attempt to analogize virtual wealth to casino chips.[262] His policy argument, that virtual income should not be taxed because it is pretend, is inconsistent with existing tax policy which keys off of the receipt of real-world value, and it raises a host of difficult line-drawing and measurement issues that defy principled or easy solutions.

IV. A NEW PROPOSAL FOR TAXING VIRTUAL INCOME

In this Part, I set forth my proposal for taxing virtual income. As described in the Introduction, I propose a two-step analysis. In the first step, one must

259. *See supra* note 39.

260. *See* Camp, *supra* note 3, at 48.

261. *See supra* Part II.

262. *See* Zarin v. Comm'r, 916 F.2d 110, 114 (3d Cir. 1990).

consider whether the receipt of virtual income in a given world increases a taxpayer's ability to pay real-world taxes. I argue that the ability to pay is a function of a taxpayer's ability to cash out. If no ability to cash out exists, then the world should be considered closed and all virtual income from that world should be excluded from the tax base. If participants can cash out, then the world should be considered open, and one must proceed to the second step in which one determines whether the different types of virtual income should be included in income under existing doctrinal rules.

Subpart A lays out the policy underlying the first step of the analysis. It focuses upon the ability to pay principle, a core tax principle that the analysis to date has largely overlooked.[263] Subpart B lays out the doctrinal analysis of the different types of virtual income that is necessary for the second step of the analysis. Subpart C sets forth my proposal.

A. Policy Analysis

At its most abstract, the definition of income encompasses the flow of satisfactions each person receives.[264] However, even the earliest thinkers on this issue realized that any attempt to tax satisfactions would be impracticable.[265] First, the flow of satisfactions from any one item would likely differ from person to person, leading to differential taxation based on each individual's unknowable utility curve.[266] Second, it would be difficult, if not impossible, to determine how much tax should be paid on items incapable of valuation. Accordingly, to create a real-world tax system, theorists narrowed their income definitions to focus on those things that could readily be assigned a value in money or money's worth.[267] However, that an item can be valued does not necessarily mean that it should automatically be included in the tax base. Rather, as I hope to demonstrate here, the ability-to-pay principle acts as a further filter when deciding what to include.

The ability-to-pay principle has many permutations. In its most prominent form, it factors into the debate over progressive taxation and whether those who earn more should pay progressively more tax than others.[268] However, it is also

---

263. Lederman touches on this issue, noting that a tax imposed on virtual income earned in World of Warcraft would likely be regressive. Lederman, *supra* note 3, at 1659, 1662. However, she is mainly concerned with the fact that those with the most leisure time are likely to be taxed the heaviest, and she does not focus on the correlation between the additional income and increased ability to pay. *Id.* at 1662.

264. *See* ROBERT M. HAIG, THE CONCEPT OF INCOME—ECONOMIC AND LEGAL ASPECTS, THE FEDERAL INCOME TAX 55 (1921), *reprinted in* READINGS IN THE ECONOMICS OF TAXATION 57 (Richard A. Musgrave & Carl S. Shoup, eds., 1959). Haig acknowledged that conceptually income is a "flow of satisfactions" that includes immeasurable intangibles, such as sunsets. *Id.*

265. *Id.* at 55–56.

266. *See id.* at 56.

267. *Id.*

268. *See generally* Walter J. Blum and Harry Kalven Jr., *The Uneasy Case for Progressive Taxation*, 19 U. CHI. L. REV. 417 (1952) (detailing objections to the major arguments in support of progressive taxation). Significant debate exists regarding the appropriate basis for allocating tax

relevant to the question of which tax base to choose.[269] For instance, one reason we do not have a head tax is that it completely disregards ability to pay.[270] A family of six with an income of $50,000 would simply have no ability to pay a head tax of $10,000 per person.[271] To avoid this problem, one could pick consumption, income or wealth as the tax base, thereby correlating tax liability to the income or assets a taxpayer actually has.

Assuming one has chosen income as the theoretical tax base, the ability-to-pay principle continues to play an important role in determining what goes into the real-world tax base. Taxes are paid in dollars, and items that cannot readily be converted into dollars may not increase one's ability to pay taxes. Thus, even if it were possible to place some value on the satisfaction derived from viewing a sunset, if such satisfaction is not readily monetized, one might think twice about including the value in the tax base. Indeed, doing so may create significant hardships.

This concern for ability to pay may help explain why we do not tax imputed income, even in instances where it can be readily and reliably valued.[272] The receipt of such income does not increase one's ability to pay real-world taxes, as imputed income does not create wealth that can be transferred to another.[273] If we

---

burdens. LIAM MURPHY & THOMAS NAGEL, THE MYTH OF OWNERSHIP: TAXES AND JUSTICE (2002) (arguing that tax systems should be evaluated based on their after-tax effects); Joseph M. Dodge, *Theories of Tax Justice: Ruminations on the Benefit, Partnership, and Ability-to-Pay Principles*, 58 TAX L. REV. 9 (2005) (rejecting the benefits principle in favor of the ability-to-pay principle, but noting that the principle is neutral on the question of progressive taxation); Deborah A. Geier, *Time to Bring Back the "Benefit" Norm?*, 33 TAX NOTES 893 (2004); Deborah A. Geier, *Incremental Versus Fundamental Tax Reform and the Top One Percent*, 56 SMU L. REV. 99 (2003) (arguing for a blend of the benefits and ability-to-pay principles of taxation); Susan Pace Hamill, *An Evaluation of Federal Tax Policy Based on Judeo-Christian Ethics*, 25 VA. TAX REV. 671 (2006) (arguing that tax policies should reflect Judeo-Christian values); *see also* Adam S. Chodorow, *Biblical Tax Systems and the Case for Progressive Taxation*, 23 J.L. & RELIGION 1 (2008) (discussing the propriety of looking to Biblical tax systems for guidance on modern tax policy). My point here is not to defend the ability-to-pay principle in this context, and I concede it may have limitations that are sometimes overlooked. *See generally* Stephen Utz, *Ability To Pay*, 23 WHITTIER L. REV. 867 *passim* (2002) (discussing the shortcomings of the ability-to-pay approach). Instead, my goal is to demonstrate how this principle provides guidance regarding the proper treatment of virtual income.

269. *See* EDWIN R. A. SELIGMAN, THE INCOME TAX: A STUDY OF HISTORY, THEORY, AND PRACTICE OF INCOME TAXATION AT HOME AND ABROAD 4 (1911) (stating "[t]he history of finance, in other words, shows the evolution of the principle of faculty or ability to pay—the principle that each individual should be held to help the state in proportion to his ability to help himself").

270. *See* WILLIAM A. KLEIN, POLICY ANALYSIS OF THE FEDERAL INCOME TAX 4 (1976) ("The appeal of the head tax diminishes rapidly, however . . . as we begin to take account of the fact that people will . . . differ in their capacities to contribute.").

271. This conclusion is somewhat overstated, as a family with high net worth but low income could pay the tax. However, the point remains that head taxes are not correlated with ability to pay.

272. Camp focuses on the administrative difficulties associated with taxing imputed income. Camp, *supra* note 3, at 38.

273. Lederman, *supra* note 3, at 1644.

knew that a shave cost $10, we would not think that a person who shaved himself was in a better position to pay his taxes than the person who failed to shave, unless we knew for a fact that the person would have actually spent $10 on a shave and now has that money in the bank.[274]

Even where someone receives something of value that can be converted to cash and used to pay taxes, concerns regarding ability to pay may arise. For instance, one of the justifications offered for retaining the realization requirement is that taxing unrealized appreciation may lead to taxpayer liquidity issues and difficulties in paying taxes due on such appreciation.[275] Liquidity concerns are also one of the justifications offered for not taxing the receipt of gifts and inheritances.[276] Thus, even though property appreciation and the receipt of gifts or an inheritance clearly constitute measurable accessions to wealth, and are therefore income under that formulation, we nonetheless exclude them from the tax base.[277] Liquidity concerns are simply a weak form of the concern regarding a taxpayer's ability to pay.

A number of other provisions found in the tax code reflect concerns regarding ability to pay. For instance, the deferral provided in I.R.C. § 83 for property subject to restrictions reflects a judgment that taxes should not be owed on compensation until the taxpayer has the ability to use that income to pay taxes.[278] Similarly, I.R.C. § 109 excludes from income improvements a lessor

---

274. To take this analysis a step further, the ability to shave oneself could be seen as having value, and arguably tax should be owed based on ability or endowment. However, endowment taxation raises the same ability to pay problems as taxation of imputed income. For a list of articles discussing endowment taxation, see David Hasen, *Liberalism and Ability Taxation*, 85 TEX. L. REV. 1057, 1059 n.1 (2007).

275. *See* Deborah H. Schenk, *A Positive Account of the Realization Rule*, 57 TAX L. REV. 355, 359 (2004). Most scholars acknowledge that the realization requirement is now a matter of administrative convenience and that Congress is free to dispense with it at will. However, at least one scholar contends that the realization requirement remains a constitutional one. *See* Henry Ordower, *Revisiting Realization: Accretion Taxation, the Constitution, Macomber, and Mark to Market*, 13 VA. TAX REV. 1, 99 (1993) (arguing that realization remains a constitutional prerequisite for the taxation of gains from property).

276. *See* Joseph M. Dodge, *A Deemed Realization Approach is Superior to Carryover Basis (and Avoids Most of the Problems of the Estate and Gift Tax)*, 54 TAX L. REV. 421 *passim* (2001) (discussing the liquidity problems a deemed realization rule would create); *see also* Adam S. Chodorow, *Maaser Kesafim and the Development of Tax Law*, 8 FLA. TAX REV. 153, 171 (2007) (identifying liquidity as one of the concerns justifying the non-taxation of gifts). This justification, of course, does not apply to cash gifts or inheritances. In actuality, the receipt of an illiquid gift might well increase someone's ability to pay under certain circumstances. Imagine someone who was planning to buy himself a $15,000 car. Next, suppose that this person's grandfather gave him the car as a birthday gift. As a result of the gift, he would have $15,000 in his bank account that he would not otherwise have had. To take liquidity concerns seriously, one would need to look at a taxpayer's intent vis-à-vis the gift.

277. *See, e.g.*, I.R.C. § 2503(b) (2000).

278. *Id.* § 83.

makes on a lessee's property that revert to the lessor when the lease terminates.[279] Such improvements are not severable from the property, and including them in income could create a liquidity problem, possibly forcing the lessor to sell the entire property to pay taxes on the improvements received.[280]

The installment sales rules found in I.R.C. § 453 also arguably reflect judgments regarding the tax base and ability to pay.[281] Taxpayers who sell their assets for a series of deferred payments have engaged in a realization event and under the normal tax laws should be required to pay tax on their gains in the year of the sale.[282] However, as they have not actually received the money for the sale, they might not have the ability to pay the tax in the year of the sale. Section 453 permits them to report their gains over time as they receive payments.[283]

In looking to the ability-to-pay principle for guidance on the proper treatment of virtual wealth, I recognize that it has its limits. In some cases, we include in income items that do not increase a person's material wealth. For instance, if an employer allows an employee to use a corporate apartment, we tax the employee on the value of use of that apartment, even though the employee has no additional money with which to pay the tax.[284] The same is true for any non-exempt fringe benefits or deferred compensation the employer provides.[285] Barter clubs provide another example. Members of a barter club may agree to earn points that they cannot redeem except within the confines of a barter club. Nonetheless, although the receipt of barter points arguably does not increase their ability to pay real-world taxes, the IRS has clearly held that the value of such points must be included in income.[286]

Congress also routinely includes items in income despite temporary liquidity concerns that in other cases appear to warrant deferral or non-taxation. For instance, a taxpayer who simply exchanges a clock for a broach is in no more of a position to pay taxes than the person who holds onto the clock. Nonetheless, except for the limited provisions of I.R.C. § 1031, we require parties to such an exchange to pay taxes on the gain inherent in the exchanged item, despite the fact that the taxpayer may not have the cash to pay the taxes.[287] The mark-to-market rules found in I.R.C. § 475 provide another example.[288] Under this provision,

279. *Id.* § 109.

280. I.R.C. § 1019 precludes the lessor from obtaining any bump in basis for the improvements, thereby preserving for a later date the recognition of income.

281. *Id.* § 453.

282. *See id.* § 61.

283. Section 453 is in the part of the code that deals with accounting and can be read as merely an accounting mechanism to be used when taxpayers receive deferred payments. *Id.* § 453. Indeed, it seems required when applied to a cash method taxpayer, in that taxes flow from cash receipts and not formal sales. *Id.* However, this section applies equally to accrual method taxpayers. *Id.* In this context, it more clearly reflects a concern about ability to pay.

284. *See, e.g.*, Dean v. Comm'r, 187 F.2d 1019, 1020 (3d Cir. 1951).

285. *See* I.R.C. § 61(a)(1).

286. Rev. Rul. 79-24, 1979-1 C.B. 60.

287. *See* I.R.C. § 1031.

288. *Id.* § 475.

securities dealers must mark their securities to market at the end of each year, including any unrealized gains or losses in income and resetting the basis of those assets.[289] Similarly, in the context of the estate tax, Congress appears to have no concern for liquidity, as all assets are included in the tax base, including value attributable to unrealized appreciation.[290]

In the absence of concerns regarding ability to pay, it might be tempting to conclude that all accessions to wealth should be included in the tax base. However, practical concerns regarding the administrative costs relative to the expected revenues, the risk of non-compliance, and the effects that may have on compliance in other areas, difficulties in valuation, and the level of government intrusion necessary to enforce the law can and sometimes do override what is the acknowledged, correct, theoretical result. The rules governing *de minimis* fringe benefits provide a good example of this phenomenon.[291] Thus, ability to pay functions as a guideline for policy makers, but it cannot determine the appropriate outcome standing alone.

In the examples above, the circumstances that seem to override the concern regarding the ability-to-pay principle include potential tax avoidance, equity between similarly situated taxpayers, or a conclusion that non-taxation would allow for too much deferral.[292] In contrast, administrative concerns appear to be foremost in cases where Congress decides not to impose tax, despite an increase in a taxpayer's ability to pay real-world taxes.

Both the concern regarding ability to pay and the considerations that generally override such concerns are at play with the income generated in virtual worlds. Clearly, virtual income that has real-world value must be considered an accession to wealth. Nonetheless, the receipt of virtual income may not increase one's ability to pay real-world taxes. However, even where it does, the revenue to be gained in light of the administrative costs may make taxation inappropriate.

Ability to pay in the virtual world context is a function of the permeability of the boundaries between reality and a given world and, in particular, the ability to exchange virtual property for real-world cash or property, i.e., the ability to cash out. Accordingly, the decision whether to include virtual income in the real-world tax base should, at least initially, key off of the nature and extent of his ability. This approach is similar in some regards to Camp's, in that it focuses on the boundaries between virtual worlds and the real world to judge whether taxation is appropriate.[293] However, Camp focuses on whether virtual reality displaces real-

289. With regard to I.R.C. § 475, Congress determined that inventories of securities are easily valued and that other methods understate income. H.R. REP. NO. 111, at 661 (1993). In addition, the market for most securities is highly liquid, reducing claims that liquidity problems would arise.

290. I.R.C. § 641.

291. *See* I.R.C. § 132(e) (2000); Treas. Reg. 1.132-6 (1992).

292. *See* Zelenak & McMahon, *supra* note 722, at 1304–05 (arguing that tax law explicitly deals with cash receipts, and the numerous rules to include property in the tax base are designed as a backstop to prevent cheating). *But see* Dodge, *supra* note 71, at 715–21 (arguing that the tax explicitly includes cash and property).

293. Camp, *supra* note 3, at 47–71.

world economic activity, citing the acceptance of virtual currencies as evidence that virtual reality was no longer pretend.[294] In contrast, I focus on the ability to cash out and its impact on a taxpayer's ability to pay his taxes.[295]

Worlds that preclude participants from cashing out should be considered closed, and, absent countervailing considerations, virtual receipts in such worlds should be excluded from income because they do not increase a participant's ability to pay real-world taxes. Because boundaries are permeable to different degrees, any rule based on permeability must grapple with the nature and significance of those differences. In contrast, worlds that permit participants to cash out should be considered open, and income earned in such worlds increases a participant's ability to pay real-world taxes and should be included in the tax base, again, unless some countervailing practical consideration exists. I discuss these issues below in Part C.

B. Doctrinal Analysis

If one determines that a virtual world is closed, and that participants cannot readily convert their virtual receipts into real-world wealth, no further doctrinal analysis is necessary. Such income should be excluded from the tax base. In contrast, if the world is deemed open such that taxation is theoretically appropriate, it is still necessary to consider whether current tax doctrines or policies would nonetheless exclude virtual income from the tax base. While it is tempting to try to sweep all virtual receipts into one category as Camp did, self-created assets, drops, sales and exchanges all present different issues. I have discussed the proper tax treatment of such receipts in the context of Lederman's and Camp's proposals, but I will summarize my conclusions here.

Under existing law, the value of real-world self-created items is excluded from income even though such items represent an accession to wealth. If Picasso were to create a new painting, he would have an undeniable accession to wealth. However, he would not be taxed until he sold it. Some would classify self-created assets as imputed income and excuse them from income as such.[296] Arguably, they are analogous to the "taken" objects Dodge describes and should be seen as an investment leading to income, but not income in-and-of-themselves.[297] Whatever the underlying theory, real-world self-created assets are not subject to tax. No reason exists to believe that virtual assets should be treated differently. Accordingly, if someone makes a virtual good, its value should be excluded from income under existing law until it is disposed of in a taxable transaction.

---

294. *Id.* at 69.

295. *C.f.* Seto *supra,* note 3, at 15–25 (arguing that convertability and redeemability are the key markers for determining the taxability of virtual income).

296. *See, e.g.,* Dodge, *supra* note 71, at 691–92 (defining one meaning of imputed income as "the market-equivalents of non-market economic activity (such as the value of self-grown crops and the rental value of self-owned assets, and possibly the value of self-performed services)").

297. *Id.* at 696 ("[t]here is no meaningful distinction between 'taken' and 'self-created' objects").

How current law treats drops is more difficult. As with self-created objects, drops reflect an accession to wealth. However, they could fall into a number of different existing tax categories, some of which require their inclusion in income and others of which excuse them. For instance, if they are considered prizes, drops will be covered by I.R.C. § 74 and included in income.[298] If they are considered treasure trove, they will be included in income under I.R.C. § 61 and Treas. Reg. § 1.61-14.[299] In contrast, if drops are considered a form of imputed income or as "taken" items, as Dodge describes that category, they should be excluded.[300] As described above in Part III.A.1.a, I believe that drops should be considered a form of prize and taxed as such because they are provided to a participant by the world's developer as a reward for completing some task. In this regard, they look more like compensation than "taken" objects or imputed income, neither of which involves third parties.

The final category is sales and exchanges. Lederman contends that the doctrinally correct answer depends on whether people are deemed to have property or use rights in their virtual goods and currency.[301] Camp avoids the issue entirely by asserting that all such income is imputed and excluded from the tax base on that ground.[302] Nonetheless, he questions whether the property/use right distinction matters because use rights can be seen as a species of property—a chose-in-action—and the exchange of such rights constitutes a realization event.[303] I concur with Camp's assessment of the nature of use rights, but reject his claim that any income generated in such exchanges should be exempt from tax as imputed income. Rather, sales and exchanges of virtual goods, whether people have property or use rights in such goods, should lead to income tax liability in the same way it does for real-world goods.

In sum, once one concedes that virtual income has value and determines that participants can readily cash out, nothing about the virtual or online nature of these transactions suggests a tax result different from what one would expect for real-world goods.[304] Acquiring property or Lindens in Second Life or PED in Entropia Universe is really no different from acquiring property or Euros in Europe, and the tax consequences should be the same.

---

298. I.R.C. § 74 (2000).

299. *Id.* § 61; Treas. Reg. 1.61-14 (1993).

300. Dodge, *supra* note 71, at 696.

301. Lederman, *supra* note 3, at 1665–66.

302. Camp, *supra* note 3, at 60–66.

303. *Id.* at 50–60.

304. Arguably, some special feature of virtual reality might warrant overriding doctrine, such as administrative concerns or legal issues that do not arise in the real-world context. One commentator explored whether people engaged in virtual reality might more easily and inadvertently fall under the tax partnership rules than actors in the real world—thus justifying exclusion of virtual income from the tax base—before ultimately concluding that this concern was unwarranted. *See* Miano, *supra* note 3, at 47–51. In addition, one could argue that policy objectives external to tax, such as a desire to nurture virtual reality in its infancy or preserve its nature as a play space, warrant non-taxation. *Id.*; Castronova, *supra* note 3. However, such arguments go beyond the basic question of whether the existing doctrines, standing alone, cover virtual income.

C. The Proposal

So, where does all of this policy and doctrinal analysis lead me? I propose that the IRS adopt two different tax regimes, depending on whether the world is deemed closed or open, as I define the terms below. I propose a cash out tax regime on closed worlds. In other words, like Camp and Lederman, I would exempt from tax in-world transactions in closed worlds. The policy justification for this proposal is that the receipt of virtual income does not sufficiently increase a recipient's ability to pay real-world taxes, such that inclusion in the tax based is appropriate. Accordingly, the type of in-world transaction involved—self-creation, drop, sale or exchange—is simply not relevant. At present, this would encompass most structured or game worlds,[305] but it could apply equally to unstructured worlds.

For open worlds, I propose that the IRS tax virtual wealth acquisition for any person or entity that earns over $600, as measured in U.S. currency, in any given year. I borrow this figure from the threshold figure the Internal Revenue Code establishes for certain information returns and convert it to a threshold for taxation.[306] Unlike floors, for instance that found in I.R.C. § 67, this threshold acts as a toggle, such that all income would become taxable once the threshold was met for a given world, not just the income in excess of the $600 amount.[307] For open worlds, the type of transaction matters. Just as in the real world, the value of self-created assets is excluded from income, so should self-created virtual assets be excluded from income. I would treat drops as prizes and therefore subject to tax. Similarly, I would include in income any gains on the sale or exchange of virtual assets, just as barter exchanges and sales of real-world items are taxed.

Finally, I propose that the IRS publish an annual list of closed and open worlds. This process would operate similarly to the IRS's oversight of non-profit organizations to which tax deductible donations can be made.[308] In this way, the IRS can make a world-by-world analysis and determine the proper treatment, based on its assessment of whether the world is truly a closed world, or whether it is simply another forum in which to conduct real world activity. To avoid *ex post facto* taxation, I propose that the list apply prospectively.

This proposal raises a number of issues. While some answers may be evident from the analysis and discussion above, I address them specifically below.

---

305. *But see* Entropia Universe, which is a structured/game world but which nonetheless permits participants to cash out.

306. *See, e.g.*, I.R.C. §§ 6050, 6060 (2000).

307. *Id.* § 67.

308. The IRS periodically issues public notices both announcing organizations that qualify as tax exempt and revoking the exempt status of organizations. *See, e.g.*, Notice 92-28, 1992-1 C.B. 515 (announcing the availability of favorable exemption letters); I.R.S. Priv. Ltr. Rul. 08-03-023 (January 18, 2008) (announcing the repeal of the exempt status of an organization previously classified as charitable).

1. Classifying Worlds

Identifying the ability to cash out as the critical factor in determining the taxability of virtual income does not end the inquiry; it is just the beginning. As described above in Part 2, virtual life is far messier than theory might suggest. Worlds do not fall neatly into two categories: closed and open. Instead, they fall along a continuum, especially as it relates to cashing out. At one extreme are Second Life and Entropia Universe, both of which facilitate the ability to cash out.[309] Other worlds may prohibit exchanges outside the context of the world, but do little to enforce the rules. Still others actively police such transactions and either confiscate property or close the accounts of those who break the rules.[310] One commentator estimated in 2004 that the secondary market for virtual goods was approximately $880 million.[311] While some of these sales may be for goods from worlds that permit external exchanges, it is almost certain that a substantial portion of these sales involve goods from worlds that prohibit such sales.[312] Even those worlds that are successful at policing their rules are not likely to be perfect, as there are always some willing to break the rules and clever enough to get away with it.

Unless one is willing to declare that all in-world transactions are subject to taxation, the policy question facing tax theorists is how much outward permeability to allow before subjecting in-world transactions to taxation.[313] One option is to look at a virtual world's formal rules (e.g., EULAs or TOSs) to determine whether in-world transactions should be taxed. If a world prohibits transactions outside the virtual space, then in-world transactions should arguably be exempt from tax. The difficulty with this approach is that it creates an incentive for worlds to prohibit such transactions formally, but, in the interests of their participants, encourages them to look the other way when they do occur. To allow non-taxation in such situations would elevate form over substance and permit virtual world developers to essentially "check the box" as to whether they

---

309. For a description of the LindeX, Second Life's exchange, see Miano, *supra* note 3, at 10. See Entropia Universe, *supra* note 31, for a description of Entropia Universe's debit card. See Zenke, *supra* note 31, for a discussion of Sony's experiment with "station exchange" on select Everquest servers.

310. Castronova, *supra* note 32, at 1.

311. Camp, *supra* note 3, at 12.

312. *See, e.g.*, Julian Dibbell, *Dragon Slayers or Tax Evaders?*, LEGAL AFF. 47 (2006), *available at* http://www.legalaffairs.org/issues/January-February-2006/feature_dibbell_janfeb06.msp.

313. Seto, *supra* note 3, at 4, 13–14, 16. While Seto identified the ability to cash out as the critical factor in determining the taxability of virtual income, he did not explore the question of what level of redeemability or convertibility would be sufficient to warrant taxation. *Id.* at 15–16. However, based on the assumption that it is not readily possible to cash out of World of Warcraft, he suggests that it should be considered a closed world. *Id.* This suggests that Seto would look to both formal and practical limits on the ability to cash out in determining whether a world was sufficiently closed so as to warrant not taxing in-world transactions.

intended in-world transactions to be tax free or taxable. It is not hard to imagine how that would play out.

Another option is to require worlds to police and actively enforce rules prohibiting the sale or exchange of assets outside the confines of the world.[314] This option raises the difficult question of how much enforcement would be necessary to satisfy the standard. The answer may well depend on one's perspective. For instance, Edward Castronova argues that worlds must police sites where such exchanges occur.[315] Indeed, insofar as extra-legal exchanges manifest themselves in-world as gratuitous transfers, he proposes that worlds might need to prohibit such exchanges to be considered "closed," and therefore immune from the tax laws.[316] However, Castronova is concerned with protecting the sanctity of play spaces and recognizes that *any* permeability in the boundaries will inexorably make a virtual space an extension of reality as the profit-seeking meme will necessarily crowd out the fantasy meme.[317] While banning gratuitous transfers makes sense in the effort to preclude real-world activities from intruding into a play space, it is not clear that tax theory requires such draconian measures.

Even for worlds that enforce their rules, gray markets will likely arise, creating a *de facto* ability to cash out. Imposing tax liability on people who follow the rules and leave their income in-world seems wrong. Moreover, the ability to cash out could be somewhat limited if the gray market is small. For example, if one person in Uzbekistan buys a virtual sword from a given world for five thousand Uzbekistan soms (UZD $5,000),[318] it seems inappropriate to subject everyone involved in that world to tax on their virtual earnings, especially if they cannot access the market. However, if the market is large enough, sufficiently developed, and accessible, the argument that the world is functionally open strengthens. Basing taxation on the existence of a grey market would require a difficult inquiry into the extent of such market and the ability of taxpayers to access it.

Accordingly, I propose that the IRS look primarily to a world's rules regarding the ability to buy and sell virtual assets outside the confines of the virtual space and the extent to which developers enforce those rules in determining whether a world should be considered closed for tax purposes. I

---

314. Castronova, *supra* note 32, at 1.

315. Castronova, *supra* note 3, at 196.

316. *Id.* at 207. While at first blush, this seems like a workable solution, it could easily be gamed through bargain sales. For instance, if someone buys a sword on eBay for $50, rather than having the seller simply give him the sword in-world, he could arrange to purchase the sword for one gold piece, far less than its in-world value. Unless the virtual world developer is willing and able to monitor the prices at which sales and exchanges are made, it would be hard to stop such evasive techniques. Such rules would also run the risk of making the experience less enjoyable, as the ability to haggle and cope with scarcity is precisely what makes these worlds interesting.

317. *Id.* at 208–09.

318. The conversion rate between U.S. dollars and Uzbekistan som is about $1 (USD)=$1,332 (UZD) Fin. Mgmt. Svc., U.S. Treas. Dept., Treas. Reporting Rates of Exch. (Sept. 30, 2008), http://fms.treas.gov/intn.html. Thus, $500 UZD amounts to about $3.

would leave the IRS with the discretion to promulgate rules, after suitable comments from interested parties, regarding what efforts are required. This process would operate similarly to the IRS's oversight of non-profit organizations to which tax deductible donations can be made. The IRS would classify each world as either open or closed. To avoid *ex post facto* taxation, I propose that the decision apply prospectively. New worlds could apply to have their status determined before they go live. Failure to do so would merely mean uncertainty for participants until the IRS issued a ruling.

As noted above, the ability-to-pay principle is not absolute, and one must look to countervailing considerations, the chief among which is the likelihood of tax evasion.[319] Taxpayers cannot simply contract their way out of taxation, as was shown in the 1970s with barter clubs.[320] At current writing, most worlds that restrict cashing out and enforce these rules are structured worlds, designed for leisure.[321] While the leisure nature of an activity is irrelevant for purposes of determining whether to include any income it produces in the tax base, it suggests nonetheless the absence of a tax avoidance motive.[322] Moreover, the amounts at issue in such worlds are sufficiently small and the administrative costs of taxation high. Thus, at present, the nature of these worlds and the activities that occur within them do not warrant overturning the initial conclusion reached by looking at whether virtual income increases one's ability to pay real-world taxes.[323] However, as virtual worlds evolve and the activities engaged in expand beyond leisure, the case for non-taxation, even in worlds that prohibit cashing out, may weaken. I would expect that the IRS would consider these issues, in addition to the ability to cash out, when deciding whether to classify a world closed or open.

2. *De Minimis* Rules for Open Worlds

The proposed $600 threshold for the open world regime appears to violate the ability-to-pay principle. For example, it shields people from taxation who earn less than $600 in a given virtual world in any give year, even though their virtual income clearly increases their ability to pay real-world taxes. As noted above, the ability-to-pay principle acts as a guide to the theoretically correct answer, but it does not and should not control the real-world results. Several practical considerations justify a deviation in this case, many of which Lederman and Camp raised in their arguments against taxing virtual income.[324]

The first relates to the difficulty of valuing virtual goods.[325] While it is certainly possible to determine the value of virtual goods, as noted by both Lederman and Camp, the market for such goods is much thinner than the market

319. Camp, *supra* note 3, at 32.
320. *Id.* at 33; Rev. Rul. 79-24, 1979-1 C.B. 61.
321. Camp, *supra* note 3, at 26.
322. *Id.* at 27.
323. *Id.*
324. *See id.* at 32; Lederman, *supra* note 3, at 1669.
325. Lederman, *supra* note 3, at 1664.

for real goods.[326] This problem is further highlighted whenever people engage in barter. The difficulty of valuing items in a barter exchange exists for real-world objects and services; it is compounded for virtual items that are not routinely bought and sold.[327]

The second relates to the cost of imposing a tax relative to the expected revenue. As a practical matter, imposing tax on the first $1 of virtual income may well impose significant burdens on taxpayers, while yielding minimal revenue. Many participants will owe no tax because their monthly participation fees will exceed the income generated. Under such circumstances, some level of exemption is appropriate. In some regards, an exemption for some amount of virtual income would function like the standard deduction found in I.R.C. § 63, which alleviates the need for taxpayers with only a few deductions to keep track and report those deductions each year.[328]

The third relates to the likelihood and effect of taxpayer non-compliance. As Camp notes, most normal people are unaware of the tax consequences of barter exchanges.[329] Those who are aware of the rules may not deem it worth their while to figure out the extent of their virtual income and simply fail to report such income. While the cost to the treasury may be small, the act of non-compliance might well lead to a lessened respect for the law and an increase in non-compliance in other, more important areas.[330] Non-compliance could be reduced through the use of detailed reporting requirements,[331] say through a Form 1099-V (for Virtual). Given the electronic nature of the transactions, tracking such income may be easier than it appears, though it would certainly impose costs on developers.[332] Currently, no such requirements exist, and world developers are likely to oppose them if proposed.[333]

The final reason relates to the disquiet most feel about taxing those engaged in purely leisure activities.[334] The leisure/work distinction is not one observed when deciding what to include in income.[335] Nonetheless, it seems wrong to tax

---

326. *Id.* at 1660, 1670; Camp, *supra* note 3, at 32.

327. Camp, *supra* note 3, at 32. One creative solution to this problem is to use the foreign currency rules found in Subchapter J. Under this system, one tallies one's starting and ending net worth in the virtual currency and then converts the difference into dollars using a weighted average exchange rate. *See* Chung, *supra* note 3, at 88–91. This approach still requires valuation, but it obviates the need to do so for each transaction.

328. *See* I.R.C. § 63(c) (2000). For taxpayers who may have deductions near the threshold level, this benefit may be apparent only, insofar as they will need to track their expenditures to see whether they should itemize. However, a significant number of taxpayers do benefit from this administrative compromise.

329. Camp, *supra* note 3, at 32.

330. *Id.* at 32–33; Lederman, *supra* note 3, at 1660.

331. Lederman, *supra* note 3, at 1669.

332. *Id.* at 1661.

333. *Id.* (quoting the CEO of Iron Realms, an online game publisher, who describes such a possibility as an "apocalypse").

334. Camp, *supra* note 3, at 60–61.

335. *Id.* at 60.

people who are simply playing in a virtual world and who may never convert their virtual wealth into real-world wealth.[336] Creating an admittedly arbitrary threshold, below which no tax should be due, has the virtue of preserving theory while accommodating the underlying intuition that taxation in this instance would be difficult and unpopular.[337]

The question next arises as to how to choose a threshold. It should be high enough to provide the benefits sought, but not so high as to negate completely the theoretically correct result. The Internal Revenue Code contains a number of different thresholds that bear on the inquiry. In particular, the information reporting rules speak to the question of how best to balance administrative costs and the possibility of non-compliance. This regime suggests that virtual income be exempted from the real-world tax base if it is below $600 in any given year.[338]

The federal income tax system relies on individuals to self-report their income. However, the government's faith in the people goes only so far. In addition to conducting audits, the government requires a wide variety of people and entities to file information returns that indicate who received payments from them and how much they received.[339] The government then matches this information against the tax returns it receives to ensure that people properly report their income.

The reporting requirements can be broken down into four main categories. The first requires reporting of all amounts. For instance, employers must report all salaries and brokers must report all proceeds from sales and exchanges, regardless of the amounts in question.[340] The second requires reporting if amounts paid out to any individual or entity aggregate over $10; for example, the payments of dividends and interest over $10 must be reported.[341]

The third reporting requirement covers amounts over $600. This threshold applies to a wide range of income events, including cancellation of debt

---

336. *Id.*; Lederman, *supra* note 3, at 1659.

337. Lederman, *supra* note 3, at 1670.

338. The withholding rules provide another point of reference when trying to establish a threshold above which virtual gains should be subject to tax. The modern withholding rules were first introduced into the Internal Revenue Code in 1943. Richard L. Doernber, *The Case Against Withholding*, 61 TEX. L. REV. 595, 601 (1982). They are designed to ensure that the government receives a steady flow of revenue throughout the year and to ensure that people do not spend all of their income as they earn it, thereby having nothing left with which to pay their taxes. *Id.* at 601–02. It has the salutary effect of requiring reporting, which aids in ensuring that people properly report their income. *Id.* at 595–96.

339. The statutory provisions requiring information returns are generally found in I.R.C. §§ 6031–6059 (2000).

340. *See id.* §§ 6045, 6051. The code requires a number of other returns regardless of income at issue, including from partnerships, *id.* § 6031; from trusts to beneficiaries, *id.* § 6034A; from S corporations, *id.* § 6037; by certain fishing boat operators, *id.* §6050A; and from those who pay certain railroad retirement benefits, *id.* § 6050G.

341. *See id.* §§ 6042, 6049. Other examples of returns due on amounts over $10 include payments of unemployment compensation, *id.* § 6050B, refunds of state or local taxes, *id.* §6050E, and royalties, *id.* § 6050N.

income,[342] certain real estate transactions,[343] the receipt of student loan interest,[344] the receipt of mortgage interest,[345] the payment of tuition and related costs of higher education,[346] and payments in the ordinary course of a trade or business, unless covered by some other provision.[347] The reporting threshold is also set at $600 for prizes won on game shows,[348] payments to physicians from certain types of organizations,[349] the purchase of fish,[350] and crop insurance proceeds.[351] Finally, the baseline threshold for reporting gambling winnings is $600.[352]

The fourth reporting requirement covers amounts over $10,000. For instance, those who receive payments of $10,000 or more in a trade or business must file returns reflecting those payments.[353] Financial institutions must report any currency transactions in excess of $10,000.[354] Similarly, those with a financial interest in a foreign account worth more than $10,000 must inform the IRS of such interest.[355]

The decision to require information reporting is not without cost. For those subject to the requirements, it is expensive to track payments, print, and submit such returns. Similarly, the government must then devote resources to creating the rules, enforcing them, and effectively dealing with the returns they receive.[356] However, there may be benefits to the returns beyond tax revenues. For instance, the existence of such returns may increase confidence in the system based on a sense that other people are paying their fair share.[357] Several studies have suggested that this sense may contribute to voluntary compliance rates.[358] Thus, the levels at which the reporting requirements kick in reflect a measured judgment regarding the cost of reporting, the amount of slippage allowable (assuming the

342. *Id.* § 6060P.

343. Treas. Reg. § 1.6045-4 (2000).

344. I.R.C. § 6050S.

345. *Id.* § 6050H.

346. *Id.* § 6050S.

347. *Id.* § 6041.

348. Treas. Reg. § 1.6041-1(d)(3) (2006).

349. *Id.* § 1.6041-1(d)(2).

350. I.R.C. § 6050R.

351. Treas. Reg. § 1.61-4(c) (1997).

352. This requirement appears to flow from I.R.C. § 6041, which does not explicitly mention gambling. Gambling winnings are reported on a Form W-2G. *See* I.R.S. 2007 Form W-2G (Certain Gambling Winnings).

353. I.R.C. § 6050I.

354. 31 C.F.R. § 103.22 (2008).

355. *Id.* § 103.24 (2008); I.R.S. Form TD F 90-22.1.

356. *See* Camp, *supra* note 3, at 22–23.

357. *See, e.g.*, Leandra Lederman, *The Interplay Between Norms and Enforcement in Tax Compliance*, 64 OHIO ST. L. J. 1453, 1477 (2003) (considering the interplay of enforcement and tax compliance norms).

358. *Id.* at 1460–62.

non-reporting below certain thresholds will lead to increased tax evasion) and the external effects reporting may have on general tax compliance.

If the task is to choose a threshold under which no taxes on virtual wealth will be owed, the "any amount" standard reflected above is obviously inapplicable. The "in excess of $10" standard is not much better. First, this standard is limited primarily to interest and dividends.[359] Second, as a practical matter, insofar as the $10 limit applies primarily to financial institutions and brokers whose customers routinely surpass these limits, it seems unlikely that such institutions will omit returns for the few who do not.[360] Thus, even though the limit exists, it seems likely that many of these institutions will send returns even to those for whom no duty exists. In other words, the $10 threshold is likely very close in operation to the "any amount" threshold.

The $10,000 standard generally applies to large transfers related to the banking system and payments from businesses.[361] It is generally used to track potential illegal activity separate from tax evasion such as currency exchanges and money laundering. It applies to a small number of categories and therefore makes little sense as a threshold.

In contrast, the $600 reporting standard seems appropriate on a number of levels. First, unlike payments to which the $10 and $10,000 limits apply, the $600 limit applies to a wide range of activities.[362] This broad application suggests that it the $600 figure is not tied to any particular activity. Second, and perhaps more important, it applies to gambling, a source of income extremely unlikely to be reported absent information returns.[363] In other words, the possibility of income not reported in an information return being omitted from a taxpayers return, and the associated revenue loss, is quite high. The $600 figure thus arguably serves as a meaningful proxy for the appropriate balance between the administrative costs of reporting and tolerable non-compliance and revenue loss.[364]

The $600 figure also works in the context of virtual worlds, in that it is high enough to exempt those who dabble in virtual reality, but low enough to capture

---

359. *See supra* note 341 and accompanying text.

360. As evidence, I refer you to the 1099-Int I received a few years back from Wells Fargo on my checking account for $1.83 (on file with the author, at least until the statute of limitations on assessment has passed). See I.R.C. § 6501 (2000).

361. *See supra* notes 353–58 and accompanying text.

362. *See, e.g., supra* notes 342-347 and accompanying text.

363. *See supra* note 352 and accompanying text.

364. This of course leads to the interesting question of the $600 figure's origin. The $600 figure first appeared as amendments to §§ 25, 51, and 147 of the Internal Revenue Code of 1939. *See* Revenue Act of 1948, ch.168 §§ 201-02, 62 Stat. 112, 114 (1948). Those sections were predecessor sections of current sections 151, 6012, and 6041. *See* I.R.C. §§ 151, 6012, 6041 (2000). On the one hand, the threshold's longstanding existence suggests that Congress has endorsed $600 as an appropriate measure. On the other hand, it may reflect legislative laziness. If the figure were adjusted for inflation, it would be $9, 443.87 in 2009 dollars. *See* Inflation Calculator Determines Change in Dollar and Rates Over Time, http://www.usinflationcalculator.com (last visited Nov. 9, 2008).

those who are making significant amounts. As Camp notes, the value of most goods is relatively low, meaning that most participants are unlikely to earn significant income in any given year.[365] While it seems likely that someone might generate $10 of income over the course of a year through his virtual activities, only those who spend considerable time in-world actively working to accumulate significant wealth are likely to reach $600. Linden Labs, the operator of Second Life, perhaps the most robust virtual world at present, reports that, of its approximately 62,000 unique users, approximately 51,000 have annual flows of Lindens in their accounts under $600.[366] Accordingly, setting the threshold at $600 is likely to provide the administrative benefits intended by the decision to adopt a threshold.

The next question is whether the threshold should apply to gross virtual income or net virtual income, and how net virtual income should be defined. I propose that the threshold apply to gross income, as that term is defined in I.R.C. § 61.[367] Such a rule is consistent with the notion that gambling winnings in excess of $600 are reported to the IRS, even though a taxpayer may have offsetting losses from other transactions. Assuming that threshold had been met, the person would be able to deduct the amounts allowed under the Internal Revenue Code from gross income. Thus, someone who earned $1,000 of virtual income could deduct the $300 of virtual income they spent to do so.

The more difficult question is whether they could also deduct the real-world costs they paid to participate in the virtual world. Under one view, that expense could be seen as permitting access, and therefore unrelated to the in-world activities required to earn virtual income.[368] In other words, the participant received the ability to play in return for his access fees. As value received is not included in income, it would amount to a double deduction if one were to allow a deduction for the expense. To make everything balance out, both the expense and the value of the right of access should be ignored. Thus, the full value of the virtual income should be considered an accession to wealth, and no deduction for the fees should be allowed. Under another view, the access fee can be considered a necessary and therefore deductible cost of earning the income. By analogy to the lottery, I would permit taxpayers to deduct their subscription fees against their virtual income, thereby avoiding any claim that their consumption was being taxed twice.

Finally, accepting, *arguendo*, Lederman's claim that virtual income should be considered consumption value and not taxed so long as there is no profit motive, using a threshold may actually be a better way to achieve that aim.[369]

---

365. Camp, *supra* note 3, at 38.

366. *See* Second Life Economic Statistics, Nov. 7, 2008, http://secondlife.com/whatis/economy_stats.php. Linden reports figures for six months. I have taken the liberty of annualizing them. As the inflows may reflect receipts that are not properly includible in income, the actual number of people with annual virtual income under $600 may be higher.

367. I.R.C. § 61.

368. *See supra* note 152.

369. Lederman, *supra* note 3, at 1659.

Assigning motives based on the type of world and whether people engage in barter transactions or sales is too crude a method to distinguish between those seeking pleasure and those seeking profit. Moreover, the dynamic response in unstructured worlds would make a mockery of any effort to impose disparate tax treatment for trades and sales. In contrast, the $600 threshold for open worlds avoids the pretense of accurately assessing motive or the nature of the activity as either consumption or profit seeking. Moreover, it is not susceptible to manipulation, except to the extent people can shift virtual income from one year to the next to stay under the $600 threshold. Thus, it may be a better measure than the one Lederman has chosen to separate taxable from tax-free transactions.

V. CONCLUSION

Whether to tax virtual income that is left in-world raises a vexing question. Most people's intuition suggests that such income should be exempt from tax, but the fact that it has real-world economic value suggests otherwise. Those who have considered the question to date have attempted to classify virtual income as outside the theoretical definition of income or as falling into a recognized category that is nonetheless exempt from tax. As described above, neither the rationales nor the proposals to date are fully satisfactory. In particular, they do violence to existing doctrine by implicitly introducing intent into the income-inclusion equation or stretching the meaning of imputed income beyond its limits. In addition, the proposals are difficult to administer or are subject to abuse by taxpayers.

In the past year, Chinese, Swedish, and South Korean tax authorities have all indicated an intent to tax virtual income left in-world.[370] Given the current budget crisis, the U.S. may not be far behind. Using the ability-to-pay principle as an initial filter achieves results consistent with intuition, while conforming to existing tax policy and doctrine. The approach produces a proposal that is more administrable than those made to date. It also allows for changes in the nature of virtual worlds, which may affect our judgment regarding the appropriate tax result. In the end, the decision whether to include virtual income in the tax base rests on equitable and administrative concerns that require evaluation. Using ability to pay as the guiding principle in making these determinations provides the IRS with sufficient flexibility to exercise such judgments, without the need to contort or stretch existing law or policy.

370. *See* Graham, *supra* note 15.

JUDICIAL (IN)DISCRETION: HOW COURTS CIRCUMVENT THE CONFRONTATION CLAUSE UNDER *CRAWFORD* AND *DAVIS*

MICHAEL D. CICCHINI *

I. INTRODUCTION

When the State attempts to prove a criminal allegation by using hearsay evidence, rather than calling the accuser to the witness stand for live testimony, the defendant's Sixth Amendment right of confrontation is implicated.[1] The right of confrontation guarantees that the accused "be confronted with the witnesses against him."[2] Historically, courts simply dispensed with the defendant's right to confront his or her accuser, provided the court first made at least a perfunctory finding that the hearsay was reliable.[3] Once the court made that finding, the theory was that actual confrontation of the witness by the defense would add little if any value to the truth-seeking process.[4] Therefore, the prosecutor was permitted to present untested, uncross-examined hearsay evidence to the jury.[5]

In the 2004 *Crawford v. Washington* decision, however, the United States Supreme Court finally addressed the dangers and unconstitutionality of allowing judges, rather than juries, to determine the reliability of hearsay evidence.[6] The Court noted that "[r]eliability is an amorphous, if not entirely subjective, concept."[7] More significantly, "judges, like other government officers, [can]not always be trusted to safeguard the rights of the people."[8] Therefore, "[a]dmitting statements deemed reliable by a judge is fundamentally at odds with the right of confrontation."[9]

Applying this new framework derived from *Crawford*, the Supreme Court determined in *Davis v. Washington* that a prosecution's use of testimonial

---

* J.D., *summa cum laude*, Marquette University (1999); C.P.A., Illinois Board of Examiners (1997); M.B.A., Marquette University (1994); B.S., University of Wisconsin—Parkside (1990). I dedicate this Article to the memory of my mother, Clare Cicchini, for her unconditional support of all my law-related endeavors. I thank my father, David Cicchini, for his continued support. I thank my friend, Amy Kushner, Ph.D., for her valuable comments on all of my legal scholarship, including this Article.

1. U.S. CONST. amend. VI.
2. *Id.*
3. Ohio v. Roberts, 448 U.S. 56, 66 (1980).
4. *See id.* at 64 (citing Chambers v. Mississippi, 410 U.S. 284, 295 (1973)).
5. *See id.* at 66.
6. Crawford v. Washington, 541 U.S. 36, 61-65 (2004).
7. *Id.* at 63.
8. *Id.* at 67.
9. *Id.* at 61.

hearsay, as opposed to nontestimonial hearsay, is what triggers the protections of the Clause.[10] Testimonial hearsay is defined, in part, as statements made during the course of police interrogations where the "primary purpose of the interrogation [was] to establish or prove past events potentially relevant to later criminal prosecution."[11]

Conversely, the Court's new framework offers no protection at all against nontestimonial hearsay. Nontestimonial hearsay includes statements made during the course of police interrogations where the "primary purpose" of the interrogation was not to investigate a past crime, but rather to gather information that would enable police to offer assistance, or respond to an "ongoing emergency."[12] Consequently, in the span of two cases – *Crawford* and *Davis* – the Court appeared to set a new course in Confrontation Clause jurisprudence: one that required the reliability of "testimonial hearsay" to be determined by juries, not by judges, and only after "testing in the crucible of cross-examination."[13]

Lower court decisions in the wake of *Crawford* and *Davis*, however, have sought to circumvent rather than apply the Court's new framework. For example, courts have grossly distorted key terminology such as "ongoing emergency" and "primary purpose," and in so doing have been able to classify hearsay as nontestimonial, thereby placing it outside the protections of the Clause altogether.[14] In other cases, courts have dramatically expanded the forfeiture doctrine, a critical exception to the right of confrontation, and in so doing have been able to dispense with cross-examination even where testimonial hearsay is at issue.[15] As a result, a defendant's right of confrontation remains as weak, malleable, and subject to judicial manipulation as it was before *Crawford* and *Davis*.

Lower courts have been successful in circumventing the Court's new framework because *Crawford* and *Davis* never actually constrained judicial discretion as the Court intended. Rather, the lower courts' discretion has merely shifted from one issue – whether the hearsay was reliable[16] – to other issues, such as whether the hearsay is testimonial[17] and, if so, whether the defendant forfeited his or her right of confrontation.[18] In the end, therefore, little has changed, and the problem of judicial manipulation remains.

Unfortunately, the most commonly proposed solutions to this problem continue to merely shift, rather than constrain, judicial discretion. This Article asserts that such approaches are constitutionally inadequate. It is true that many

10. Davis v. Washington, 547 U.S. 813, 822 (2006).
11. *Id.*
12. *Id.*
13. *Crawford*, 541 U.S. at 61.
14. *See* discussion *infra* Part IV.A.-B.
15. *See* discussion *infra* Part IV.C.
16. *See* discussion *infra* Part V.
17. *See* discussion *infra* Part VI.A.
18. *See* discussion *infra* Part V.

judges – perhaps even most judges – honestly attempt to uphold the Constitution. However, the cases discussed herein will show that many do not. The purpose of this Article, then, is to address this serious problem directly and openly.

The only viable solution to the problem of judicial manipulation is to *constrain* judicial discretion. In the context of defining testimonial hearsay, such constraint can only be achieved by focusing on how a statement is used at trial, not the manner in which the statement was gathered or obtained or the circumstances under which the statement was made.[19] Only a bright-line, trial-based rule will restore the right of confrontation, as the Court intended in *Crawford* and *Davis*.

II. CONFRONTATION: A BRIEF AND RECENT HISTORY

The Sixth Amendment's Confrontation Clause states rather simply and clearly that "[i]n all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against him."[20] Surprisingly, this simple and clear mandate has met with much resistance from the courts and has historically been the subject of much debate.[21] For purposes of this Article, we need only to recount the recent history of the Clause.

A. Ohio v. Roberts: Interpreting the Clause

From 1980 to 2004, the leading Supreme Court case interpreting the Confrontation Clause was *Ohio v. Roberts*.[22] In *Roberts*, the defendant was charged with possession of stolen property, among other crimes.[23] At his preliminary hearing, a witness had testified and denied giving the defendant permission to possess the property.[24] At trial, the witness was unavailable for live testimony, and the State had attempted to introduce the transcript of her preliminary hearing testimony; the defendant objected on confrontation grounds.[25]

---

19. *See* discussion *infra* Part VI.B.

20. U.S. CONST. amend. VI.

21. *See, e.g.*, Randolph N. Jonakait, *"Witnesses" in the Confrontation Clause:* Crawford v. Washington, *Noah Webster & Compulsory Process*, 79 TEMP. L. REV. 155, 197 (2006); Ellen Liang Yee, *Confronting the "Ongoing Emergency": A Pragmatic Approach to Hearsay Evidence in the Context of the Sixth Amendment*, 35 FLA. ST. U. L. REV. 729, 735-38 (2008); Fred O. Smith, Jr., Note, Crawford*'s Aftershock: Aligning the Regulation of Nontestimonial Hearsay with the History and Purposes of the Confrontation Clause*, 60 STAN. L. REV. 1497, 1501-07 (2008).

22. Ohio v. Roberts, 448 U.S. 56 (1980).

23. *Id.* at 58. According to the Court, "[The defendant] was charged with forgery of a check . . . and with possession of stolen credit cards." *Id.*

24. *Id.*

25. *Id.* at 59. There was much argument about whether the witness was actually "unavailable" for trial. *Id.* at 60-61. Although outside the scope of this Article, the concept of unavailability is crucial. If the witness is actually "available," the State must call the witness to

The Court acknowledged the importance of a defendant's right to confront a witness at trial, stating that cross-examination is crucial in "testing the recollection and sifting the conscience of the witness" and aiding the jury to decide "by his demeanor upon the stand and the manner in which he gives his testimony whether he is worthy of belief."[26] However, the Court also stated that a literal reading of the Clause "would require, on objection, the exclusion of any statement made by a declarant not present at trial."[27] The Court rejected this literal approach because it believed that a variety of other competing interests "may warrant dispensing with confrontation at trial."[28]

The *Roberts* Court therefore held that once a witness is shown to be unavailable for live testimony at a criminal trial, the witness's prior hearsay statement may be used by the State, and the defendant's right of confrontation may be dispensed with, if the proffered statement bears "indicia of reliability."[29] If the trial judge determined that the hearsay is reliable, then cross-examination of the declarant on the stand would serve little purpose, and confrontation must give way to competing interests.

Under the *Roberts* analysis, trial courts could make this finding of reliability in two ways: (1) if the court found that the hearsay fell within a "firmly rooted hearsay exception," it would be deemed reliable *per se*, and therefore admissible;[30] or (2) if the hearsay did not fall within such an exception, the trial court could still find that the prior statement carried "particularized guarantees of trustworthiness" and was therefore reliable and admissible.[31] When using this second, disjunctive prong of the reliability analysis, courts had wide latitude to consider virtually any imaginable fact or circumstance that surrounded both the declarant and the hearsay itself.[32]

At the preliminary hearing in *Roberts*, defense counsel had the opportunity for cross-examination, which "was not 'significantly limited in any way in [its]

---

testify at trial, regardless of whether the defendant had a prior opportunity for cross-examination. California v. Green, 399 U.S. 149, 182-83 (1970).

26. *Id.* at 64 (citing Mattox v. United States, 156 U.S. 237, 242-43 (1895)).

27. *Id.* at 63.

28. *Id.* at 64. Interestingly, the Court once believed that each jurisdiction considered "the development and precise formulation of the rules of evidence applicable in criminal proceedings" as one such competing interest that warranted dispensing with the right of confrontation. *Id.* (citing Snyder v. Massachusetts, 291 U.S. 97, 107 (1934)). This position was reversed years later when the Court stated that "[w]here testimonial statements were involved, we do not think the Framers meant to leave the Sixth Amendment's protection to the vagaries of the rules of evidence . . . ." Crawford v. Washington, 541 U.S. 36, 61 (2004).

29. *Roberts*, 448 U.S. at 66.

30. *Id.*

31. *Id.* This Article does not address the historical underpinnings of the various hearsay exceptions. Consequently, the search for particularized guarantees of trustworthiness will simply be referred to, more generally, as the "reliability test" or "reliability analysis."

32. *See* discussion *infra* Part II.B. The past tense verb "had" is used because the *Roberts* analysis was overturned in *Crawford v. Washington*, 541 U.S. 36 (2004) and *Davis v. Washington*, 547 U.S. 813, 822 (2006).

scope or nature . . . .'"[33] Defense counsel asked leading questions and even had some success in exposing the declarant's "ulterior personal reasons for unfairly casting blame."[34] For these reasons, the hearsay was deemed reliable by the Court and was therefore admitted into evidence.[35]

At first glance, the *Roberts* Court's reliability test seemed promising and, at least under the specific facts of that case, seemed to result in a holding that was immune from any serious criticism. The application of this test in the 26 years to come, however, would showcase the perils of allowing judges, rather than juries, to determine the reliability of evidence.

B. Judicial Discretion in Action

The *Roberts* reliability test would prove, in hindsight, to be quite ill-conceived. Applying the test to the particular hearsay in *Roberts* – the prior, sworn, and fully cross-examined preliminary hearing testimony – was easy and straightforward. The facts presented in *Roberts* left little room for any real judicial discretion or meaningful debate.

Applying the reliability test to the facts of other cases, however, produced highly inconsistent, irreconcilable, and even bizarre results. Judges were able to manipulate facts to reach their desired outcomes, largely because of the tremendous amount of judicial discretion inherent in the *Roberts* reliability test. Thus, two phenomena—highly inconsistent and irreconcilable results without manipulative intent, and outright judicial manipulation—are evident to varying degrees when comparing relevant court decisions across states, within a single state, and even within a single court.[36]

1. Inter-State Analysis

A citizen's constitutional rights should not vary depending on the state in which that person was accused of a crime.[37] Nonetheless, mere random variation is the most innocent of the possible explanations for discrepancies across

---

33. *Roberts*, 448 U.S. at 71 (citing California v. Green, 399 U.S. 149, 166 (1970)).

34. *Id.*

35. *Id.* at 73. Furthermore, "[i]n *Green* the Court found guarantees of trustworthiness in the accouterments [sic] of the preliminary hearing itself." *Id.* As in *Green*, the *Roberts* court found that the defendant had "an adequate opportunity to cross-examine [the witness]," leaving no reason to hold that the hearsay testimony in *Roberts* was unreliable. *Id.* (quoting Mancusi v. Stubbs, 408 U.S. 204, 216 (1972)).

36. The state court cases discussed in the following sections were cited by the Court in *Crawford*, when criticizing the *Roberts* reliability test. *Crawford*, 541 U.S. at 63. However, the cases are discussed in greater detail in this Article than in *Crawford*; therefore, citations herein are made directly to the state court cases themselves. *See* notes and discussion *infra* Parts II.B-VI.B.

37. This statement incorporates the presumption of innocence. For those less inclined to respect the presumption, a modified version of the statement still holds true: a citizen's constitutional rights should not vary depending on the state in which he or she has *committed* a crime.

jurisdictions. A less innocent (and perhaps more likely) explanation is that judges may have attached any meaning they wished to any given factor to reach their desired outcome: the admission of untested, uncross-examined hearsay evidence against defendants.

For example, in a Virginia case *Nowlin v. Commonwealth*, the defendant was charged with illegally possessing a firearm.[38] At trial, the State introduced the hearsay statement of the defendant's wife who was unavailable for live testimony.[39] She had previously stated to police, "[h]is bedroom door was locked. He keeps his bedroom door locked because we've got guns in there."[40] This statement was offered to prove that the defendant *knowingly* possessed a firearm, which was an essential element of the charge.[41]

After doing a facts-and-circumstances reliability analysis, the court found that the wife's hearsay statement to the police was reliable, largely because the wife made the statement after she had been arrested, taken into custody, and interrogated regarding her *own* criminal activity.[42] Ironically, the wife was being investigated for shooting at her husband with the very firearm that he was accused of unlawfully possessing.[43] As justification for its finding, the court reiterated, "[a]gain, we note that [the defendant's] wife made the statement while she was *in police custody, charged with shooting at [the defendant]*."[44]

In contrast, in the Wisconsin case *State v. Bintz*, the defendant was charged with homicide for allegedly killing a bartender at a tavern.[45] At trial, the State introduced the hearsay statement of the defendant's brother, who was unavailable for live testimony.[46] The brother previously said to the police, among other things, that he and the defendant were at the tavern the night of the murder.[47] Of course, this statement was offered to place the defendant at the scene of the crime.[48]

After its facts-and-circumstances reliability analysis, the court found that the brother's hearsay statement to police was reliable because the brother had *not* been arrested, was *not* in custody, and was *not* a suspect in the murder or in any other crime when he made the statement.[49] In making its reliability finding, the court stated that "there is no evidence [the brother] was told he was a suspect.

---

38. Nowlin v. Commonwealth, 579 S.E.2d 367, 369 (Va. Ct. App. 2003).

39. *Id.* at 370.

40. *Id.*

41. *See id.*

42. *Id.* at 372. The statement by the defendant's wife was deemed to be reliable because the investigating officer "advised [the defendant's] wife of her rights to remain silent and that what she said might be used against her." *Id.*

43. *Id.* at 371.

44. *Id.* at 372 (emphasis added).

45. State v. Bintz, 650 N.W.2d 913, 915 (Wis. Ct. App. 2002).

46. *Id.* at 916. The declarant had invoked his Fifth Amendment privilege, thus the court found him unavailable for trial. *Id.*

47. *Id.* at 915.

48. *See id.* at 917.

49. *Id.*

[He] was *not in custody* and there is *no indication he was threatened with prosecution* or asked leading questions."[50]

These decisions illustrate an irreconcilable contradiction. In Virginia, when a declarant is *in custody* and *accused* of a crime, any statement he or she made is likely to be found reliable and therefore admissible against a defendant, even if the declarant became unavailable for trial and could not be cross-examined.[51] Conversely the opposite was true in Wisconsin. When a declarant was *not* in custody and *not* accused of a crime, any statement he or she made was likely be found reliable and therefore admissible against a defendant, even if the declarant became unavailable for trial and could not be cross-examined.[52]

Further, this discrepancy across jurisdictions was not an aberration. For example, a Colorado court determined that a hearsay statement was reliable and therefore properly admitted because it was detailed.[53] The declarant "provided detailed descriptions of the events and conversations that occurred, the surroundings at each stage of the criminal episode, and the actions attributable to each party."[54] Conversely, the Fourth Circuit Court of Appeals determined that a hearsay statement was reliable and therefore properly admitted in a Virginia case because the declarant's statement contained little detail.[55] The court was particularly impressed that the statement, in its relevant part, "was fleeting at best."[56]

Variations between states may have been attributable to causes other than judicial manipulation. These causes may include, for example, differences in the wording of two states' facts-and-circumstances reliability tests. Of course, given that the right of confrontation is guaranteed by the United States Constitution, even this innocent explanation is simply unacceptable. A citizen accused of a crime should not have his or her freedom hinge merely on the state in which the citizen has the misfortune of being charged.

However, an alternative explanation is that courts were willing and able to manipulate any set of facts and circumstances, no matter how diametrically opposed, to reach the same, predetermined outcome. If this explanation were the case, then the right of confrontation had effectively been eliminated. An analysis of cases within a single state, rather than across states, will test this hypothesis.

2. Intra-State Analysis

Further analysis of the *Roberts*' reliability test also shows tremendous variations not only between states, but also *within* any given state. These

50. *Id.* at 918 (emphasis added).

51. Nowlin v. Commonwealth, 579 S.E.2d 367, 372 (Va. Ct. App. 2003).

52. *Bintz*, 650 N.W.2d at 918.

53. People v. Farrell, 34 P.3d 401, 407 (Colo. 2001).

54. *Id.*

55. *See* United States v. Photogrammetric Data Servs., Inc., 259 F.3d 229, 245 (4th Cir. 2001).

56. *Id.*

variations are most easily exposed when analyzing a single case as it proceeds vertically in the hierarchy of a state court system.

In a Washington case *State v. Crawford*, the defendant was charged criminally for stabbing a man, but maintained that he did so in self-defense.[57] At trial, the State introduced the hearsay statement of the defendant's wife, a witness who was unavailable for live testimony.[58] The wife, who was a suspected accomplice to the crime, had been arrested and interrogated by police.[59] During the interrogation, she described in great detail how the defendant stabbed the alleged victim.[60]

After concluding a facts-and-circumstances reliability analysis, the trial court found that the wife's hearsay statement to police was reliable and therefore admissible.[61] The court found that the statement "was against her penal interest"[62] because her "admissions could give rise to accomplice liability."[63] Moreover, the absence of police coercion or offers of leniency coupled with the wife's "apparent motive" to help the defendant also led the court to conclude that the statement was reliable.[64]

The appellate court reversed the trial court, finding that much of the wife's statement was "*not* against her penal interest."[65] Further, the court found that the wife gave conflicting versions of her statement, that her statements were not spontaneous but were the product of structured police interrogation, and that she even admitted to closing her eyes during the stabbing.[66] All of these factors indicated that the statement was not reliable and therefore should not have been admitted against the defendant.[67]

Nevertheless, the Washington courts were not through with this case. The Washington Supreme Court analyzed the reliability of the wife's statements and reversed the appellate court because of two factors.[68] The court held that the wife's statements were reliable because they were self-inculpatory.[69] Moreover,

---

57. State v. Crawford, No. 25307-1-II, 2001 Wash. App. LEXIS 1723, at *1 (Wash. Ct. App. Jul. 30, 2001). This Washington Court of Appeals decision led to the Washington Supreme Court decision *State v. Crawford*, 54 P.3d 656 (Wash. 2002). Eventually, that decision led to the United States Supreme Court decision *Crawford v. Washington*, 541 U.S. 36 (2004), discussed at length in Parts III.A and V of this Article. All citations in this section, however, are to the state court decisions.

58. *Crawford*, 2001 Wash. App. LEXIS 1723, at *2. The defendant invoked the marital privilege to prevent his wife from testifying. *Id.*

59. *Id.*

60. *Id.*

61. *See id.* at *9.

62. *Id.* at *9.

63. *Id.*

64. *Id.* at *9-10.

65. *Id.* at *10 (emphasis in original).

66. *Id.* at *13-15.

67. *Id.* at *16.

68. *See* State v. Crawford, 54 P.3d 656, 664 (Wash. 2002).

69. *Id.* at 662.

the court held that, because the hearsay statement by the wife "was virtually identical" to the statement given by the defendant, the statements interlocked and thus the hearsay was reliable and had been properly admitted by the trial court.[70]

The issue of interlocking statements, however, had already been analyzed by the appellate court. In its analysis, the appellate court acknowledged many similarities between the statements, including what the husband and wife did and where they went *before* the alleged crime.[71] Yet the appellate court decided that the two statements were different with regard to the only critical issue in the case: whether the alleged victim was armed at the time he was stabbed.[72] As the Washington Supreme Court even conceded, "[s]elf-defense is at issue in this case, so admittedly the timing of [the] possession of a weapon is significant."[73] Nonetheless, the state supreme court overruled the thorough, well-reasoned decision of the appellate court and reinstated the conviction in an effort to uphold the admission of the hearsay.[74]

Just as with inter-state variations, these intra-state variations are far from aberrational.[75] Judicial flip-flopping from court to court, combined with disagreement within a given court, diminishes citizens' confidence in any judicial system. Constitutional protections should not be so easily malleable and subject to the whim of the judges who preside over a case on a given day.

This intra-state example shows that allowing judges to determine the reliability of hearsay produced random and highly inconsistent outcomes. Additionally, because all of the courts were in the *same* state and were applying the *same* reliability test, this intra-state example may also rule out any innocent explanation for the variation. A further examination of a single court within a state will provide strong evidence that judicial manipulation is the most likely explanation.

3. Intra-Court Analysis

Trial courts from different states could have innocently and unintentionally used completely *opposite* facts to reach the *same* finding of reliability in two different cases. Also, different levels of courts within the same state could have innocently and unintentionally used the *identical* facts to reach *different* findings of reliability in the same case.

Even if devoid of manipulative intent such inter- and intra-state variations are not acceptable under the United States Constitution. Nonetheless, a further analysis of a single state, and a single court within that state, may rule out innocent explanations altogether. More specifically, when a single court reaches

---

70. *Id.* at 664.

71. *Crawford*, 2001 Wash. App. LEXIS 1723, at *17-18.

72. *Id.*

73. *Crawford*, 54 P.3d at 664.

74. *Id.*

75. *See, e.g.*, People v. Farrell, 34 P.3d 401, 402 (Colo. 2001) (reinstating the trial court's finding of reliability, and therefore admissibility, of hearsay evidence).

the same finding of reliability with completely opposite facts in two different cases, the most likely explanation is judicial manipulation.

This point is well illustrated by the Colorado Supreme Court in *People v. Farrell*,[76] where the defendant was charged with murder, among other crimes.[77] At trial, the State introduced the hearsay statement of the co-defendant who was unavailable for live testimony.[78] The co-defendant had previously given a statement to police that inculpated himself and the defendant in the crimes.[79]

The court determined that the co-defendant's statement was reliable and therefore admissible against the defendant for several now familiar reasons. The court found that the co-defendant gave a great level of detail and did not make the statement in exchange for "any offers of leniency or special deals."[80] Also, the court found that the co-defendant, who was in police custody, "made his statement *immediately after* the criminal episode," which further enhanced its reliability.[81]

Conversely, the very same court less than four months earlier decided *Stevens v. People*,[82] where the defendant was also charged with murder, among other crimes.[83] At trial, the State also introduced the hearsay statement of a co-defendant who was unavailable for live testimony.[84] This co-defendant had also previously given a statement to police that inculpated himself and the defendant in the crimes.[85]

In *Stevens*, the court also determined that the co-defendant's statement was reliable and therefore admissible against the defendant for several reasons. Just as in *Farrell*, the court found that the co-defendant provided many details about the murder and did "not receive any deals in exchange for his statement."[86] Unlike *Farrell*, however, the court found enhanced reliability in the co-defendant's statement because he was not in police custody at the time. Moreover, he made his statement after "*two years had passed*" from the time of the murder.[87]

This contradiction offends not only the Constitution, but also the fundamental concepts of consistency and logic. In one case, the court found a statement reliable and therefore admissible in large part because it was made *immediately*

---

76. *Id.*

77. *Id.* A jury convicted the defendant of intentional first-degree murder, felony first-degree murder, robbery of an at-risk adult, aggravated robbery, second degree kidnapping, two counts of second-degree burglary, theft, first-degree criminal trespass, and two counts of conspiracy. *Id.*

78. *See id.* at 404.

79. *See id.* at 403.

80. *Id.* at 407.

81. *Id.* (emphasis added).

82. Stevens v. People, 29 P.3d 305 (Colo. 2001).

83. *Id.* at 308. The defendant was convicted of first-degree murder, conspiracy to commit first-degree murder, and solicitation to commit first-degree murder. *Id.*

84. *See id.* at 310.

85. *See id.*

86. *Id.* at 316.

87. *Id.* (emphasis added).

after the alleged crime.[88] In another case, the same court found a statement reliable and therefore admissible in large part because it was made *two years after* the alleged crime.[89] Given that these two decisions were the product of the same court and were issued within four months of each other, one may reasonably conclude that judicial manipulation lies at the heart of the inconsistency.

III. THE "SEA CHANGE": *CRAWFORD* AND *DAVIS*

Although it took twenty-four years, the Court finally revisited the *Roberts* reliability test that had seemingly allowed judges to dispense with confrontation rights on mere whim. This change in course came in the form of two cases—*Crawford* and *Davis*—thought to be so significant that many courts and commentators have hailed them as a "sea change" in Confrontation Clause jurisprudence.[90] Other commentators have been equally dramatic, claiming a "Copernican shift in federal constitutional law" and a "revolutionary decision in the law of evidence."[91] Although the ultimate accuracy of these claims is the subject of Part V of this Article, the Court's bold language in *Crawford* and *Davis* supported these grand predictions of the day.

A. Crawford v. Washington—The Paradigm Shift

Regardless of whether irreconcilable lower court rulings were due to judicial manipulation or a more benign explanation, the Court in *Crawford v. Washington* finally acknowledged that the *Roberts* reliability test was a failure.[92] In *Crawford*, the Court was loud and clear in its criticism, stating that "we do not think the Framers meant to leave the Sixth Amendment's protection to . . . amorphous notions of 'reliability.'"[93] Further, "[a]dmitting statements deemed reliable by a judge is fundamentally at odds with the right of confrontation."[94] Instead, the

88. People v. Farrell, 34 P.3d 401, 407 (Colo. 2001).

89. *Stevens*, 29 P.3d at 316.

90. *See* State v. Grace, 111 P.3d 28, 36 (Haw. Ct. App. 2005) ("[e]ffecting a sea change in our understanding of the [C]onfrontation [C]lause . . ."); Chris Hutton, *Sir Walter Raleigh Revived: The Supreme Court Re-vamps Two Decades of Confrontation Clause Precedent in* Crawford v. Washington, 50 S.D. L. REV. 41, 61 (2005) ("[t]his is a sea change for prosecution of cases involving child witnesses."); Andrew King-Ries, State v. Mizenko*: The Montana Supreme Court Wades into the Post-*Crawford *Waters*, 67 MONT. L. REV. 275, 313 (2006) ("*Mizenko*, therefore, recognizes *Crawford's* sea-change in confrontation rights, ...")

91. Jerome C. Latimer, *Confrontation After* Crawford*: The Decision's Impact on How Hearsay is Analyzed Under the Confrontation Clause*, 36 SETON HALL L. REV. 327, 329 (2006) (citations omitted) (citing a lengthy collection of comments and catch-phrases used to describe the *Crawford* decision).

92. *See* Crawford v. Washington, 541 U.S. 36, 60 (2004).

93. *Id.* at 61.

94. *Id.*

Clause demands "that reliability be assessed in a particular manner: by testing in the crucible of cross-examination."[95]

On a practical level, the Court's chief criticism was that the *Roberts* reliability test was "so unpredictable that it fail[ed] to provide meaningful protection from even core confrontation violations."[96] These core violations consisted of admitting into evidence hearsay statements made to police while the declarant was in police custody and being interrogated.[97] The involvement of police in the production of hearsay evidence against a defendant is "the principal evil at which the Confrontation Clause was directed,"[98] and the *Roberts* reliability test simply offered no consistent or substantial protection.

In addition, the Court acknowledged a more fundamental problem. Put simply, "[v]ague standards are manipulable,"[99] and "judges, like other government officers, could not always be trusted to safeguard the rights of the people."[100] Despite this problem, "[t]he *Roberts* test allow[ed] a jury to hear evidence, untested by the adversary process, based on a mere judicial determination of reliability."[101] Succinctly stated, "[d]ispensing with confrontation because testimony is obviously reliable is akin to dispensing with jury trial because a defendant is obviously guilty. This is not what the Sixth Amendment prescribes."[102]

The Court then replaced the *Roberts* reliability test with a new test.[103] Under *Crawford*, the State may use the prior testimony of a witness who is unavailable for trial only if the defendant has had an opportunity to cross-examine that statement.[104] Absent that opportunity, the Confrontation Clause requires that the statement be excluded from evidence.[105]

This summary, however, is an oversimplification in at least one regard: the new rule actually only applies to "testimonial" hearsay.[106] If the hearsay being offered at trial is nontestimonial, the defendant is not afforded such constitutional protection.[107] Deserved criticism was directed at the Court for its failure to define

---

95. *Id.*
96. *Id.* at 63.
97. *Id.* at 63.
98. *Id.* at 50.
99. *Id.* at 68.
100. *Id.* at 67.
101. *Id.* at 62.
102. *Id.*
103. *Id.* at 68.
104. *Id.*
105. *Id.*
106. *Id.*
107. *See id.* ("Where nontestimonial hearsay is at issue, it is wholly consistent with the Framers' design to afford the States flexibility in their development of hearsay law . . . ."). Actually, in the years immediately following *Crawford*, the debate raged as to whether nontestimonial hearsay was afforded any residual protection under the old *Roberts* test. Today, however, most jurisdictions have abandoned *Roberts* completely based on the Court's dicta in *Davis*, as well as other subsequent case law.

the term "testimonial" but the Court did hold that "[s]tatements taken by police officers in the course of interrogations are [] testimonial under even a narrow standard."[108] This vagueness, in turn, led to much debate about the meaning of the term "interrogation."[109] Two years later, the Court decided *Davis v. Washington* in which it developed a potentially workable framework to distinguish between testimonial and nontestimonial hearsay.

B. Davis v. Washington—The Emerging Framework

Although the Court in *Davis v. Washington* finally expanded *Crawford's* framework, the two cases, even in combination, offer limited guidance regarding the vast majority of potential scenarios in the universe of hearsay evidence.[110] Instead, the focus of *Davis* is on hearsay statements that are produced through police interrogation of the declarant.[111] Within this focus, the Court in *Davis* developed a potentially workable framework to determine what types of hearsay are testimonial, and therefore must be excluded by the Confrontation Clause, and what types are nontestimonial, and therefore are not affected by the Clause.[112]

In *Davis*, the Court heard two consolidated cases and, between the two sets of facts, developed one rule.[113] The *Davis* Court held that hearsay is testimonial and must be excluded under the Clause when "the primary purpose of the interrogation is to establish or prove past events potentially relevant to later criminal prosecution."[114] Conversely, hearsay is nontestimonial and therefore not affected by the Clause when the statements are "made in the course of police interrogation under circumstances objectively indicating that the primary purpose of the interrogation is to enable police assistance to meet an ongoing emergency."[115]

This holding seemed to conform to the Court's statement in *Crawford* about the dangers of government-manufactured hearsay.[116] If the police were responding to an ongoing emergency, their goal would be to protect crime victims, and presumably would have no opportunity to fabricate, mold, or

---

108. *Crawford*, 541 U.S. at 52.

109. *See, e.g.*, Michael D. Cicchini & Vincent Rust, *Confrontation After* Crawford v. Washington*: Defining "Testimonial"*, 10 LEWIS & CLARK L. REV. 531, 554 (2006) (advocating a solution providing that all accusatory hearsay be defined as testimonial to comport with the goal of judicial constraint put forth in *Crawford*); Josephine Ross, *After* Crawford *Double-Speak: "Testimony" Does Not Mean Testimony and "Witness" Does Not Mean Witness*, 97 J. CRIM. L. & CRIMINOLOGY 147, 215-17 (2006) (advocating, in substance, for the same solution).

110. *See Crawford*, 541 U.S. at 75; *see also* Davis v. Washington, 547 U.S. 813, 822 (2006) (where court does not "attempt [ ] to produce list of conceivable statements in response to police interrogation.")

111. *See Davis*, 547 U.S. at 822 n.1.

112. *Id.* at 821-22.

113. *Id.* at 817-19.

114. *Id.* at 822.

115. *Id.*

116. *See Crawford*, 541 U.S. at 56 n.7.

manipulate statements to suit the prosecution.[117] Conversely, if the police are in an investigative mode and looking to "establish or prove past events,"[118] the danger of manipulation would be quite real, and statements gathered under these circumstances should be excluded by the Clause.[119]

The Court then applied this new rule to the two factual scenarios before it. In one of the consolidated cases, *Davis v. Washington*, the alleged victim called 911 to report that the defendant was, at the time of the call, jumping on her and striking her with his fists.[120] After more dialogue, she then reported that the defendant had left the residence with a third party.[121] The conversation continued, and the 911 operator obtained additional information.[122] At trial, the alleged victim was unavailable, and the State introduced her hearsay statements to the 911 operator over the defendant's confrontation objection.[123]

The Court held that under these facts, the initial statements to the 911 operator were nontestimonial and therefore properly admitted because the alleged victim "was speaking about events as *they were actually happening*, rather than 'describ[ing] past events.'"[124] Her 911 call was "a call for help against a bona fide physical threat," and her statements "were necessary to be able to resolve the present emergency."[125] However, the Court also stated that "the emergency appears to have ended (when [the defendant] drove away from the premises)."[126] After that point in time, the alleged victim's statements "were testimonial, not unlike the 'structured police questioning' that occurred in *Crawford*."[127]

In *Hammon v. Indiana*, the other of the consolidated cases decided in *Davis*, the police responded in person to the home of an alleged victim.[128] Upon arrival, they found her "somewhat frightened" on the porch.[129] She and the police went into the living room, while the defendant remained in the kitchen, and she told police that the defendant had battered her.[130] During this time, the defendant attempted to intervene in the interview and "became angry when [the officer] insisted that [he] stay separated" from the alleged victim.[131] At trial, the alleged

117. *Id.*

118. *Davis*, 547 U.S. at 822.

119. *See id.* at 832 n.6.

120. *Id.* at 817.

121. *Id.* at 818.

122. *Id.* ("[The 911 operator] then gathered more information about Davis (including his birthday), and learned that Davis had told [her] that his purpose in coming to the house was 'to get his stuff,' since [she] was moving. [She then] described the context of the assault.").

123. *Id.* at 819.

124. *See Davis*, 547 U.S. at 827-28 (quoting Lilly v. Virginia, 527 U.S. 116, 137 (1999)).

125. *Davis*, 547 U.S. at 827 (emphasis omitted).

126. *Id.* at 828.

127. *Id.* at 829 (quoting Crawford v. Washington, 541 U.S. 36, 53 n.4 (2004)).

128. *Davis*, 547 U.S. at 819.

129. *Id.* (quoting Hammon v. Indiana, 829 N.E.2d 444, 446 (Ind. 2005)).

130. *Davis*, 547 U.S. at 819.

131. *Id.* at 819-20.

victim was unavailable, and the State introduced her hearsay statements to the police over the defendant's confrontation objection.[132]

The Court held that under these facts, all statements to the police were testimonial and therefore were admitted in error because "the interrogation was part of an investigation into possibly criminal past conduct," and "[t]here was no emergency in progress."[133] Even though the defendant became angry and the "officers forcibly prevented [him] from participating in the interrogation," there simply was "no immediate threat" to the alleged victim.[134] The Court explained that "[o]bjectively viewed, the primary, if not indeed the sole, purpose of the interrogation was to investigate a possible crime . . . ."[135]

A fair reading of *Crawford* and *Davis* would lead an objective, dispassionate person to conclude that when someone is calling out for help and reporting an ongoing crime, that statement is nontestimonial and may be admitted.[136] However, if statements are made after the emergency dissipates—e.g., after the perpetrator leaves the scene or after the police arrive and begin questioning the alleged victim about what happened—the statement is testimonial and must be excluded.[137]

The problem, however, is that judges cannot be assumed to be objective and dispassionate. In fact, the Court in *Crawford* had already rearticulated what the Framers knew long ago: "judges, like other government officers, [can] not always be trusted to safeguard the rights of the people."[138] This wisdom has become painfully obvious in lower court decisions post-*Davis*.

IV. THE LOWER COURTS: CIRCUMVENTING *CRAWFORD* AND *DAVIS*

The post-*Davis* years have produced perhaps the most poorly reasoned and disingenuous court decisions in Confrontation Clause jurisprudence. Repeatedly, courts completely distort the Clause—as interpreted in *Crawford* and *Davis*—in order to accomplish a predetermined goal of admitting hearsay evidence against defendants.[139] The means by which courts accomplish this result are limited only by judicial imagination and creativity. However, the most common judicial approaches include expanding the ongoing emergency, distorting the primary purpose test, and expanding the forfeiture doctrine.

---

132. *Id.*
133. *Id.* at 829.
134. *Id.* at 829-30.
135. *Id.* at 830.
136. *See id.* at 832.
137. *See id.* at 830.
138. *Crawford*, 541 U.S. at 67.
139. *See* Richard D. Friedman, Crawford, Davis *and Way Beyond*, 15 J.L. & POL'Y 553, 563 (2007) (stating that courts will look "for whatever toehold they can find to admit accusatory statements that were made absent an opportunity for confrontation").

A. Expanding the Ongoing Emergency

Perhaps the most common way courts have distorted the plain language of *Crawford* and *Davis* in order to dispense with the right of confrontation is simply to expand the concept of the ongoing emergency. If courts can somehow find that a declarant's statements to police were made within an ongoing emergency situation, then the statements can be labeled as nontestimonial and thus admitted into evidence without ever being cross-examined.

For example, in the Oregon case *State v. Camarena*, the defendant was accused of striking the alleged victim in the eye.[140] The alleged victim then told the defendant that she was going to call the police, and the defendant promptly left the apartment and drove away in a car.[141] After about one minute, the alleged victim called 911 but hung up.[142] The operator called back and questioned the alleged victim, who then accused the defendant.[143] After being told that the defendant hit the alleged victim, the operator asked "[w]here is he at now?"[144] The alleged victim responded, "I don't know. He took the car and he left."[145] She then answered a number of other questions about the alleged assault and in the process provided the defendant's "name and driver's license number."[146] The alleged victim was unavailable at trial, but the court held her statements to be nontestimonial and therefore admissible.[147]

In analyzing the case, the court relied on *Davis* in acknowledging that if "the primary purpose of the interrogation [was] to establish or prove past events potentially relevant to later criminal prosecution," then the statements would be testimonial and inadmissible.[148] The court went on to find that the alleged victim's statements were "referring to past events" and were not describing events that were ongoing.[149] In fact, the accusations were made *after* the defendant had left the residence in a car.[150]

However, the court still found the statements to be nontestimonial and therefore admissible because it was *possible* that the domestic assault could conceivably have been renewed before the police arrived.[151] The court found that, even though the defendant had just driven away from the residence, "the danger of a renewal of the domestic assault had *not necessarily* or fully abated."[152] The

140. State v. Camarena, 145 P.3d 267, 269 (Or. Ct. App. 2006).
141. *Id.*
142. *Id.*
143. *Id.*
144. *Id.*
145. *Id.*
146. Id. at 269-70.
147. *Id.* at 274.
148. *Id.* at 272.
149. *Id.* at 275 (internal quotations omitted).
150. *Id.* at 269.
151. *Id.* at 275.
152. *Id.* (emphasis added).

court believed that this theoretical possibility of future criminal activity against the declarant was enough to constitute an ongoing emergency during the time at which she accused the defendant.[153]

Similarly, in the Minnesota case of *State v. Warsame*, the alleged victim accused the defendant of domestic violence by describing past events.[154] These allegations occurred both after the alleged crime and after the defendant drove away from the scene.[155] In this case, however, the alleged victim was actually in the presence of at least two police officers at the time she accused the defendant of battery.[156]

In finding the statements nontestimonial and therefore admissible, the Minnesota Court of Appeals expanded the meaning of ongoing emergency.[157] However, in this case, the court could not assert that the defendant might renew the alleged domestic attack; the alleged victim was in the actual presence of police officers when making her statements.[158] Instead, the court expanded the concept of ongoing emergency to include other people and other locations.[159]

The court found that it was conceivable that the defendant posed a danger to others, including the car passenger with whom he drove away.[160] Therefore, the court reasoned that the alleged victim's statements about what had happened to *her* should be classified as nontestimonial.[161] The court explained, "[w]e conclude that the 'ongoing emergency' referred to in *Davis* . . . need not be limited to the complainant's predicament or the location where she is questioned by police."[162]

These two decisions—*Camarena* and *Warsame*—should be openly criticized as judicial manipulation designed to reach a predetermined outcome. First in applying the *Davis* analysis to the facts of these two cases, the alleged victims were describing "past events," rather than "events as they were actually happening."[163] Consequently, because there was "no immediate threat"[164] to the alleged victims, the statements were testimonial and should have been excluded.

Second, the Court in *Davis* stated that even when an alleged victim does describe ongoing events that constitute a true emergency—which was *not* the case in either *Camarena* or *Warsame*—the emergency ends when the defendant drives "away from the premises."[165] Just as in *Davis*, the defendants in both *Camarena* and *Warsame* had already driven away from the supposed crime scenes *before* the

153. *Id.*
154. State v. Warsame, 723 N.W.2d 637, 638-39 (Minn. Ct. App. 2006).
155. *Id.*
156. *Id.*
157. *Id.* at 643
158. *Id.* at 638.
159. *Id.* at 641-42.
160. *Id.* at 641.
161. *Id.*
162. *Id.*
163. Davis v. Washington, 547 U.S. 813, 827 (2006).
164. *Id.* at 830.
165. *Id.* at 828.

allegations were made.[166] Once again, for this reason, the statements were testimonial and should have been excluded.

Under the rationale of *Davis*, an ongoing emergency cannot exist when the defendant has left the scene and the alleged victim calls the police to report past events.[167] Furthermore, with *Camarena*, *Warsame*, and *Davis* all dealing with domestic violence allegations, the facts of these cases do not indicate a random crime spree where other members of the community might be at risk. Instead, the police had not even a hint of evidence in any of the cases that another crime was remotely likely.

Was it not just as likely that the defendant in *Davis* could have renewed his alleged assault after he drove away from the residence but before the police arrived? Of course it was, but this theoretical possibility did not create an ongoing emergency in *Davis*, nor should it have in *Camarena*. Was it not just as likely that the defendant in *Davis* could have committed a crime against the passenger in his car with whom he left? Of course it was, but again, this theoretical possibility did not create an ongoing emergency in *Davis*, nor should it have in *Warsame*.

The courts' analyses in *Camarena* and *Warsame* are disingenuous and intellectually deficient. If situations can be upgraded to ongoing emergencies simply because a defendant might commit an unspecified crime at some unspecified time in the future against an unspecified victim, then every situation will be automatically transformed into an ongoing emergency. In that case, every factual scenario would swallow the *Davis* rule whole, as no statement could ever be classified as testimonial. As the Connecticut Supreme Court stated in a similar context:

> Even if we were to accept the State's contention that the complainant was hysterical and in need of medical assistance, those portions of the call explaining what had happened to her at the hands of the defendant did not point to an ongoing emergency, but rather to an explanation of past events. Put differently, accepting the State's arguments on this point would render meaningless the distinction drawn by the United States Supreme Court, *as they would render virtually any telephone report of a past violent crime in which a suspect was still at large, no matter the timing of the call, into the report of a "public safety emergency."*[168]

Unfortunately, the examples cited herein are not anomalous. Numerous other state courts—including those in Wisconsin, Texas and Nevada—have been highly creative in defying reason, logic, and the holding of *Davis* in order to make any statement part of a fictional, ongoing emergency.[169] Once that is

166. *Camarena*, 145 P.3d at 269; *Warsame*, 723 N.W.2d at 639.

167. *Davis*, 547 U.S. at 827.

168. State v. Kirby, 908 A.2d 506, 523-24 n.19 (Conn. 2006) (emphasis added).

169. *See, e.g.*, Harkins v. State, 143 P.3d 706, 715 (Nev. 2006) (holding that allegations made during a 911 call, where there was no proof that the defendant "left the area completely at that point," were part of an ongoing emergency); Vinson v. State, 221 S.W.3d 256, 264-65 (Tex. App. 2006) (holding that the police asking the alleged victim "what happened?" was the functional

accomplished, the hearsay statements are labeled nontestimonial and, more significantly, admitted without cross-examination.

B. Distorting the Primary Purpose Test

Even when there is no possibility of finding an "ongoing emergency"—e.g., in cases where the hearsay statement was made days or weeks after the alleged crime—courts still have other means of labeling hearsay as nontestimonial, thereby placing it beyond the reach of the Clause. One such way is to find that the statement was made, or obtained, for some "primary purpose" other than the investigation of a crime.[170] This tactic is very common in cases involving medical professionals who act on behalf of, or in concert with, police.[171]

For example, in the Ohio case *State v. Stahl*, the alleged victim reported to police that the defendant had "orally raped her" the previous day.[172] After taking the alleged victim's statement, an officer took her to the Developing Options for Violent Emergencies ("DOVE") unit, a medical facility that is funded by the state Attorney General's Office and gathers and retains physical evidence for use in prosecuting crimes.[173] At the DOVE unit, the alleged victim signed a waiver stating, "I authorize the release of evidence, *information* (including protected health information), clothing, colposcope photos, and photography documentation of injuries to law enforcement agency *for use only in the investigation and prosecution of this crime*."[174]

The alleged victim was then interviewed by a Sexual Assault Nurse Examiner ("SANE nurse") to whom she repeated the allegations she made to the officer.[175] The officer was also present when the SANE nurse interviewed the alleged victim in the examination room.[176] The SANE nurse then took photos of the alleged victim's mouth, collected "nail scrapings, oral swabbings, and material retrieved with dental floss."[177] She also collected a napkin from the alleged victim's pocket

---

equivalent of asking "whether an emergency existed," and therefore was part of an ongoing emergency); State v. Rodriguez, 722 N.W.2d 136, 147-48 (Wis. Ct. App. 2006) (holding that questioning by police the day following the incident, when police responded to return some property, was part of an ongoing emergency); *see also* State v. Washington, 725 N.W.2d 125, 132-33 (Minn. Ct. App. 2006) (holding that allegations made after the defendant left the area were part of an ongoing emergency because the alleged victim asked the police to "watch her when she left the apartment to make sure that she was not assaulted again").

170. *See, e.g.*, State v. Stahl, 855 N.E.2d 834, 836 (Ohio 2006).

171. *See* Elizabeth J. Stevens, Comment, *Deputy-Doctors: The Medical Treatment Exception after* Davis v. Washington, 43 CAL. W. L. REV. 451, 472 (2007) ("[U]nder *Davis*, courts should treat health care providers as agents of the police and their interactions with the declarant as police interrogation" based on principles of agency law.).

172. *Stahl*, 855 N.E.2d at 836.

173. *Id.*

174. *Id.* at 837 (emphasis added).

175. *Id.*

176. *Id.*

177. *Id.*

thought to contain physical evidence related to the assault.[178] No medical doctor ever treated, or even saw, the alleged victim.[179] The alleged victim was unavailable at trial, but the court inexplicably held that her statements to the SANE nurse were nontestimonial and therefore admissible.[180]

However, SANE nurses, or even medical doctors or any other medical professionals, can act as agents of the police, just as the 911 operator did in *Davis*.[181] The relevant inquiry is whether the SANE nurse was performing the function of, or acting on behalf of, the police or prosecutor.[182] For example, in *Davis*, the 911 operator took the call, listened to allegations, and asked follow-up questions of the alleged victim.[183] All of this questioning was done on behalf of police, who then took the information and responded to the scene.[184]

Similarly, in *Stahl*, the SANE nurse met with the officer and the alleged victim, listened to the allegations, took photographs, and collected evidence from the alleged victim.[185] The DOVE unit's release form even stated that the information and evidence would be used in "the investigation and prosecution of this crime."[186] In fact, not only was the SANE nurse an agent of police,[187] but her primary purpose was to collect information and evidence "to establish or prove past events potentially relevant to later criminal prosecution."[188]

However, the *Stahl* court ignored the overwhelming evidence of the SANE nurse's primary purpose. Instead, it relied on rank speculation as to the alleged victim's *possible* expectations, ultimately finding the statements to the SANE nurse nontestimonial and therefore admissible.[189] The court stated that because the alleged victim first spoke to police, and *then* was taken to the DOVE unit, she "*could reasonably have assumed* that repeating the same information to a nurse or other medical professional served a separate and distinct medical purpose."[190]

If the legal community values the Constitution in any significant way, the *Stahl* court's conclusion must be openly and widely criticized. The only reasonable conclusion is that the primary, if not sole, purpose of the interview by the SANE nurse – whether viewed from the perspective of the SANE Nurse, the alleged victim, the police officer, or a hypothetical independent observer – was to

178. *Id.*
179. *Id.*
180. *Id.* at 845-47.
181. Stevens, *supra* note 171, at 479-81.
182. *Id.* at 472.
183. Davis v. Washington, 547 U.S. 813, 817 (2006).
184. *Id.* at 818.
185. *Stahl*, 855 N.E.2d at 837.
186. *Id.*
187. Stevens, *supra* note 171, at 479-80 (arguing that SANE nurses "should categorically be treated as police agents" due in part because of their intimate involvement with the prosecutor, the state, and the criminal process on multiple levels).
188. *Stahl*, 855 N.E.2d at 841 (quoting *Davis*, 547 U.S. at 813-14).
189. *Id.*
190. *Id.* at 846 (emphasis added).

collect and preserve evidence of past events for future criminal prosecution.[191] The alleged victim's statements to the SANE nurse were therefore testimonial and should have been excluded.

Unfortunately, police and prosecutors can take steps to make it even easier for judges to bypass the Clause. For example, in the Minnesota case *In re A.J.A.*, parents reported to police that their child informed them that she was sexually assaulted.[192] The police then purposely took several steps to *avoid* doing anything that remotely resembled a police investigation so that they could distance themselves from any hearsay statements.[193]

First, a detective responded to the home of the parents and the alleged victim "dressed in plain clothes" and purposely avoided contact with the alleged victim.[194] The detective then contacted the district attorney's office for information on "the appropriate person to examine" the alleged victim.[195] The detective provided that information to the parents and instructed *them* to arrange for an examination.[196] A nurse subsequently interviewed the alleged victim who identified the defendant.[197]

However, as the police and prosecutor knew under Minnesota law, the nurse was a mandatory reporter.[198] Therefore, after hearing the accusations of abuse, the nurse simply repeated the information to the police, who initiated the criminal prosecution.[199] Further, when the alleged victim later became unavailable for trial, the State argued that because the interview was conducted by a nurse, rather than by police, the primary purpose of the interview was not to investigate a past crime, but rather to render medical aid.[200]

The trial court, however, applied a substance-over-form analysis and correctly held that the nurse "was simply a surrogate for police investigation and interview."[201] As a result, the hearsay was testimonial and therefore inadmissible.[202] Nonetheless, the appellate court reversed and found the statements nontestimonial, and therefore admissible, for a completely irrelevant

---

191. *Id.* at 847-48 (Lanzinger, J., dissenting) (stating that the "primary purpose for the police to take [the alleged victim] to the DOVE unit was for collection of evidence, not medical treatment," and that "[t]he forensic aspects of the DOVE unit are clear from the precise nature of the nurse's activities in collecting evidence").

192. *In re* A.J.A., No. A06-479, 2006 Minn. App. LEXIS 988, at *6 (Minn. Ct. App. Aug. 29, 2006).

193. *See id.* at *7.

194. *Id.*

195. *Id.*

196. *Id.*

197. *Id.* at *8.

198. *Id.* at *7 (citing Minn. Stat. § 626.556, subd. 3(a)(1) (2004)).

199. *Id.* at *8-9. *See* Stevens, *supra* note 171, at 478-79 (arguing that the mandatory reporter laws, which are often enforced by threat of criminal penalty, "effectively deputize all medical practitioners").

200. *See In re A.J.A.*, 2006 Minn. App. LEXIS 988, at *12.

201. *Id.*

202. *Id.*

reason: that the nurse "expressly rejected the suggestion that police could influence her examination of her patients" and therefore "was not acting as an agent of or in concert with the government."[203]

Can judges really dispense with the Clause this easily? The State, with the blessing of the courts, is simply bypassing the Clause by shifting investigative duties away from police to mandatory reporters who, by law,[204] must immediately report back to police. Consequently, the police still obtain, indirectly, the allegations they need to initiate a criminal prosecution. Additionally, should the alleged victim later become unavailable for trial, the statements can be admitted without any cross-examination because they were first made to a nurse, rather than to the police. Indeed "[i]f testimonial evidence can be admitted through this 'middleman' mechanism, then the Confrontation Clause's renewed vigor post-*Crawford* is a sham."[205]

Interestingly, a comparison of the two cases discussed in this section—one from Ohio and one from Minnesota—invokes memories of the inconsistent and irreconcilable rulings from earlier cases that applied the *Roberts* reliability test. That is, in Ohio, statements will be labeled nontestimonial and therefore admissible if the police *first* take a statement from the alleged victim and *then* take her to a nurse to repeat the statement.[206] Conversely, in Minnesota, the opposite is true.[207] Statements will be labeled nontestimonial and therefore admissible if the police *avoid* taking a statement from the alleged victim but rather send her *directly* to a nurse.[208] Once again, outright judicial manipulation designed to reach a predetermined outcome leads to an irreconcilable contradiction.

C. Expanding the Forfeiture Doctrine

Even when courts are unable to fabricate an ongoing emergency, or when police cannot use surrogates in their place, courts have found other ways to bypass the Confrontation Clause. One such way is to expand the scope of the forfeiture doctrine. This doctrine is essentially an exception to the general ban on testimonial hearsay and allows even testimonial hearsay to be admitted if a court finds that the defendant acted in such a way as to forfeit his right of confrontation.[209]

---

203. *Id.* at *11.

204. *Id.* at *9 (citing Minn. Stat. § 626.556, subd. 3(a)(1) (2004).

205. Stevens, *supra* note 171, at 476.

206. *See* State v. Stahl, 855 N.E.2d 834 (Ohio 2006).

207. *See* State v. Washington, 725 N.W.2d 125 (Minn. Ct. App. 2006).

208. *Stahl*, 855 N.E.2d at 834.

209. James F. Flanagan, *Foreshadowing the Future of Forfeiture/Estoppel by Wrongdoing:* Davis v. Washington *and the Necessity of the Defendant's Intent to Intimidate the Witness*, 15 J.L. & POL'Y 863, 864-65 (2007).

> This [forfeiture] doctrine provides that a defendant who deliberately acts to prevent a witness from testifying loses any right to object to the admission of the witnesses' testimonial hearsay statement.... This doctrine has always required that the defendant *specifically intend to prevent the witness from testifying*, and was previously limited to cases of deliberate *witness tampering*.[210]

More specifically, the doctrine has "always required the defendant's knowledge of the declarant's status as a witness and intentional efforts to prevent that witness from testifying."[211] In the wake of *Crawford* and *Davis*, however, and in an effort to gain more convictions, courts have expanded the scope of this doctrine.[212] Even more alarming is the bold nature in which courts have made this move. In fact, courts have openly stated that they are seeking not to apply *Crawford* and *Davis*, but rather to circumvent them.

For example, in the Wisconsin case *State v. Jensen*, the defendant was charged with murdering his wife, and the State attempted to introduce his wife's hearsay statements into evidence.[213] The trial court, and eventually the state supreme court, agreed with the defendant that the statements were, in fact, testimonial.[214] Further, because the wife was deceased long before any criminal action had begun, there was no possibility that the killer had murdered her for the purpose of preventing her from testifying in court.[215] This obvious fact precluded the possibility that the defendant could have forfeited his right of confrontation.

However, this reality did not prevent the state supreme court from finding a way to admit the statements. The court expanded the forfeiture doctrine by requiring the trial court to determine, by a mere preponderance of the evidence, *not* whether the defendant acted to prevent the witness from testifying, but whether the defendant was guilty of the underlying murder.[216] Without the murder, the reasoning continues, the wife would hypothetically be available to testify at trial—albeit a trial that would not even take place if she were alive.[217] Therefore, if the judge believed that the defendant murdered his wife, the defendant forfeited his right of confrontation.[218]

The state supreme court acknowledged that "[r]equiring the court to decide by a preponderance of the evidence the very question for which the defendant is on trial may seem, at first glance, troublesome."[219] However, it still boldly declared that "[i]n essence, we believe that in a post-*Crawford* world the broad

---

210. *Id.* (emphasis altered).
211. *Id.* at 874.
212. *See id.* at 877-78.
213. State v. Jensen, 727 N.W.2d 518, 520-21 (Wis. 2007).
214. *Id.* at 521.
215. *Id.*
216. *Id.* at 536.
217. *See id.*
218. *Id.*
219. *Id.* at 535 (citing United States v. Mayhew, 380 F. Supp. 2d 961, 967 (S.D. Ohio 2005)).

view of forfeiture by wrongdoing espoused by [Professor] Friedman and utilized by various jurisdictions since *Crawford's* release *is essential.*"[220]

The court's declaration is nothing short of a plain and open admission that it was expanding the forfeiture doctrine to circumvent *Crawford's* strengthened confrontation right. In so doing, the court found comforting that "[s]ince the release of *Crawford*, many jurisdictions have either adopted the forfeiture by wrongdoing doctrine if they had not done so before, or *they have expanded the doctrine . . . .*"[221]

However, the court should have found its new rule quite troublesome. First, in its rush to find a way to admit evidence against the defendant, the court merely abandoned the "substantive doctrine that was adopted by the founders."[222] Second, as the trial court had previously warned, the expanded forfeiture doctrine—then espoused by the prosecutor and later adopted by the supreme court—"would render superfluous the doctrine of dying declarations."[223] Third, the court only selectively adopted the views of Professor Friedman, whom it cited in support of its new rule, while ignoring his views that worked against the adoption of its new rule.[224] Fourth, and most significantly, as Justice Butler stated in his dissent:

> [A]pplying the forfeiture doctrine to admit testimonial evidence when the defendant is on trial for the crime that rendered the witness unavailable, absent any showing that the defendant's purpose was to procure the absence of the witness to keep him or her from testifying at trial, *places the cart before the horse.*[225]

The far-reaching implications of expanding the forfeiture doctrine should not be underestimated. For example, the obvious consequence of the court's new expanded doctrine is that the trial judge must make a finding, by the preponderance of the evidence, on whether the defendant is guilty of the underlying crime.[226] However, the often forgotten consequence is that *after*

---

220. *Id.* (emphasis added).

221. *Id.* at 533 (emphasis added).

222. *Id.* at 545 (Butler, Jr., J., dissenting in part).

223. *Id.* at 546 (Butler, Jr., J., dissenting in part).

224. *Id.* at 545 n.9 (Butler, Jr., J., dissenting in part) ("The majority declines, however, to adopt Professor Friedman's recommendation" to apply a higher burden of proof before finding that the defendant forfeited the right to confront his accuser.).

225. *Id.* at 546 (Butler, Jr., J., dissenting in part) (emphasis added). As it turns out, Justice Butler was correct. After this Article was written, but before it was published, the United States Supreme Court decided *Giles v. California*, 128 S. Ct. 2678 (2008), where it held, with regard to the expanded forfeiture doctrine, that "[w]e decline to approve an exception to the Confrontation Clause unheard of at the time of the founding or for 200 years thereafter." *Id.* at 2693. Further, "[t]he notion that judges may strip the defendant of a right that the Constitution deems essential to a fair trial, on the basis of a prior *judicial* assessment that the defendant is guilty as charged, does not sit well with the right to trial by jury." *Id.* at 2686 (emphasis in original).

226. *See* Stevens, *supra* note 171, at 495 ("A forfeiture hearing should not become a mini-trial

finding the defendant guilty, the very same trial judge must then preside over the defendant's trial.[227] This same predicament in other contexts is a constitutional due process violation.

For example, in *Franklin v. McCaughtry*, the trial judge—coincidentally the very same trial judge that presided in *Jensen*—"took the highly unusual step of filing a memorandum" expressing an opinion on the defendant's guilt while his case was still pending.[228] The United States Court of Appeals for the Seventh Circuit held that "[t]he memorandum demonstrates that [the judge] decided the issue of [the defendant's] guilt long before trial."[229] The problem with this pretrial determination of guilt, of course, is that due process guarantees "a defendant's right to be tried *by an impartial judge*."[230] Therefore, the judge's opinion on the defendant's guilt, as expressed in his memorandum issued *before* the trial, was "a clear violation of [the defendant's] due process rights."[231] Consequently, the conviction was reversed.[232]

Is there any substantive difference between expressing an opinion of guilt in a pretrial memorandum, as in *Franklin*, and expressing an opinion of guilt in a pretrial ruling under the forfeiture doctrine, as in *Jensen*? In both cases, the trial judge has formed an opinion before trial that the defendant is guilty, but then must preside over the defendant's trial. In both cases, the defendant's due process rights are *necessarily* violated. However, in its haste to find a way to reach its predetermined outcome—the admission of hearsay statements against the defendant—the Wisconsin Supreme Court was too short-sighted to see the ramifications of the rule that it rushed to adopt.

Further, and unfortunately, this expanded forfeiture doctrine is not anomalous. Other state courts—including those in Colorado, Kansas, Texas and Ohio—have also expanded the forfeiture doctrine by requiring the trial judge to first make a finding of guilt on the underlying crime, and then preside over the defendant's trial on that very same criminal allegation.[233]

---

at which the judge effectively adjudicates the defendant's guilt under a lesser standard of proof. If 'allow[ing] a jury to hear evidence, untested by the adversary process, based on a mere judicial determination of reliability' was bad, allowing the jury to hear such evidence based on a judicial predetermination of guilt would be far worse.") (alteration in original) (internal footnotes omitted).

227. This pretrial finding of the defendant's guilt on the underlying offense should not be confused with other pretrial findings that are routine, and constitutionally permissible, in criminal law. For example, a judge's belief and finding that the defendant intimidated a witness has nothing to do with whether the defendant committed the underlying crime for which he is standing trial. As another example, a judge's belief and finding that police did *not* violate a defendant's privacy rights during a search of his home has nothing to do with whether the items recovered during the search were illegal or even knowingly possessed by the defendant. Therefore, in these examples, the judge is not making a finding of ultimate guilt.

228. Franklin v. McCaughtry, 398 F.3d 955, 957 (7th Cir. 2005).

229. *Id.* at 961.

230. *Id.* at 959 (emphasis added).

231. *Id.* at 962.

232. *Id.* at 957.

233. *See, e.g.*, United States v. Mayhew, 380 F. Supp. 2d 961, 969 (S.D. Ohio 2005); State v.

V. WHAT WENT WRONG? A CLOSER LOOK AT *CRAWFORD* AND *DAVIS*

Crawford and *Davis* have failed to live up to their billing as a great "sea change" in Confrontation Clause jurisprudence.[234] While the Court claimed to have grand intentions of constraining judicial discretion, post-*Davis* cases in lower courts have shown that little has changed. In fact, as illustrated in Part IV, judicial manipulation is as prevalent today as it was under the old *Roberts* reliability test.[235]

Earlier, however, Part III of this Article purposely cast the *Crawford* and *Davis* decisions in a fairly positive light.[236] The reason for this moderate praise was that if applied by any objective and dispassionate person, the policies and rules developed in the cases would probably produce correct results the majority of the time. In that limited regard, then, the problem lies not with the Court, but with the lower courts' distortion of *Crawford* and *Davis*.

The fundamental problem, however, still lies with the United States Supreme Court. The Supreme Court specifically acknowledged that judges often are *not* objective and dispassionate, yet it still did nothing to constrain their discretion. This lack of resolve is evident in *Crawford* itself, where the Court in one sentence acknowledges that "judges, like other government officers, could not always be trusted to safeguard the rights of the people."[237] However, in the very same paragraph, the Court concludes perplexingly that "[w]e have no doubt that the courts below were acting in utmost good faith when they found reliability."[238]

Unfortunately, this lack of resolve resulted in the implementation of another vague and fact-intensive test by carrying over the broad concepts of *Crawford* into the more specific framework outlined in *Davis*.[239] Now, instead of finding that a statement is reliable, and therefore admissible, trial judges may simply find that the statement is nontestimonial—e.g., that it was made in the course of an ongoing emergency—and is therefore admissible.[240] All the Court did in *Davis* and *Crawford* was change the label for the judicial discretion, while doing nothing to remove the judicial discretion itself, the very thing that the Court (weakly) acknowledged was the underlying problem.[241] The new framework still

---

Meeks, 88 P.3d 789, 794 (Kan. 2004); People v. Moore, 117 P.3d 1, 5 (Colo. Ct. App. 2004); Gonzalez v. State, 155 S.W.3d 603, 609-10 (Tex. App. 2004).

234. See *supra* notes 90 and 91.

235. *See* discussion *supra* Part IV.

236. *See* discussion *supra* Part III.

237. *Crawford*, 541 U.S. at 67.

238. *Id.*

239. *See* Ross, *supra* note 109, at 193 (The "Court may have unconsciously descended down the *Roberts* reliability path the justices so recently abandoned.").

240. Davis v. Washington, 547 U.S. 813, 822 (2006).

241. *See* Lisa Kern Griffin, *Circling Around the Confrontation Clause: Redefined Reach But Not a Robust Right*, 105 MICH. L. REV. FIRST IMPRESSIONS 16, 18 (2006) ("Ironically, the *Davis* formulation is a totality of the circumstances test in many ways more similar to the *Roberts* reliability inquiry . . . ."); Gregory M. O'Neil, Comment, Davis & Hammon*: Redefining the Constitutional Right to Confrontation*, 40 CONN. L. REV. 511, 536 (2007) ("The fact-specific nature

"allows a jury to hear evidence, untested by the adversary process, based on a mere judicial determination."[242]

The new *Davis* test, much like the *Roberts* test, is in practice nothing more than a manipulable, open-ended balancing test that *Crawford* condemned.[243] In essence, "[t]he multiple factors for determining whether a statement is testimonial invite a lack of uniformity in applying the new test. Trial judges may weigh the factors *in any way they wish* to support their conclusion . . . ."[244] Moreover, when deciding the primary purpose of an interrogation, judges *may chose the very factors* they wish to weigh, including whether the declarant's statement was made in the past or present tense, whether the declarant was seeking assistance, the lapse of time between the alleged crime and the statement, and the level of formality in the interrogation.[245]

Additionally, under *Davis*, the "primary purpose" may actually "evolve" at a judicially determined point of the interrogation.[246] This evolution allows a judicial parsing of the statements[247] and serves as yet another tool to admit whatever hearsay statements the court sees fit. As Justice Thomas stated in his dissent in *Davis*, the *Davis* framework, when compared to the *Roberts* framework, is "an equally unpredictable test, under which district courts are charged with divining the 'primary purpose' of police interrogations."[248] Further, the new *Davis* framework "is neither workable nor a targeted attempt to reach the abuses forbidden by the Clause."[249]

Equally significant with regard to the forfeiture doctrine, the Court stated that "[w]e take no position on the standards necessary to demonstrate such forfeiture."[250] This lack of resolve essentially granted the lower courts full discretion to expand the forfeiture doctrine in any way they wished.[251] As a result, courts now routinely make pre-trial findings of guilt on the underlying crime with which the defendant is charged—a level of discretion unknown pre-*Crawford*—in order to admit untested, uncross-examined hearsay.[252]

---

of the Court's inquiry in *Davis* seems to afford courts broad discretion over defendant's confrontation right, similar to that allowed under the *Roberts* test."); Stevens, *supra* note 171, at 495 (warning that continued judicial discretion under *Davis* will "resurrect *Roberts*'s critical flaw in new guise.").

242. *Crawford*, 541 U.S. at 62.

243. Friedman, *supra* note 139, at 563 ("[a] test relying on the terms 'primary purpose' and 'ongoing emergency' is extremely ambiguous . . .").

244. Ross, *supra* note 109, at 193 (emphasis added).

245. *Davis*, 547 U.S. at 827. *See* O'Neil, *supra* note 241, at 543 ("Courts are given the power to create and consider a host of factors in deciding whether an out of court statement should be admissible against the defendant.").

246. *Davis*, 547 U.S. at 828.

247. *See id.* at 829.

248. *Id.* at 834 (Thomas, J., dissenting in part).

249. *Id.* at 842 (Thomas, J., dissenting in part).

250. *Id.* at 833.

251. *See* State v. Jensen, 727 N.W.2d 518, 536 (Wis. 2007).

252. *See id.* This lower court practice, however, should now be curtailed by the Court's

Obviously the *Davis* test was born not only out of the Court's lack of resolve, but also out of its naiveté as to the workings of an actual criminal trial. For example, the Court states that "[w]hile prosecutors may hope that inculpatory 'nontestimonial' evidence is gathered, this is essentially beyond police control. *Their saying that an emergency exists cannot make it be so.*"[253] This statement, however, is simply wrong. In reality, a police officer's saying that an emergency exists *is* precisely the thing that makes it so.[254]

In such circumstances, it is important to keep in mind that the declarant of the statement is necessarily absent from trial, or he or she would simply testify and these confrontation issues would not even arise in the first place. Further, as the cases discussed in this Article illustrate, the defendant will rarely be a witness to the declarant's statement. This leaves only the police officer or other government agent to testify about the purpose of the interrogation, the circumstances surrounding the statement, and even the very *content* of the statement. This uncontradicted police testimony, combined with the lower courts' eagerness to upgrade every situation to emergency status, *is* precisely what makes the emergency.

In sum, then, there is ample blame to go around for the state of post-*Davis* confrontation law. It is true that the lower courts are responsible for distorting the reasonably clear purposes of *Crawford* and *Davis*.[255] However, the Supreme Court knew, and even explicitly stated, that the lower courts "could not always be trusted to safeguard the rights of the people."[256] Despite this awareness, the Court made no serious attempt to constrain judicial discretion. The Court, therefore, must bear the blame for the lower courts' manipulation of the *Davis* test.

VI. EVALUATING POTENTIAL SOLUTIONS

In forums for legal writing, there has been no shortage of opinions on the Confrontation Clause, especially with regard to defining "testimonial hearsay."

---

decision in *Giles v. California*, 128 S. Ct. 2678 (2008), where it struck down the expanded forfeiture doctrine, holding that "[w]e decline to approve an exception to the Confrontation Clause unheard of at the time of the founding or for 200 years thereafter." *Id.* at 2693.

253. *Davis*, 547 U.S. at 832 n.6 (emphasis added).

254. *See* Cicchini & Rust, *supra* note 112, at 547-48 (analogizing to Fourth Amendment violations and illustrating how easily police could, after the fact, convince a court that a statement was freely offered, rather than extracted through a formal interrogation); *see also* Andrew C. Fine, *Refining* Crawford*: The Confrontation Clause After* Davis v. Washington *and* Hammon v. Indiana, 105 MICH. L. REV. FIRST IMPRESSIONS 11, 12 (2006), http://www.michiganlawreview.org/firstimpressions/vol105/find.pdf ("When determining the 'primary purpose' of questioning, it will be difficult for courts to ignore an officer's claim that he believed the emergency to be ongoing when he questioned the declarant," similar to the difficulty in "Fourth Amendment issues.").

255. *See* Friedman, *supra* note 139, at 563 (explaining how courts will look "for whatever toehold they can find to admit accusatory statements that were made absent an opportunity for confrontation.").

256. *Crawford*, 541 U.S. at 67.

However, rather than evaluating each and every potential solution on an individual basis, it is important to understand on a fundamental level what will *not* work and *why* it will not work. Only after this fundamental recognition can a sound and workable solution be developed.

A. Identifying Ineffective Solutions

Perhaps the most common proposal for interpreting the Clause, and more specifically for defining the term testimonial hearsay, is to suggest that courts focus on the objective intent of the declarant, rather than of the questioner or interrogator.[257] It is argued that this standard is less easily manipulated by police, prosecutors, and judges and is also a more accurate test of whether the declarant actually bore testimony, and thereby created testimonial hearsay.[258]

However, the problem with this approach, and with all of the imaginable variations of it, is that it does not even mask the problem, let alone resolve it. The real problem is not whether the courts should focus on the declarant or the questioner, assuming such a distinction offers even a theoretical benefit. The problem, rather, is the use of judicial discretion itself, combined with the potential for manipulation by police.[259] It is just as easy for a police officer to manipulate the objective intent of the declarant, who is necessarily absent from trial and cannot speak for himself or herself, than it is to misrepresent the officer's own intent.[260] For example, it *is* very easy for an officer to testify that he was personally concerned for the safety of the alleged victim or others, and therefore was asking questions to help him assess the situation and not to investigate a past

---

257. *See* O'Neil, *supra* note 241, at 545 ("For the sake of uniform application of a defendant's right to confrontation, courts should confine their constitutional analyses to the witness's perspective of the events at the time the statement was made."); Friedman, *supra* note 139, at 560 (arguing that the perspective of the declarant, and not the questioner, is the relevant focus); Tom Lininger, Davis *and* Hammon*: A Step Forward, or a Step Back?*, 105 MICH. L. REV. FIRST IMPRESSIONS 28, 29 (2006), http://www.michiganlawreview.org/firstimpressions/vol105/lininger.pdf (arguing that an objective test from the standpoint of the declarant is the relevant focus).

258. *See* O'Neil, *supra* note 241, at 547 (arguing that the focus should be on the "declarant's reasonable expectation, something the officers cannot control through their actions or observations."); Lininger, *supra* note 257, at 29 (arguing that an objective test, from the standpoint of the declarant, will "minimize the ability of police to manipulate this test").

259. Ross, *supra* note 109, at 205 ("One problem with focusing on how evidence is gathered . . . is that it permits manipulation by police and police agents.").

260. In fact, police and prosecutors are already well trained in manipulating the intent of the declarant; they do it often at preliminary hearings and even at trials to fit hearsay statements into the excited utterance exception to the hearsay rule. For example, in *State v. Searcy*, 709 N.W.2d 497, 502 (Wis. Ct. App. 2005), an officer testified that the declarant, at the time she made her statement, was "[u]m, rather excited." The court accepted this self-serving, conclusory testimony, and found that the statement was admissible under the excited utterance exception. *Id.*

crime.[261] Therefore, focusing on the interrogator's intent or expectations—whether subjectively or objectively—*is* rife with possibilities for manipulation.[262]

However, it is just as easy for a police officer to testify that the declarant, at the time he or she made the statement, appeared to be scared, spoke frantically, and in the present tense. The officer could also testify that, at the time of the statement, the defendant's location was unknown and the declarant acted as though he or she feared the assault may be renewed. To make matters even simpler, an officer could testify that the alleged victim, after giving the statement, asked the officer to "watch her when she left the apartment to make sure that she was not assaulted again."[263] Any of these simple tactics will effectively manipulate the objective intent of the absent declarant.[264] The statements will then be labeled as nontestimonial and admitted into evidence.[265]

Another excellent example of manipulating the objective intent of the declarant is *State v. Stahl*, discussed in Part IV.B., where the court found the declarant reasonably could have believed that the interview was conducted to obtain medical help, and not to assist in the prosecution of the crime.[266] This particular judicial finding did not even require any overt manipulation by the police, and was made despite the declarant's acknowledgment that her statements would be used "in the investigation and prosecution of this crime."[267]

These examples establish that the problem lies not in the particular focus of the judicial discretion, but rather in the judicial discretion itself. Any proposed solution that merely shifts the subject of the judge's discretion without constraining it will fail.[268] This assertion is true not only with regard to defining testimonial hearsay, but also with regard to applying the forfeiture doctrine. Solutions that continue to rely on judicial discretion, regardless of the label under which it is used, will be no more effective in fixing *Davis* than *Davis* was in fixing *Roberts*.

B. Putting "Confrontation" Back in the Clause

The solution to the problem is simple and can therefore be stated briefly: the Court must constrain, rather than merely shift, judicial discretion. With regard to the forfeiture doctrine, lower courts must *not* be permitted to find a defendant guilty of the underlying the crime with which he or she is charged—the ultimate

261. *See* Ross, *supra* note 109, at 183.

262. *See id.* (discussing how the "intent of the officer's rationale allows manipulation by police.").

263. State v. Washington, 725 N.W.2d 125, 132 (Minn. Ct. App. 2006).

264. *See* Ross, *supra* note 109, at 183.

265. *See id.* at 133.

266. *Stahl*, 855 N.E.2d at 846.

267. *Id.* at 837.

268. *See* Ross, *supra* note 109, at 206-07 (discussing how the proposals of two well-known professors fail to cure the problem because they do not eliminate the opportunity for manipulation).

exercise of discretion—and then use that pretrial finding of guilt to admit uncross-examined testimonial hearsay at trial.

Fortunately, the forfeiture doctrine need not be reinvented or even discussed at great length. Rather, the doctrine's pre-*Crawford* application, which was limited specifically to cases of witness tampering,[269] should be explicitly adopted by the Court.[270] This interpretation is not only true to the doctrine's constitutional origins, but it also serves to constrain judicial discretion and prevent the perverse "cart before the horse" logic of the lower courts' post-*Davis* holdings such as *State v. Jensen*.[271]

With regard to defining testimonial hearsay, its meaning must *not* depend on what the interrogator claims his or her primary purpose was in asking the questions, or what the court thinks the declarant's expectations might have been regarding how the statement might be used in the future. First, as Josephine Ross has stated, the right of confrontation is a trial right and does not depend on the means or techniques used in gathering the statement, but rather on how the statement is used at trial.[272]

Second, any definition of testimonial hearsay that hinges on theoretical subtleties—e.g., whether the focus should be on listener or speaker, or whether the test for intent should be objective or subjective—will be entirely ineffective, unworkable, and subject to manipulation.[273] As the cases discussed in this Article demonstrate, linguistic dances about distinctions without a practical difference provide busy work for commentators but are easily sidestepped by prosecutors and judges in the courtroom.

The concept of testimonial must therefore focus on the statement's use at trial. As Vincent Rust and I wrote before *Davis* was published:

> The term testimonial should be defined as all accusatory hearsay, i.e., hearsay that tends to establish in any way an element of the crime or the identification of the defendant. To adopt a narrower definition . . . would necessarily require a tremendous amount of judicial discretion under a facts-and-circumstances analysis. Although such an analysis would be under the heading of testimonial,

---

269. *See* Flanagan, *supra* note 209, at 864-65.

270. As noted earlier, after this Article was written, but before it was published, the United States Supreme Court decided *Giles v. California*, 128 S. Ct. 2678 (2008), where it struck down the expanded forfeiture doctrine. The Court held that "[t]he notion that judges may strip the defendant of a right that the Constitution deems essential to a fair trial, on the basis of a prior *judicial* assessment that the defendant is guilty as charged, does not sit well with the right to trial by jury." *Id.* at 2686 (emphasis in original).

271. State v. Jensen, 727 N.W.2d 518, 546 (Wis. 2007) (Butler, Jr., J., dissenting in part).

272. Ross, *supra* note 109, at 196-97 (arguing that the right of confrontation is a trial right, and the focus should *not* be on how the statements were gathered, but rather on "how the out-of-court words are being used in the particular trial").

273. *See id.* at 171-72 (arguing that under fact-intensive tests that focus on the method of creating or collecting the hearsay, "[p]olice and prosecutors will be encouraged to alter their methods of gathering evidence and describing the investigation in such a way that statements will be deemed responses to emergency situations . . .").

rather than reliability, the end result would be the same: judges would still be deciding which hearsay is admissible.[274]

Most significantly, this proposed definition would prevent the admission of core testimonial hearsay, which is "the principal evil at which the Confrontation Clause was directed."[275] Because the focus would be on how the statement is used at trial rather than how it was obtained, there would no longer be any incentive for government manipulation. For example, the police would no longer have an incentive to channel statements through surrogates, rather than conduct the investigations themselves, in hopes of circumventing the defendant's confrontation rights should the declarant later become unavailable.[276] The pretrial manipulation and gamesmanship created by the current definition of testimonial would simply become irrelevant.

Realistically, however, it took the Court twenty four years to change course from *Roberts*, and the Court is now only four years into its new course with *Crawford*. If recent history is any indication, the Court will not make such firm and sweeping changes anytime soon. Additionally, the Court's trepidation in deciding even basic issues—e.g., whether 911 operators are agents of the police[277]—makes it even less likely that swift change is anywhere on the near horizon. Therefore, if reform is going to take place it may have to come from the state courts.[278]

However, reform by the Court is not completely without hope. The Court has shifted to bright line, workable rules in other areas of constitutional jurisprudence, such as in search and seizure cases. In *Chimel v. California*,[279] for example, the

---

274. Cicchini & Rust, *supra* note 109, at 543-44; *see also* Ross, *supra* note 109, at 196 (arguing similarly that "the term testimonial should apply to all statements repeated at court that are accusatory in the context of the criminal trial, that are introduced for the truth of their assertion, and where the reliability of the declarant could affect the truth of the charges in that particular case"). Admittedly, while it is possible to constrain judicial discretion, it is never possible to eliminate it entirely. For example, using Ross' definition, a court could simply find that the statement was being offered for a purpose *other than* the truth of the matter asserted, and could attempt to bypass the Clause in this manner. However, there is a much greater body of case law defining what constitutes the "truth of the matter asserted" than there is defining what constitutes an "ongoing emergency," for example. Therefore, long-standing and well-established legal precedent would severely temper judicial manipulation in such instances.

275. *Crawford*, 541 U.S. at 50.

276. *See* discussion *supra* Part IV.B.

277. *See* Davis v. Washington, 547 U.S. 813, 823 n.2 (stating that "[f]or purposes of this opinion (and without decided the point), we consider [911 operators'] acts to be acts of the police").

278. *See, e.g.*, State v. Knapp, 700 N.W.2d 899, 914 (Wis. 2005) (citing State v. Doe, 254 N.W.2d 210, 216 (Wis. 1977)) (holding that "[t]his court 'will not be bound by the minimums which are imposed by the Supreme Court of the United States if it is the judgment of this court that the Constitution of Wisconsin and the laws of this state require that greater protection of citizens' liberties ought to be afforded.'")

279. Chimel v. California, 395 U.S. 752 (1969).

Court held that police are allowed to search all areas "within [the arrestee's] immediate control" in order to prevent an arrestee from destroying potential evidence or using a weapon.[280] This was a fact-intensive test, much like the test in *Roberts* and *Davis*, and depended on the facts surrounding the arrest, including the distance between the arrestee and the area being searched.[281]

In *New York v. Belton*,[282] however, the Court changed course and abandoned its facts-and-circumstances analysis to adopt a new bright-line rule.[283] The new rule allowed police to search an arrestee's automobile under all circumstances incident to arrest.[284] The Wisconsin Supreme Court applied this rule and upheld a police search of an arrestee's vehicle even when he was arrested outside of and away from his automobile, was handcuffed, secured in a squad car, and guarded by officers.[285] Clearly, if a person was not near his or her automobile when arrested and then was searched, cuffed, and locked in a guarded squad car, there is *no possibility* that the person will obtain a weapon from, or destroy potential evidence in, his automobile.

Nonetheless, the United States Supreme Court adopted this bright line rule to give police the right to search an automobile in all cases incident to arrest, even where there was no real or imaginary risk of the arrestee destroying evidence or obtaining a weapon. The Court, as well as the lower courts, found the previous, fact-intensive analysis "unworkable" because it was supposedly too difficult to determine when an arrestee could actually gain access to his automobile.[286] Therefore, the police are no longer required to *prove* or even *believe* that the automobile was accessible; rather, they may simply *assume* accessibility, even in cases where accessibility is literally impossible.[287]

This example proves that the Court is quite capable of drawing bright line rules. Admittedly, this particular bright line rule was drawn for the benefit of police rather than defendants; it would be somewhat naive to assume that the Court would employ the same underlying reasoning in cases where the rule would benefit the citizenry, rather than the government. Nonetheless, if the Court is willing to adopt such rules as a matter of police convenience in search and seizure law, it is not too much to demand that the Court do so to protect fundamental constitutional rights. Therefore, a similar bright line rule, focusing on the statement's use at trial rather than the method in which it was obtained, should be the hallmark of Confrontation Clause jurisprudence.

---

280. *Id.* at 762-63 (1969).

281. *Id.*

282. New York v. Belton, 453 U.S. 454, 460 (1981).

283. *Id.*

284. *Id.*

285. *See* State v. Fry, 388 N.W.2d 565, 565-75 (Wis. 1986).

286. *Id.* at 574. ("The only other alternative to the *Belton* rule would be to permit searches on a case-by-case basis when the police believe that a suspect may escape from their control and regain access to an automobile. This alternative is unworkable, however, because such momentary escapes are not predictable.").

287. *Id.*

VII. CONCLUSION

In *Crawford* and *Davis* the Court finally acknowledged the dangers and unconstitutionality of allowing judges to substitute their discretion in place of actual confrontation.[288] Unfortunately, however, the Court's new framework—at first thought by many to be a "sea change" in Confrontation Clause jurisprudence[289]—has done little to constrain the very judicial discretion that the Court condemned.[290] Instead, the framework merely shifted the judicial discretion from one issue—whether the hearsay was reliable—to other issues, such as whether the hearsay is testimonial and, if so, whether the defendant forfeited his right of confrontation.[291]

The Court's lack of resolve has left the lower courts with as much or more discretion in admitting untested hearsay evidence as they had under the old *Roberts* reliability test. Consequently, lower courts have easily circumvented *Crawford* and *Davis* and have done so in numerous, creative ways.[292] Most commonly, courts have expanded the scope of the ongoing emergency[293] and distorted the primary purpose test,[294] both of which result in labeling hearsay as nontestimonial, thus causing it to fall outside the scope of the Clause altogether.[295] Even in cases of testimonial hearsay, courts have simply expanded the scope of the forfeiture doctrine in order to make a finding that the defendant forfeited his right of confrontation.[296]

Regardless of the tactic employed, however, the end result is just as it was under the old *Roberts* test: judges, with the help of police and other government agents, are easily finding ways to admit untested and uncross-examined hearsay against defendants. Any proposed solution that attempts to cure these judicial abuses by merely shifting judicial discretion from one issue to another will be no more effective in fixing *Davis* than *Davis* was in fixing *Roberts*.[297]

The only viable solution to the problem is to constrain judicial discretion. With regard to the forfeiture doctrine, its reach must be limited to cases of alleged witness tampering.[298] With regard to testimonial hearsay, the focus must be on the hearsay's use at trial, rather than the manner in which it was given by the

---

288. *See* discussion *supra* Part III.

289. *See* State v. Grace, 111 P.3d 28, 36 (Haw. Ct. App. 2005); Hutton, *supra* note 90, at 61; King-Ries, *supra* note 90, at 313.

290. *See* discussion *supra* Part V.

291. *See id.*

292. *See* discussion *supra* Part IV.

293. *See* State v. Washington, 725 N.W.2d 125 (Minn. Ct. App. 2006); State v. Warsame, 723 N.W.2d 637 (Minn. Ct. App. 2006); Harkins v. State, 143 P.3d 706 (Nev. 2006); State v. Camarena, 145 P.3d 267 (Or. Ct. App. 2006); Vinson v. State, 221 S.W.3d 256 (Tex. App. 2006); State v. Rodriguez, 722 N.W.2d 136 (Wis. Ct. App. 2006).

294. *See* State v. Stahl, 855 N.E.2d 834 (Ohio 2006).

295. *See* discussion *supra* Parts IV.A.-B.

296. *See* discussion *supra* Part IV.C.

297. *See* discussion *supra* Part VI.A.

298. *See* discussion *supra* Part VI.B.

declarant or obtained by the police.[299] This bright line, trial-based rule is not only mandated by the plain language of the Confrontation Clause, but it would also be consistent with the Court's reasoning and holdings in other areas of constitutional jurisprudence.[300] Only this approach will constrain judicial discretion and ensure the constitutional right of confrontation.

299. *See id.*
300. *See id.*

THE NEW DOCTRINALISM IN CONSTITUTIONAL SCHOLARSHIP AND *DISTRICT OF COLUMBIA V. HELLER*

BRANNON P. DENNING*

In the run-up to oral arguments in *District of Columbia v. Heller*,[1] the Bush Administration outraged many gun rights advocates when the Solicitor General filed the Government's *amicus* brief.[2] Though it endorsed an individual right interpretation, the Administration's brief urged the U.S. Supreme Court to overturn the D.C. Circuit's decision, which invalidated the strict ban on handguns,[3] and remand for reconsideration under a more lenient standard of review.[4] Gun rights supporters viewed the Government's position as a betrayal, in part, because of a sense that where an individual right is recognized, nothing less than "strict scrutiny"—requiring the government to demonstrate that a law is "narrowly tailored to achieve a compelling governmental interest"[5]—follows from that recognition.[6] At oral argument,[7] and in the opinion itself,[8] there was considerable discussion of the proper standard of review.

---

* Professor of Law and Director of Faculty Development, Cumberland School of Law, Samford University. Thanks to Richard Fallon, Glenn Reynolds, Kim Roosevelt, Ted Ruger, Miguel Schor, and Norman Williams for their excellent suggestions.

1. 128 S. Ct. 2783 (2008).

2. *See, e.g.*, *Government files amicus -- on DC's Side!*, Of Arms & the Law, http://armsandthelaw.com/archives/2008/01/government_file_1.php (Jan. 11, 2008, 20:41 EST).

3. Parker v. District of Columbia, 478 F.3d 370 (D.C. Cir. 2007), *aff'd sub nom.* District of Columbia v. Heller, 128 S. Ct. 2783 (2008).

4. For a summary of the Bush Administration's argument, see Brief for the United States as Amicus Curiae at 8, *Heller*, 128 S. Ct. 2783 (Jan. 11, 2008) (No. 07-209), 2008 WL 15701:

> [T]he Second Amendment, properly construed, allows for reasonable regulation of firearms, must be interpreted in light of context and history, and is subject to important exceptions .... Nothing in the Second Amendment properly understood—and certainly no principle necessary to decide this case—calls for invalidation of the numerous federal laws regulating firearms. When ... a law directly limits the private possession of 'Arms' in a way that has no grounding in Framing-era practice, the Second Amendment requires that the law be subject to heightened scrutiny that considers (a) the practical impact of the challenged restriction on the plaintiff's ability to possess firearms for lawful purposes (which depends in turn on the nature and functional adequacy of available alternatives), and (b) the strength of the government's interest in enforcement of the relevant restriction.

5. *See Heller*, 128 S. Ct. at 2851 (Breyer, J., dissenting) (quoting Abrams v. Johnson, 521 U.S. 74, 82 (1997)).

6. *See, e.g.*, Charles Bloomer, *The Second Amendment a Second Class Right?*, ENTER STAGE RIGHT, Jan. 21, 2008, http://www.enterstageright.com/archive/articles/0108/0108secamd.htm ("The argument the Solicitor General makes that *Heller vs. DC* should only be

The controversy over the proper standard of review in *Heller* coincides with a renewed interest among scholars in the formation and application of doctrine.[9] What I am calling the "New Doctrinalism" in constitutional scholarship focuses less on controversies over the fixing of constitutional meaning and more on the rules courts develop to "implement," as Richard Fallon described it, those constitutional commands.[10] In this brief essay, I describe the New Doctrinalism and explain the potential payoff for constitutional law using *Heller* as a case study.

DEFINING THE NEW DOCTRINALISM

With much of the focus on empiricism and interdisciplinarity,[11] to suggest that doctrinal scholarship, especially in constitutional law, is worthy of study requires some explanation and a defense. The scholars engaged in the New Doctrinalism are not simply describing, with critical analysis, cases that make up particular areas of constitutional law. Rather, the New Doctrinalism's focus is on macro-level doctrinal formation—how the Court takes general propositions about constitutional meaning and constructs rules that enable it to decide particular cases.

In his 2001 book, for example, Richard Fallon states that in most cases, the Court does not choose between originalist and nonoriginalist interpretive methods; rather, it applies rules developed in prior cases or adjusts those rules to

---

afforded 'intermediate scrutiny' is pure governmental arrogance. Anytime *any* law is considered that potentially restricts *any* of our Constitutionally guaranteed rights, whether in the legislative process or in judicial review, the principle of 'strict scrutiny' should apply.").

7. Transcript of Oral Argument at 18, 22, 35, 40, 41–43, 45, 48, 53, 66, 72, 74, *Heller*, 128 S. Ct. 2783 (No. 07-290), http://www.supremecourtus.gov/oral_arguments/argument_transcripts/07-290.pdf.

8. *Heller*, 128 S. Ct. at 2817–18, 2817 n.27, 2821; *id.* at 2850–53 (Breyer, J., dissenting).

9. *See, e.g.*, RICHARD H. FALLON, JR., IMPLEMENTING THE CONSTITUTION (2001).

10. *Id.* at 5 ("The term *implementation* invites recognition that the function of putting the Constitution effectively into practice is a necessarily collaborative one, which often requires compromise and accommodation. It also emphasizes the practical, frequently strategic aspects of the Court's work [including] the formulation of constitutional rules, formulas, and tests . . . .").

11. For a current instantiation of this debate, see *Interdisciplinarity, Multidisciplinarity and the Future of the Legal Academy*, Legal Theory Blog, http://lsolum.typepad.com/legaltheory/2008/01/interdisciplina.html (Jan. 18, 2008, 15:07 EST).

take account of new circumstances.[12] Professor Fallon's book provides a veritable safari of the types of doctrinal tests employed in constitutional cases.[13]

Fallon's work was followed by a number of articles that developed and extended his insights. In 2004, for example, Mitchell Berman's *Constitutional Decision Rules* urged scholars to reconceive constitutional interpretation as a two-stage process.[14] In stage one, judges fix constitutional meaning, using any one of a number of interpretive methodologies recognized as legitimate.[15] Then, once the "constitutional operative proposition" is established, "decision rules" are generated; these rules form the backbone of doctrine that permits the implementation of the operative proposition.[16]

Building on Berman's work, Kermit Roosevelt discussed "constitutional calcification," which explored how doctrine degrades over time, requiring adjustments or, in some cases, radical revision as doctrinal rules become unsuitable to implement constitutional commands.[17] Occasionally, Roosevelt noted, the rules and the operative propositions are conflated, with the result that rules are taken as constitutional commands.[18] Thus, he argues, does the Constitution meld with the Court's gloss on it.[19] Careful separation of operative propositions and rules, he argues further, can sensitize scholars and judges to the danger of conflating the Constitution with doctrine and prevent internalization of constitutional "commands" that are no more than tools used by the judiciary to decide cases.[20]

Other scholars are now focusing on particular doctrinal areas. For example, Richard Fallon and Adam Winkler have taken a close look at "strict scrutiny."[21]

---

12. FALLON, *supra* note 9, at 5 ("Especially in formulating tests such as these, the Court does not characteristically engage in historical or moral philosophical analysis, nor does it attempt to determine whether particular events in the world come within the semantic meaning of a constitutional norm. Rather, the Court devises and then implements strategies for enforcing constitutional values.").

13. *Id.* at 76–101.

14. Mitchell N. Berman, *Constitutional Decision Rules*, 90 VA. L. REV. 1, 15 (2004).

15. *Id.* at 9.

16. *Id.* at 13 ("I will argue that judges, scholars, and litigators should make greater efforts to distinguish whether a constitutional rule is an announcement of constitutional meaning (i.e., a constitutional operative proposition) or, instead, is a constitutional decision rule, and should pay attention, in the making of constitutional decision rules, to the particular considerations that might justify its construction.").

17. Kermit Roosevelt III, *Constitutional Calcification: How the Law Becomes What the Court Does*, 91 VA. L. REV. 1649, 1657–58, 1686–89 (2005).

18. *Id.* at 1652 ("In a striking number of cases the Court has forgotten the reasons behind particular rules and has come to treat them as nothing more than statements of constitutional requirements.").

19. *Id.* ("This mistaken equation of judicial doctrine and constitutional command tends to warp doctrine, frequently at significant cost to constitutional values; it also distorts the relationship the Court has to other governmental actors and to the American people.").

20. *Id.* at 1686–1719.

21. Richard H. Fallon, Jr., *Strict Judicial Scrutiny*, 54 UCLA L. REV. 1267 (2007)

They have discovered what closer observers of the Court have noticed in opinions: that there are one, two, many forms of strict scrutiny.[22] And not all are "fatal in fact," as Gerry Gunther famously put it.[23] Winkler has also done a study demonstrating further that finding a "fundamental right" or an "individual right" does not *ipso facto* result in a court applying strict scrutiny.[24] Calvin Massey has suggested that the Court's haphazardly-applied equal protection standards of review have made doctrine in that area unstable,[25] and he recently examined the role that inquiries into governmental purpose play in constitutional law.[26]

Other contributions to doctrinal scholarship include Dan Coenen's examination of what he terms "subconstitutional rules" that promote interbranch dialogue.[27] Paul Horwitz has examined judicial deference in a variety of constitutional contexts.[28] David Strauss has forcefully argued that doctrinalism has normative advantages over other forms of constitutional interpretation.[29] And, in G. Edward White, doctrine has found its own historian.[30] This is, of course, an incomplete list:[31] Akhil Amar,[32] Mike Dorf,[33] Charles Fried,[34] Elena

---

[hereinafter Fallon, *Strict Scrutiny*]; Adam Winkler, *Fatal in Theory and Strict in Fact: An Empirical Analysis of Strict Scrutiny in the Federal Courts*, 59 VAND. L. REV. 793 (2006); *see also* Richard H. Fallon, Jr., *Judicially Manageable Standards and Constitutional Meaning*, 119 HARV. L. REV. 1274 (2006) [hereinafter Fallon, *Judicially Manageable Standards*]; Richard H. Fallon, Jr., *The Linkage Between Justiciability and Remedies—and Their Connections to Substantive Rights*, 92 VA. L. REV. 633 (2006) [hereinafter Fallon, *Justiciability*].

22. Fallon, *Strict Scrutiny*, *supra* note 21, at 1302–15; Winkler, *supra* note 21, at 828–69.

23. Gerald Gunther, *The Supreme Court, 1971 Term—Foreword: In Search of Evolving Doctrine on a Changing Court: A Model for a Newer Equal Protection*, 86 HARV. L. REV. 1, 8 (1972).

24. Adam Winkler, *Fundamentally Wrong About Fundamental Rights*, 23 CONST. COMMENT. 227, 227–28 (2006).

25. Calvin Massey, *The New Formalism: Requiem for Tiered Scrutiny?*, 6 U. PA. J. CONST. L. 945 (2004).

26. Calvin Massey, *The Role of Governmental Purpose in Constitutional Judicial Review*, 59 S.C. L. REV. 1 (2007).

27. Dan T. Coenen, *The Rehnquist Court, Structural Due Process, and Semisubstantive Constitutional Review*, 75 S. CAL. L. REV. 1281, 1336 n.301 (2002); Dan T. Coenen, *A Constitution of Collaboration: Protecting Fundamental Values with Second-Look Rules of Interbranch Dialogue*, 42 WM. & MARY L. REV. 1575 (2001).

28. Paul Horwitz, *Three Faces of Deference*, 83 NOTRE DAME L. REV. 1061 (2008).

29. David A. Strauss, *Common Law Constitutional Interpretation*, 63 U. CHI. L. REV. 877, 879 (1996) ("The common law approach restrains judges more effectively, is more justifiable in abstract terms than textualism or originalism, and provides a far better account of our practices."); *see also* David A. Strauss, *Common Law, Common Ground, and Jefferson's Principle*, 112 YALE L.J. 1717 (2003); David A. Strauss, *The Common Law Genius of the Warren Court* (Univ. of Chi. Law Sch., Public Law and Legal Theory Working Paper No. 25, 2002), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=315682.

30. G. Edward White, *Historicizing Judicial Scrutiny*, 57 S.C. L. REV. 1 (2005).

31. My intent is to be illustrative, not comprehensive. My apologies to any other New Doctrinalists or fellow travelers not listed here. For my own contribution, see Brannon P. Denning, *Reconstructing the Dormant Commerce Clause Doctrine*, 50 WM. & MARY L. REV. 417 (2008).

Kagan,[35] and Kathleen Sullivan[36] have also contributed to our understanding of constitutional doctrine in particular areas. And if the New Doctrinalism's practitioners constitute a veritable pantheon of constitutional law's leading scholars, then Laurence Tribe looms, Zeus-like, over them all![37]

THE CASE FOR THE NEW DOCTRINALISM—AND THE CASE AGAINST

This renewed interest in doctrine should be welcome for a number of reasons. First, the New Doctrinalism provides a focal point for discussions of the Court's work product by scholars with different views on the proper method for interpreting the Constitution. It provides a mechanism for implementing what Cass Sunstein described as "incompletely theorized agreements" about meaning,[38] and a way for those agnostic on the originalism vs. nonoriginalism debate, or those uninterested in arid historical arguments about constitutional meaning, to discuss the Court and its work.

For example, suppose two scholars agree the Commerce Clause not only empowers Congress to act, but also implicitly limits the ability of states to discriminate against interstate commerce. They *disagree*, however, on the reason for why that implicit limit exists. Scholar A might think that it exists because the Framers intended the Commerce Clause to limit state power in that way; Scholar B, though, might simply regard anti-discrimination as a necessary check on parochialism that maximizes utility for all states and for the country as a whole.

---

32. Akhil Reed Amar, *The Supreme Court, 1999 Term—Foreword: The Document and the Doctrine*, 114 HARV. L. REV. 26 (2000).

33. Michael C. Dorf, *Incidental Burdens on Fundamental Rights*, 109 HARV. L. REV. 1175 (1996).

34. Charles Fried, *Constitutional Doctrine*, 107 HARV. L. REV. 1140 (1994); Charles Fried, *Types*, 14 CONST. COMMENT. 55 (1997).

35. *See, e.g.*, Elena Kagan, *Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine*, 63 U. CHI. L. REV. 413 (1996).

36. *See, e.g.*, Kathleen M. Sullivan, *Unconstitutional Conditions*, 102 HARV. L. REV. 1413 (1989).

37. 1 LAURENCE H. TRIBE, AMERICAN CONSTITUTIONAL LAW (3d ed. 2000). Dean Sager's work on underenforcement deserves mention as an Ur-text of the New Doctrinalism. *See* Lawrence Gene Sager, *Fair Measure: The Legal Status of Underenforced Constitutional Norms*, 81 HARV. L. REV. 1212 (1978).

38. Cass R. Sunstein, *Incompletely Theorized Agreements in Constitutional Law*, SOC. RES. (forthcoming) (Univ. of Chi. Law Sch., John M. Olin Law & Economics Working Paper No. 322, Public Law and Legal Theory Working Paper No. 147, 2007), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=957369. Sunstein's minimalism project is at least a cousin of the New Doctrinalism, insofar as it seeks to confine disagreements about ultimate meaning and resolve cases narrowly, using doctrinal rules that produce narrow or "minimalist" opinions that leave room for the democratic resolution of first-order issues. *See* CASS R. SUNSTEIN, ONE CASE AT A TIME: JUDICIAL MINIMALISM ON THE SUPREME COURT (1999).

The New Doctrinalism offers a way for both scholars to focus on what form the decision rules implementing the anti-discrimination principle should take. If nothing else, recent doctrinal scholarship advances constitutional law beyond the stalemated (and stale) debates between originalists and nonoriginalists that have dominated constitutional theory for more than a generation.

Second, a renewed focus on doctrine also harkens back to the initial project of the Legal Realists—to test "paper" rules against the "real" rules courts applied to decide cases, and then adjust the former accordingly if there is a gap between the two.[39] The Legal Realists were primarily concerned with private law subjects; but just as Herman Oliphant, in the words of his famous essay, hoped close study could enable "a return to *stare decisis*," the same might be true of a close study of constitutional doctrine as well.[40] Any meaningful normative critique of the Court's work must depend on an accurate descriptive account of what, in fact, the Court does when it decides cases. And describing precisely what the Court does with its decision rules has already yielded some surprises, e.g., that strict scrutiny can take different forms in different situation types.[41]

Third, good doctrinal scholarship also represents a form of empiricism, which is a refreshing change of pace from some of the constitutional theory that, in the past, was so abstract as to border on the ethereal. John Hart Ely once quipped that such scholarship seemed animated by the expectation of reading a Court opinion holding, "We like Rawls, you like Nozick. We win, 6-3."[42] The empiricism may be a little soft, but it's a start.

Fourth, by separating the oft-conflated fixing of meaning and design of standards of review, the New Doctrinalism offers an opportunity to reflect on the choices open to judges when creating rules of decision and to articulate and debate the proper criteria for such rules. These criteria, not surprisingly, figure prominently in the recent writings.[43] Debates over doctrinal rules, moreover, can

---

39. *See* Karl N. Llewellyn, *A Realistic Jurisprudence—The Next Step*, 30 COLUM. L. REV. 431, 447–54 (1930) (describing and distinguishing "paper" rules and "real" rules and explaining their application to doctrine); G. Edward White, *From Sociological Jurisprudence to Realism: Jurisprudence and Social Change in Early Twentieth-Century America*, 58 VA. L. REV. 999 (1972) (describing realism in legal thought and the shift to realism from sociological jurisprudence in the early twentieth-century).

40. Herman Oliphant, *A Return to Stare Decisis*, 14 A.B.A. J. 71 (1928).

41. *See supra* note 21.

42. JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW 58 (1980).

43. *See, e.g.*, Berman, *supra* note 14, at 93–96 (discussing "six analytically distinct factors or families of factors that might appeal to a judge considering whether, and how, to form a constitutional decision rule—considerations I label *adjudicatory, deterrent, protective, fiscal, institutional,* and *substantive*"); Roosevelt, *supra* note 17, at 1658–67 (discussing "institutional competence," "costs of error," "frequency of unconstitutional action," "legislative pathologies," "enforcement costs," and "guidance for other governmental actors" as factors in the formation of decision rules).

furnish a convenient platform for consideration of larger questions of institutional competence and judicial review that are too often posed in the abstract.[44]

Fifth, because the New Doctrinalism focuses on the work product of the courts and the tools judges fashion to dispose of constitutional cases, it carries with it the possibility of bridging the much-discussed gap between judges and lawyers on the one hand, and academics on the other.[45] It might even create a useful feedback mechanism between the Supreme Court and lower courts. Doctrinal scholarship's focus on the craft of rule formation can sensitize lower courts to the scope and content of Supreme Court doctrine, highlight choices made or avoided by the doctrine, and alert judges to gaps in the doctrine.

Further, if the New Doctrinalism is extended, as it should be, to the lower courts, then scholars can furnish evidence of the lower courts' application (or lack of application) of decision rules, highlight doctrinal areas in need of clarification, and identify rules that require alteration. It might also alert the Supreme Court to areas in which lower courts are declining to implement the Court's own doctrine.[46] This lower court resistance might be caused in part by a shrinking case load that reduces the chance lower courts will be called to account for ignoring the Court.[47] Doctrinal scholarship could also assist practitioners in framing arguments to courts. Exposing practitioners to the different standards of review that can arise in various situation types might enable them to frame questions and tests more precisely.

The New Doctrinalism may even have some payoff for *teachers* of constitutional law. Stressing the Court's role of creating—and lower courts' roles in applying—decision rules can alert students to the discretion that Justices and judges have in enforcing constitutional commands. Examining lower court cases shows that much often remains after "blockbuster" Supreme Court cases, and that fudging the standard of review can create confusion in the lower courts. Stressing decision rules also furnishes a way to introduce topics like institutional competence and requires students to think critically about the factors that do (or should) influence a court's choice in doctrine. Moreover, such skills are transferable and offer a response to the student who asks (as they often do) whether the course has any connection to their future practice.

---

44. For discussions of institutional competence in constitutional adjudication, see ADRIAN VERMEULE, JUDGING UNDER UNCERTAINTY: AN INSTITUTIONAL THEORY OF LEGAL INTERPRETATION (2006), and Fallon, *Judicially Manageable Standards*, *supra* note 21, at 1280–96, 1309–12.

45. Harry T. Edwards, *The Growing Disjunction Between Legal Education and the Legal Profession*, 91 MICH. L. REV. 34 (1992).

46. Glenn Reynolds and I tried to do something like this with *Lopez* and *Morrison*. *See* Brannon P. Denning & Glenn H. Reynolds, *Rulings and Resistance: The New Commerce Clause Jurisprudence Encounters the Lower Courts*, 55 ARK. L. REV. 1253 (2003) [hereinafter Denning & Reynolds, *Rulings and Resistance*]; Glenn H. Reynolds & Brannon P. Denning, *Lower Court Readings of* Lopez, *or What if the Supreme Court Held a Constitutional Revolution and Nobody Came?*, 2000 WIS. L. REV. 369 (2000). We plan to undertake a similar project with *Heller*.

47. *See* Denning & Reynolds, *Rulings and Resistance*, *supra* note 46, at 1308.

Finally, if we are ever to create space for members of the elected branches (and ordinary voters) to make constitutional arguments, it is important that no one feel obligated to imitate Court-speak.[48] Attention to the values of the "thin Constitution" should not depend on the ability to recite and apply all of the Court's various doctrinal tests that thicken it.[49] "Judicial overhang" can stunt consideration of constitutional questions by legislatures and stymie the creation of "populist constitutional law."[50] Separation of the "doctrine" from "the document," as Akhil Amar once put it,[51] can make clear that not all constitutional conversations must occur in a judicial voice.

Of course, what would thesis be without antithesis? The New Doctrinalism is already producing critical literature. Responding to one of Richard Fallon's papers, Roderick Hills argued that meaning and doctrine are not easily separated and that New Doctrinalist efforts to convince us otherwise are "seductively misleading" because they represent another attempt to achieve "noninstrumental certainty in the law."[52] For Hills, "the meaning of a constitutional provision *is* in its implementation," and to pretend otherwise is to "talk in empty, metaphysical abstractions."[53] Daryl Levinson similarly argues against a "rights essentialism" that "assumes a process of constitutional adjudication that begins with judicial identification of a pure constitutional value," which is "corrupted by being forced into a remedial apparatus that translates the right into an operational rule applied to the facts of the real world."[54]

Levinson wrote his article before the latest wave of New Doctrinalist scholarship, which avoids the "rights essentialism" trap because it does not view rule formation or implementation as "corrupting" the constitutional commands.[55] Both the fixing of meaning and the formation of implementing rules are seen as essential parts of adjudication.

As for Hills's critique, I don't think Fallon or anyone else views operative propositions and decision rules as wholly unrelated. The rules shape the judicially-enforced content of the operative proposition, just as the presence or absence of legal remedies shapes the perception of a "right." And yet it also seems to me that that assessing meaning—say, from an historical perspective—is

---

48. *See* MARK TUSHNET, TAKING THE CONSTITUTION AWAY FROM THE COURTS 194 (1999).

49. For the contrast between the "thin" and "thick" Constitutions, see *id.* at 9–14.

50. *Id.* at 57–65 (discussing ways in which "judicial overhang" inhibits legislative consideration of constitutional values).

51. Amar, *supra* note 32, at 26.

52. Roderick M. Hills, Jr., *The Pragmatist's View of Constitutional Implementation and Constitutional Meaning*, 119 HARV. L. REV. F. 173, 174 (2006) (responding to Fallon, *Judicially Manageable Standards*, *supra* note 21).

53. *Id.* at 175.

54. Daryl J. Levinson, *Rights Essentialism and Remedial Equilibration*, 99 COLUM. L. REV. 857, 858 (1999).

55. *See* Roosevelt, *supra* note 17, at 1711–12. Roosevelt does point out that meaning can be "warped" by a conflation of rules and operative propositions—by mistaking the rules themselves for the commands. *Id.*

a different enterprise than deciding how a judge should deploy that meaning in concrete cases—even if the rules necessarily shape the nature of that *judicial* meaning.

At the very least, the extent to which meaning and doctrine are separable or inextricably intertwined strikes me as a debate worth having. Simply by provoking a critique, the New Doctrinalism has opened up a new discussion in constitutional scholarship, which many think has grown rather stale over the years.

HELLER AND THE NEW DOCTRINALISM

Which brings us back to *Heller*. For years, the debate over the Second Amendment rarely progressed beyond the historical: did the Framers intend the Amendment to guarantee an individual right or a "collective" or "state's right"?[56] Since the courts were disinclined to engage the constitutional dimension of gun control, this debate was largely academic; thus, no one gave much thought to the proper standard of review.[57] To the extent anyone considered it, there seemed to be an assumption that "strict scrutiny" would follow from the individual rights reading.[58] Because it was widely assumed that "strict" scrutiny was well-nigh fatal to any law to which it was applied,[59] gun control proponents and opponents both had a stake in settling the individual/collective rights question.

At first, the initial posture of the debate over the Amendment seems to confirm Hills' and Levinson's argument that constitutional meaning is determined by rules, rather than the other way around. But as New Doctrinalist scholarship has shown, strict scrutiny is *not* necessarily the default standard of review for provisions of the Bill of Rights that indubitably guarantee "individual" rights. [60] Moreover, as Adam Winkler argued, in states where individual gun rights are clearly established in state constitutions, strict scrutiny has *not* followed as a matter of course.[61] In addition, studies have shown that several different tests travel under the label "strict scrutiny," but courts do not apply them in the same way.[62] Winkler's argument was that one could support the individual rights interpretation of the Second Amendment without dooming

---

56. For an excellent summary of the debate, circa the mid-1990s, when the debate heated up again, see Glenn Harlan Reynolds, *A Critical Guide to the Second Amendment*, 62 TENN. L. REV. 461 (1995).

57. *But see, e.g.*, Nelson Lund, *The Past and the Future of the Individual's Right to Arms*, 31 GA. L. REV. 1, 67–75 (1996).

58. *See id.* at 72.

59. *See* Gunther, *supra* note 23.

60. *See* Winkler, *supra* note 24.

61. Adam Winkler, *Scrutinizing the Second Amendment*, 105 MICH. L. REV. 683, 715–26 (2007).

62. Fallon, *Strict Scrutiny*, *supra* note 21, at 1302–15; Winkler, *supra* note 21, at 828–69.

"reasonable" gun control legislation.[63] The Bush Justice Department agreed in part, arguing for something less than strict scrutiny to apply to federal gun laws[64]—to the dismay of many gun rights supporters.[65]

The first draft of this Essay proceeded on the assumption that the majority was likely to be explicit about the standard of review that was or ought to have been applied to the District's ban on handguns, and either affirm or reverse on that basis. My first clue that something else might occur was during oral argument when Chief Justice Roberts asked whether prescribing a standard of review was really necessary.[66] Apparently the Chief Justice convinced the majority, because its opinion was not explicit about the standard that it used to evaluate the District's gun ban or about the standard that the lower courts ought to employ in future cases.[67] That said, I think that one might be inferred from hints in the opinion. Consider the following:

First, perhaps obviously, the Court affirmed the lower court,[68] and did not—as the Government had requested[69]—remand for reconsideration under a different standard of review.[70]

The majority rejected the use of any rational basis standard of review, calling it inappropriate for a "specific, enumerated right" like the Second Amendment.[71] "If all that was required," Justice Scalia wrote, "to overcome the right to keep and bear arms was a rational basis, the Second Amendment would be redundant with the separate constitutional prohibitions on irrational laws, and would have no effect."[72]The majority also rejected the balancing approach of Justice Breyer, who discussed the appropriate standard of review extensively.[73] The majority wrote:

> We know of no other enumerated constitutional right whose core protection has been subject to a freestanding 'interest-balancing' approach. The very enumeration of the right takes out of the hands of government—even the Third Branch of Government—the power to decide on a case-by-case basis whether the right is really worth insisting upon.[74]

---

63. Winkler, *supra* note 61, at 732.

64. *See* Brief of the United States, *supra* note 4, at 8. The Solicitor General's brief was apparently geared toward ensuring that nothing in *Heller* endanger the federal ban on machine guns. *See, e.g.*, Michael P. O'Shea, *The Right to Defensive Arms after* D.C. v. Heller, W. VA. L. REV. (forthcoming 2008).

65. *See supra* note 2.

66. Transcript of Oral Argument, *supra* note 7, at 44 (asking "why in this case we have to articulate an all-encompassing standard").

67. *See* District of Columbia v. Heller, 128 S. Ct. 2783, 2821–22 (2008).

68. *Id.*

69. *See* Brief for the United States, *supra* note 4, at 9–10.

70. *See Heller*, 128 S. Ct. at 2821–22.

71. *Id.* at 2817–18, 2817 n.27.

72. *Id.* at 2818 n.27.

73. *Id.* at 2821.

74. *Id.*

The majority refused to defer to findings by the District regarding the problem of gun violence and the efficacy of strict gun control as a means of stanching that violence.[75] "[T]he enshrinement of constitutional rights," Justice Scalia wrote, "necessarily takes certain policy choices off the table."[76]

And yet, as the majority made clear, not all gun control laws are presumptively unconstitutional.[77] Specifically, the majority mentioned "longstanding prohibitions on the possession of firearms by felons and the mentally ill, or laws forbidding the carrying of firearms in sensitive places such as schools and government buildings, or laws imposing conditions and qualifications on the commercial sale of arms."[78] The Court also limited the right by defining "arms" to exclude "dangerous and unusual weapons."[79]

Responding to Justice Breyer's criticism of the Court's reticence to explicitly embrace a particular standard of review, Justice Scalia defended the majority's position on the ground that *Heller* "represents this Court's first in-depth examination of the Second Amendment."[80] The pieces of the opinion, however, suggest that the Court at least adopted a form of intermediate scrutiny, if not strict scrutiny tempered by categorical exclusions of certain persons and arms (e.g., felons and machine guns or other "unusual" weapons) from the scope of the right.[81]

But there is a cost to the Court's ambiguity on this point. Lower courts may refuse to infer a standard from the clues the Court provided, as I have here, and may find it easier to narrow *Heller* or even avoid it altogether. Unless the Court is willing to monitor the lower courts by accepting cases for review, the failure of the Court to articulate an explicit standard of review could impair the robust implementation—in the Fallonian sense—of the right.[82]

In contrast to the majority's reticence to talk about a standard of review, Justice Breyer's dissent laid his approach out in some detail.[83] Criticizing the respondents' call for strict scrutiny as "impossible,"[84] Justice Breyer would adopt an "interest-balancing inquiry" that he explains would avoid the rigid dichotomy of strict scrutiny versus rational basis.[85] He would, in other words, balance the

---

75. *Id.* at 2822.

76. *Id.*

77. *Id.* at 2816–17.

78. *Id.*

79. *Id.* at 2817.

80. *Id.* at 2821. Therefore, Justice Scalia added, "one should not expect it to clarify the entire field." *Id.*

81. *Id.* at 2817.

82. For more on this point, see Glenn Harlan Reynolds & Brannon P. Denning, Heller*'s Future in the Lower Courts*, 102 NW. U. L. REV. 2035 (2008).

83. *See Heller*, 128 S. Ct. at 2847–70 (Breyer, J., dissenting).

84. *Id.* at 2851 (arguing "true" strict scrutiny is "impossible" because "almost every gun-control regulation will seek to advance . . . a 'primary concern of every government—a concern for the safety and indeed the lives of its citizens'" that would justify restrictions on civil liberties) (quoting United States v. Salerno, 481 U.S. 739, 755 (1987)).

85. *Id.* at 2852.

interest of those who wish to own guns for self-defense against the interest of the government in restricting gun ownership to combat violent crime, noting that perhaps violent crime feeds the need of otherwise law-abiding citizens to own guns for self-defense in the first place.[86] But, and this was especially important in *Heller*, Justice Breyer would "defer[] to a legislature's empirical judgment in matters where a legislature is likely to have greater expertise and greater institutional factfinding capacity."[87]

Given the deference in Justice Breyer's formulation, he would find the District's ban passed muster.[88] After reviewing the highly contested empirical evidence, Justice Breyer concluded that the evidence consisted of "studies and counterstudies that, at most, could leave a judge uncertain about the proper policy conclusion."[89] But, he concluded, "legislators, not judges, have primary responsibility for drawing policy conclusions from empirical fact."[90] For that reason, Justice Breyer's dissenting opinion was not a concurring opinion, especially because, in his view, private self-defense was only a secondary aim of the Amendment,[91] and there were no superior alternatives to the ban.[92] There was no "clearly superior, less restrictive alternative," wrote Justice Breyer, because "the ban's very objective is to reduce significantly the number of handguns in the District."[93]

Justice Breyer was correct to criticize the majority for the lacunae in the reasoning that accompanied its conclusion that although the District's gun did not pass muster, not all gun controls were unconstitutional.[94] At the same time, his methodology is open to criticism. For example, as scholars have noted, balancing tests, like Justice Breyer's, often appear more transparent than they actually are.[95] Further, it is odd to "balance" a right against presumed governmental interests. Rights are usually thought of by citizens as trumps that aren't to be merely balanced away in an opaque judicial equation.[96] If the District is not required to exert effort to justify its draconian gun policy, then one wonders why Justice

86. *Id.* ("Any answer would take account both of the statute's effects upon the competing interests and the existence of any clearly superior less restrictive alternative.").

87. *Id.*

88. *Id.* at 2870.

89. *Id.* at 2860.

90. *Id.*

91. *Id.* at 2847.

92. *Id.* at 2864.

93. *Id.*

94. *Id.* at 2868–69 (arguing that "[t]he majority's methodology is . . . substantially less transparent than" the interest-balancing approach); *id.* at 2869–70 (asking what justifies the majority's presumption that certain gun controls were constitutional or that certain arms were excluded).

95. *See, e.g.*, T. Alexander Aleinikoff, *Constitutional Law in the Age of Balancing*, 96 YALE L.J. 943 (1987); Denning, *supra* note 31 (criticizing the Court's use of balancing in the dormant Commerce Clause doctrine).

96. *See, e.g.*, Brannon P. Denning & Glenn H. Reynolds, *Five Takes on* District of Columbia v. Heller, OHIO ST. L.J. (forthcoming 2008) (criticizing Justice Breyer's dissent on these grounds).

Breyer made a point of referring to the Second Amendment as embodying an individual right in the first place.

But ultimately the point is neither whether Winkler's—or any other scholar's—conclusions are unassailable;[97] nor whether the Bush Administration's, Justice Breyer's, or the *Heller* majority's standard of review is the best one to apply to cases implicating Second Amendment rights. The point is simply that *Heller* and the Second Amendment furnish an unusually clear case of the importance of choosing standards of review and of the enormous discretion courts have in designing and applying them—especially where a majority of the Supreme Court has declined to specify a single standard.

General statements about constitutional commands do not decide cases; courts must develop tools for implementing those commands. The New Doctrinalism promises study of those commands and offers, I submit, a useful framework for analyzing and critiquing the Supreme Court and the constitutional law it makes.

97. *See, e.g.*, Glenn Harlan Reynolds, *Guns and Gay Sex: Some Notes on Firearms, the Second Amendment, and "Reasonable Regulation,"* 75 TENN. L. REV. 137 (2007) (responding to Winkler). For example, I have doubts about some of Strauss's claims about the normative superiority of the common law method to fix constitutional meaning. *See supra* note 29.

TORT LAW—PRODUCTS LIABILITY—PREMARKET APPROVAL UNDER THE MEDICAL DEVICE AMENDMENTS OF 1976 AND ITS PREEMPTIVE EFFECT ON STATE COMMON-LAW TORT CLAIMS

Riegel v. Medtronic, Inc., 128 S. Ct. 999 (2008).

I. INTRODUCTION

On May 10, 1996, Charles Riegel, plaintiff, underwent a coronary angioplasty.[1] The surgeon, Dr. Eric Roccario, attempted to dilate Mr. Riegel's diseased and calcified right coronary artery.[2] Dr. Roccario used an Evergreen Balloon Catheter, a medical device marketed by Medtronic, Inc., the defendant.[3] The catheter, a Class III medical device, received premarket approval pursuant to the provisions set forth in the Medical Device Amendments of 1976 ("MDA") to the Federal Food, Drug, and Cosmetic Act.[4] The catheter's label, approved as part of its premarket application, stated that its use was "contraindicated" for patients, like Mr. Riegel, with "diffuse or calcified stenoses."[5] The label also warned against inflating the catheter beyond a burst pressure of eight atmospheres.[6] Dr. Roccario, however, inflated the catheter five times, equaling a pressure of ten atmospheres.[7] On the fifth inflation the catheter ruptured, causing Mr. Riegel to develop a heart block.[8] Despite undergoing emergency bypass surgery, Mr. Riegel suffered "severe and permanent personal injuries and disabilities."[9]

---

1. *See* Riegel v. Medtronic, Inc., 128 S. Ct. 999, 1005 (2008); Riegel v. Medtronic, Inc., 451 F.3d 104, 107 (2d Cir. 2006).

2. *See Riegel*, 128 S. Ct. at 1005; *Riegel*, 451 F.3d at 107.

3. *Riegel*, 128 S. Ct. at 1005.

4. *See id.*; *Riegel*, 451 F.3d at 107. A Class III medical device is classified by the Food and Drug Administration ("FDA") as a device which is "purported or represented to be for a use in supporting or sustaining human life or for a use which is of substantial importance in preventing impairment of human health" or a device that "presents a potential unreasonable risk of illness or injury." 21 U.S.C. § 360c(a)(1)(C)(ii)(I)–(II) (2000).
Under the MDA, a Class III medical device may enter the market through either the premarket approval process or the § 510(k) process. *See Riegel*, 128 S. Ct. at 1004. The § 510(k) process allows a device to forego premarket approval if it is "substantially equivalent" to one already on the market before the adoption of the MDA. *Id.*; *see also* 21 U.S.C. § 360c(f)(1)(A)(i)–(ii) (guidelines for classifying a device as a Class III).

5. *Riegel*, 128 S. Ct. at 1005.

6. *Id.*

7. *Id.*

8. *Id.*

9. *Riegel*, 451 F.3d at 107.

In 1999, Mr. Riegel and his wife, Donna Riegel, brought suit against the defendant in the United States District Court for the Northern District of New York.[10] The plaintiffs alleged, among other claims, that the catheter's design, labeling, and manufacture violated New York common law.[11] Defendant raised the affirmative defense of federal preemption pursuant to § 360k(a) of the MDA.[12] The district court "held that the MDA pre-empted Riegel's claims of strict liability; breach of implied warranty; and negligence in the design, testing, inspection, distribution, labeling, marketing, and sale of the catheter."[13]

On appeal, the Second Circuit affirmed the findings of the district court.[14] The court reasoned that the plaintiffs' claims were preempted because a judgment in the plaintiffs' favor would impose state requirements that differed from, or added to, the device-specific requirements in the catheter's premarket approval application.[15] On certiorari to the United States Supreme Court, *held*, affirmed.[16] The preemption clause in the Medical Device Amendments of 1976 bars state common-law tort claims challenging the safety and effectiveness of a medical device which has been given premarket approval by the Food and Drug Administration. *Riegel v. Medtronic, Inc.*, 128 S. Ct. 999 (2008).

II. PREEMPTION OF STATE COMMON LAW CLAIMS UNDER 21 U.S.C. § 360K(A)

The United States Supreme Court's last major analysis of § 360k(a) of the MDA was in *Medtronic, Inc. v. Lohr*.[17] In *Lohr*, the Court examined the

---

10. *Riegel*, 128 S. Ct. at 1005.

11. *Id.* The plaintiffs' specific allegations were: "(1) negligence in the design, testing, inspection, manufacture, distribution, labeling, marketing, and sale of the Evergreen Balloon Catheter; (2) strict liability; (3) breach of express warranty; (4) breach of implied warranty; and (5) loss of consortium." *Riegel*, 451 F.3d at 107.

12. *Riegel*, 451 F.3d at 107. The preemption clause of the MDA provides that no state "may establish or continue . . . any requirement . . . which is different from, or in addition to" a FDA-promulgated requirement "which relates to the safety or effectiveness of the device or to any other matter included in a requirement . . . under this chapter." 21 U.S.C. § 360k(a) (2000).

13. *Riegel*, 128 S. Ct. at 1005–06. The district court granted summary judgment on the preempted claims. *Riegel*, 451 F.3d at 108. The breach of express warranty, negligent manufacturing, and loss of consortium claims were allowed to proceed to discovery, but summary judgment was eventually granted on these as well. *Id.* at 107–08. The court concluded that the catheter's instructions "clearly disclaimed any express warranty" and the evidence as to the negligent manufacturing claim was insufficient to warrant a determination that the catheter's burst was caused by negligence, as opposed to over-inflation, a puncture from a calcium spicule in the plaintiff's artery, or some combination thereof. *Id.* at 108.

14. *Riegel*, 128 S. Ct. at 1006; *Riegel*, 451 F.3d at 127.

15. *Riegel*, 451 F.3d at 122.

16. *Riegel*, 128 S. Ct. at 1011. Mr. Riegel died before the case reached the Supreme Court, but his wife continued as petitioner on behalf of herself and as administrator of her husband's estate. *Id.* at 1006 n.3.

17. 518 U.S. 470 (1996).

preemptive scope of § 360k(a) in the context of claims brought against the manufacturer of a medical device marketed pursuant to the § 510(k) process[18] and attempted to address a split between the circuits concerning the effect of the clause on state common-law tort claims.[19]

Both the plurality and dissenting opinions in *Lohr* found that the plaintiffs' state law design claims were not preempted because the generality of § 510(k) clearance did not set out "requirements" for the design of the device that conflicted with state law.[20] The Court, however, disagreed on the meaning of "requirements" as the term applied to the plaintiffs' manufacturing and labeling claims.[21]

The four-Justice plurality determined that state statutes or regulations typically constitute "requirements" for the purposes of the MDA.[22] In contrast, the four dissenting Justices found that the plain language of the MDA preempted any state common law action imposing requirements differing from, or in addition to, the standards promulgated by the FDA, even if those requirements were applicable to all devices in general.[23] In his concurrence, Justice Breyer agreed with the dissent's interpretation of "requirements," stating that the term could be taken to mean legal requirements resulting from state tort law application.[24] Although Justice Breyer's concurrence in the judgment allowed the plaintiffs' claims to escape preemption,[25] scholars have noted that "[n]o fully articulated reading of the preemption provision gained the support of a majority of the Court."[26]

---

18. *Id.* at 480–81.

19. *See id.* at 484. *See generally* Kennedy v. Collagen Corp., 67 F.3d 1453, 1458–59 (9th Cir. 1995) (holding that the MDA did not preempt state common law claims of general applicability, regardless of the "class" of the device); Lohr v. Medtronic, Inc., 56 F.3d 1335, 1347–52 (11th Cir. 1995), *aff'd in part, rev'd in part*, 518 U.S. 470 (1996) (holding that some, but not all, state tort claims were preempted by the MDA); King v. Collagen Corp., 983 F.2d 1130, 1135–37 (1st Cir. 1993) (holding that all state tort claims brought against Class III medical devices were preempted by the MDA).

20. *See Lohr*, 518 U.S. at 493–94 (finding that the § 510(k) application process did not "require" the device to take a specific form). Justice O'Connor reasoned that the § 510(k) process demonstrated that a post-1976 device was equivalent to a device grandfathered in under the MDA and did not place any new "requirements" on the device. *Id.* at 513 (O'Connor, J., concurring in part and dissenting in part).

21. *See id.* at 497–500 (majority opinion); *id.* at 505–07 (Breyer, J., concurring in part and concurring in the judgment); *id.* at 511–14 (O'Connor, J., concurring in part and dissenting in part); *see also infra* text accompanying notes 109–117 (discussing the Court's different views regarding the meaning of the term "requirements").

22. *See Lohr*, 518 U.S. at 489 (plurality opinion).

23. *Id.* at 511, 513–14 (O'Connor, J., concurring in part and dissenting in part) (stating that the "statutory language does not indicate that a 'requirement' must be 'specific'").

24. *Id.* at 503–05 (Breyer, J., concurring in part and concurring in the judgment).

25. *See id.* at 508.

26. Robert B. Leflar & Robert S. Adler, *The Preemption Pentad: Federal Preemption of Products Liability Claims After* Medtronic, 64 TENN. L. REV. 691, 701 (1997).

After *Lohr*, lower courts generally "narrowed their treatment of preemption under [the MDA]," particularly where devices marketed under § 510(k) were concerned.[27] However, the question remained whether state common-law claims were preempted if alleged in the context of a device with premarket approval pursuant to § 360k(a).[28] Although a majority of the lower courts found that such claims were completely barred, there was enough inconsistency among the lower courts to warrant the Court's clarification of the issue.[29]

Riegel presented the Court with the opportunity to define the preemptive breadth of § 360k(a) when state common-law tort claims are asserted against the manufacturer of a medical device that has received premarket approval.[30] The two specific questions before the Court were: (1) whether the MDA's premarket approval process established federal "requirements" specific to an individual medical device; and (2) whether state common-law causes of action challenging the safety and efficacy of an FDA-approved medical device established "requirements" preempted by the MDA.[31]

A. The Basis of Federal Preemption

The doctrine of federal preemption is rooted in the Supremacy Clause of the United States Constitution.[32] The doctrine "explicitly grants priority to federal law when the state law interferes with, or is contrary to, federal law."[33] However, "[t]he critical question in any preemption analysis is whether Congress intended that federal regulation supersede state law," and any analysis "starts with the basic assumption that Congress did not intend to displace the state law."[34]

---

27. *Id.* at 710 (noting that since *Lohr*, lower courts have used various means in their analyses of preemption defenses under the MDA).

28. *Id.*

29. *See Riegel*, 451 F.3d at 117. *Compare* McMullen v. Medtronic, Inc., 421 F.3d 482, 490 (7th Cir. 2005) (holding that state common-law tort claims for failure to warn were preempted by the MDA); Horn v. Thoratec Corp., 376 F.3d 163, 180 (3d Cir. 2004) (holding that state common-law tort claims were preempted by the MDA); Brooks v. Howmedica, Inc., 273 F.3d 785, 799 (8th Cir. 2001) (holding that state common-law tort claims were preempted by the MDA); Martin v. Medtronic, Inc., 254 F.3d 573, 585 (5th Cir. 2001) (holding that state common-law tort claims were preempted by the MDA), *with* Goodlin v. Medtronic, Inc., 167 F.3d 1367, 1382 (11th Cir. 1999) (holding that premarket approval alone does not impose specific federal requirements on a medical device and has no preemptive effect under § 360k(a)).

30. *Riegel*, 128 S. Ct. at 1006.

31. *Id.* at 1006–07.

32. U.S. CONST. art. VI, cl. 2 ("[T]he Laws of the United States . . . shall be the supreme Law of the Land; and the Judges in every State shall be bound thereby, any Thing in the . . . Laws of any State to the Contrary notwithstanding.").

33. Anne-Marie Dega, Comment, *The Battle over Medical Device Regulation: Do the Federal Medical Device Amendments Preempt State Tort Law Claims?*, 27 LOY. U. CHI. L.J. 615, 619 (1996).

34. 16 AM. JUR. 2D *Constitutional Law* § 53 (1998).

B. Federal Preemption as a Defense to State Common-Law Claims: The Supreme Court's Interpretation

Federal preemption as a defense to state common law tort claims is a relatively recent use of the doctrine.[35] *Silkwood v. Kerr-McGee Corp.* is an early example of its application in this setting.[36] There, the Court held that an award of punitive damages for radiation injuries was not preempted by the Atomic Energy Act of 1954,[37] an act establishing regulations for nuclear safety standards.[38] The Court acknowledged the existence of a "tension" between the federal law's exclusive jurisdiction over nuclear safety regulations and the states' abilities to award damages based on tort liability.[39] However, the Court stated that "Congress intended to stand by both concepts and to tolerate whatever tension there was between them" so that plaintiffs injured by nuclear hazards could be compensated with punitive damages.[40] Although the Court focused on the compensatory function of the state claims, thereby nullifying the defendant's preemption defense,[41] "*Silkwood* had the effect of calling manufacturers' attention to the possibility that preemption could serve as a defense to tort claims."[42]

The next important case to address the preemption issue, *Cipollone v. Liggett Group, Inc.*,[43] was a "landmark decision" that laid the fundamental groundwork for the Court's current preemption analysis of state common-law claims.[44] In *Cipollone*, the plaintiff brought state common-law claims against various cigarette manufacturers, alleging that the manufacturers were responsible for his mother's death from lung cancer.[45] The defendants invoked federal preemption, claiming that the Federal Cigarette Labeling and Advertising Act of

---

35. *See* Richard C. Ausness, *"After You, My Dear Alphonse!": Should the Courts Defer to the FDA's New Interpretation of § 360k(a) of the Medical Device Amendments?*, 80 TUL. L. REV. 727, 737 (2006).

36. 464 U.S. 238 (1984).

37. *Silkwood*, 464 U.S. at 250–56 (stating that there is no evidence that Congress intended for the Atomic Energy Act of 1954 or the Price-Anderson Act to preempt state punitive damage awards).

38. 42 U.S.C. § 2011 (2000).

39. *Silkwood*, 464 U.S. at 256.

40. *Id.* (discussing the Price-Anderson Act, 42 U.S.C. § 2210 (2000), as evidence that Congress did not intend for federal law to preempt state law remedies for those "injured by nuclear incidents"); *see also id.* at 258 (holding that the plaintiff's claim for punitive damages was not preempted by federal law).

41. Catherine M. Sharkey, *Products Liability Preemption: An Institutional Approach*, 76 GEO. WASH. L. REV. 449, 467 (2008).

42. Leflar & Adler, *supra* note 26, at 696.

43. 505 U.S. 504 (1992).

44. Dega, *supra* note 33, at 630 (noting that although *Cipollone* focused on a broad examination of the federal preemption issue regarding state common-law tort claims, many courts have used the *Cipollone* analysis to consider preemption issues in MDA cases).

45. *Cipollone*, 505 U.S. at 508.

1965 ("1965 Act"),[46] as amended by the Public Health Cigarette Smoking Act of 1969 ("1969 Act"),[47] protected them from liability.[48] This defense prompted the Court to consider whether either statute could preempt state common law claims.[49] The original 1965 Act provided that "[n]o statement relating to smoking and health . . . shall be required" on cigarette packaging or in cigarette advertising if the packages were "labeled in conformity with the provisions of this Act."[50] The 1969 amendment stated that "[n]o requirement or prohibition based on smoking and health shall be imposed under State law with respect to the advertising or promotion" of cigarettes if the packages complied with the Act's provisions.[51]

In *Cipollone*, seven Justices found that the provision in the 1965 Act did not preempt state common-law damage claims as it "merely prohibited state and federal rulemaking bodies from mandating particular cautionary statements on cigarette labels or in cigarette advertisements."[52] The Court stated that the language of the statute was "best read as having superseded only positive enactments by legislatures or administrative agencies that mandate particular warning labels."[53]

The Court fractured as to the preemptive scope of the provision in the 1969 Act. Justice Stevens, writing for a plurality of four, determined that the 1969 Act "sweeps broadly and suggests no distinction between positive enactments and common law."[54] Although the plurality found that some of the plaintiff's claims were "requirements" that were preempted under the 1969 Act, it qualified this finding, observing that even though "the preemptive scope of § 5(b) cannot be limited to positive enactments[, this] does not mean that [it] preempts all common law claims."[55] The plurality, however, did not provide guidance in determining when a preemptive clause would bar all or only some state common-law claims.[56]

Concurring and dissenting in part, Justice Scalia agreed with the plurality's conclusion that "requirements" could encompass more than statutes or regulations to include state common-law tort duties.[57] However, he firmly disagreed with the "narrow" construction of the provision in the 1969 Act, stating

---

46. Pub. L. No. 89-92, 79 Stat. 282 (1965).

47. Pub. L. No. 91-222, 84 Stat. 87 (1970).

48. *Cipollone*, 505 U.S. at 510.

49. *Id.* at 510–12.

50. *Id.* at 514 (quoting the Federal Cigarette Labeling and Advertising Act of 1965 § 5(a)–(b)).

51. *Id.* at 515 (quoting the Public Health Cigarette Smoking Act of 1969 § 5(b)).

52. *Id.* at 518 (citations omitted).

53. *Id.* at 518–19.

54. *Id.* at 521 (plurality opinion).

55. *Id.* at 523.

56. *See* Leflar & Adler, *supra* note 26, at 697-98.

57. *Cipollone*, 505 U.S. at 548–49 (Scalia, J., concurring in the judgment in part and dissenting in part).

that the Court's "job is to interpret Congress's decrees of pre-emption neither narrowly nor broadly, but in accordance with their apparent meaning."[58] He concluded that if the 1969 Act was accorded its plain meanings, all of the plaintiff's claims would be preempted.[59] Justice Scalia closed his dissent with a prescient question which would not be clarified for almost sixteen years: "For pre-emption purposes, does 'state law' include legal duties imposed on voluntary acts . . . or does it not . . . ?"[60]

Cipollone opened the door for the use of the preemption doctrine as a defense, specifically in the realm of products liability.[61] Likewise, the Court's indication that "requirements" went beyond positive enactments "provided a springboard for subsequent courts to hold injured consumers' claims preempted under other statutes."[62]

The next significant case to address the issue of preemption was *Geier v. American Honda Motor Co.*,[63] which illustrated the Court's use of preemption in preventing "tort as regulation" in the context of state common-law claims.[64] The issue before the Court was whether the National Traffic and Motor Vehicle Safety Act of 1966 ("MVSA") preempted a state common-law tort claim that a vehicle's design should have included airbags, despite the defendant's compliance with the Federal Motor Vehicle Safety Standard ("FMVSS 208").[65] The 5-4 opinion held that the plaintiff's lawsuit "conflict[ed] with the objectives of FMVSS 208, a standard authorized by the Act, and [was] therefore preempted by the Act."[66]

The Court based its holding on the DOT objectives: to gradually implement a variety of passive restraints, to "help develop information about the comparative effectiveness of different systems," and to "promote public acceptance."[67] The plaintiff's tort action, which essentially claimed that "manufacturers had a duty to install an airbag," would require manufacturers to choose airbags over other passive restraint systems, thereby presenting "an obstacle to the variety and mix of devices that the federal regulation sought."[68] In holding that the plaintiff's action was preempted, the Court "evinced little desire to balance the states' interest in compensating victims of commercial behavior that transgresses local

58. *Id.* at 544.

59. *Id.*

60. *Id.* at 555–56.

61. Dega, *supra* note 33, at 632–33.

62. Leflar & Adler, *supra* note 26, at 698.

63. 529 U.S. 861 (2000).

64. Sharkey, *supra* note 41, at 461–62.

65. *Geier*, 529 U.S at 864-65. FMVSS 208, promulgated by the Department of Transportation ("DOT"), required automobile manufacturers to provide "passive restraints" in "some but not all of their 1987 vehicles." *Id.*

66. *Id.* at 866.

67. *Id.* at 879.

68. *Id.* at 881.

norms against the drawbacks associated with the existence of nonuniform tort law in a national automobile market."[69]

In his dissent, Justice Stevens argued that the term "safety standard" used in the Act and in FMVSS 208 referred to affirmative legislative or regulatory acts, rather than "case-specific decisions . . . resolv[ing] common-law claims."[70] He characterized this distinction made by Congress as rational because "common law liability—unlike most legislative or administrative rulemaking—necessarily performs an important remedial role in compensating accident victims."[71] The dissent "raised the specter of the forgotten remedial, compensatory role of tort" from which the Court seemed to be drifting away from when regulatory interests were at stake.[72]

The conflicting viewpoints within the Court were further illustrated in *Bates v. Dow Agrosciences LLC*,[73] where the Court again determined that a "requirement" could include state common-law duties, but "an event, such as a jury verdict, that merely motivates an optional decision is not a requirement."[74] Justice Stevens, writing for a seven-Justice majority, held that § 136v(b) of the Federal Insecticide, Fungicide, and Rodenticide Act ("FIFRA") did not preempt state common-law complaints that were "fully consistent" with federal requirements.[75] In *Bates*, the Court appeared to strike a balance between "the compensation function of tort law" and a "respect for state sovereignty."[76] Justice Stevens distinguished the FIFRA statute from the one at issue in *Cipollone*, finding that the latter "prescribed certain immutable warning statements," while the former "contemplate[d] that pesticide labels [would] evolve over time" and that tort liability could serve as a "catalyst" in the development of safer products by revealing unknown safety hazards.[77]

Justices Thomas and Scalia, concurring in the judgment in part and dissenting in part, echoed the view from past opinions that state common-law claims, "even if not specific to labeling, nevertheless impose[] . . . requirement[s] 'in addition to or different from' FIFRA's," thereby warranting preemption.[78] The dissent concluded that when Congress included express language, as in

---

69. Sharkey, *supra* note 41, at 462.

70. *Geier*, 529 U.S. at 896.

71. *Id.* (citing Silkwood v. Kerr-McGee Corp., 464 U.S. 238, 251, 256 (1984)).

72. Sharkey, *supra* note 41, at 462.

73. 544 U.S. 431 (2005).

74. *Id.* at 445.

75. *Id.* at 452. The FIFRA prohibited states from imposing "requirements" on a pesticide's labeling or packaging that were "in addition to or different from" any federally mandated requirements. *Id.* at 439.

76. Sharkey, *supra* note 41, at 470–71 (examining Justice Stevens' analysis of the history of tort litigation, the continued availability of state remedies in pesticide litigation, and the general concern that farmers would not have a remedy if tort suits for misbranding and failure to warn were completely preempted).

77. *Bates*, 544 U.S. at 451.

78. *Id.* at 456 (Thomas, J., concurring in judgment in part and dissenting in part).

FIFRA, the majority's "presumption against pre-emption" was not applicable.[79] The dissent disagreed with the majority's statement that "[p]rivate remedies that enforce federal . . . requirements would seem to aid, rather than hinder, the functioning of FIFRA,"[80] countering with the argument that "it is for Congress, not this Court, to strike a balance between state tort suits and federal regulation."[81]

C. Preemption of State Common-Law Claims and § 360k of the Medical Device Amendments of 1976

I. The History and Purpose of the MDA

The relationship between the Court's preemption framework as it applies to § 360k of the MDA must be viewed through the historical background of medical device regulation and Congress' goals in promulgating the legislation. Until the MDA's enactment in 1976, the FDA's only regulatory role regarding medical devices was the authority to monitor the devices once they were placed on the market.[82] Following World War II, rapid advances in medical technology led to the development of devices that were more complex and potentially dangerous to the public,[83] prompting legislators to realize it was time "to require that all medical devices are safe and effective *before* they are allowed in the marketplace."[84] The urgency to act increased after the disastrous events concerning the Dalkon Shield, an intrauterine device utilized by millions of women before it was discovered to be highly defective.[85] The legislative hearings held on the MDA highlighted the need for preventative measures on the front end of medical device regulation so that similar injuries and deaths could be avoided.[86]

---

79. *Id.* at 457 (citing Cipollone v. Liggett Group, Inc., 505 U.S. 504, 545–46 (1992) (Scalia, J., concurring in judgment in part and dissenting in part)).

80. *Id.* at 451 (majority opinion).

81. *Id.* at 458.

82. *See* S. REP. NO. 94-33, at 2 (1975), *as reprinted in* 1976 U.S.C.C.A.N. 1070-71; Dega, *supra* note 33, at 623 (stating that the FDA only had the authority to monitor medical devices after the device had been introduced into interstate commerce).

83. S. REP. NO. 94-33, at 5 (1975), *as reprinted in* 1976 U.S.C.C.A.N. 1070, 1075.

84. S. REP. NO. 94-33, at 1 (1975), *as reprinted in* 1976 U.S.C.C.A.N. 1070 (emphasis added).

85. *See In re* N. Dist. Cal. Dalkon Shield IUD Prods. Liab. Litig., 693 F.2d 847 (1982); *see also* S. REP. NO. 94-33, at 1, *as reprinted in* 1976 U.S.C.C.A.N. 1070, 1071. The Dalkon Shield, available from 1970 until 1974, caused "uterine perforations, infections, ectopic and uterine pregnancies, spontaneous abortions, fetal injuries and birth defects, sterility, and hysterectomies," as well as several deaths. *In re Dalkon Shield*, 693 F.2d at 848–49. These injuries lead to the filing of over 3,000 lawsuits against the manufacturer. *Id.*

86. H.R. REP. NO. 94-853, at 8 (1976).

The resulting legislation placed medical devices into three classes, with the most stringent regulations applying to Class III.[87] Class III devices, which include heart valves and pacemakers, cannot be marketed under either Class I or Class II regulations because there is "insufficient information" from which to determine that either the general or special controls governing the lesser classes would "provide reasonable assurance of the safety and effectiveness of the device."[88]

In order to effectuate the goals of "safety and effectiveness", Congress included a premarket approval process specific to Class III devices.[89] Premarket approval requires that the manufacturer submit an application with information about the device's clinical investigations, components, manufacturing methods, labeling, and any other pertinent information or data relative to the efficacy of the device.[90] If the application is approved, the FDA sends an approval order to the manufacturer listing any additional condition the manufacturer must comply with prior to marketing.[91] Any changes to a device that could affect the design's "safety and effectiveness" must be approved through a supplemental application evaluated under the same standards as the original application.[92]

In drafting this legislation, Congress determined it was not feasible to require devices already on the market to undergo the premarket approval process.[93] However, such an exemption would afford an unfair competitive advantage to devices marketed prior to 1976.[94] In order to deal with this issue, Congress included two exceptions to the premarket approval process: a "'grandfathering' provision" for devices pre-dating 1976[95] and a "substantially equivalent"

---

87. *See* 21 U.S.C. § 360c(a) (2000). Class I devices are regulated through "general controls" and include items such as "surgeon's gloves, eye pads, and ice bags." Ausness, *supra* note 35, at 730–31; *see also* 21 U.S.C. § 360(a)(1)(A)(i)–(ii).

Class II devices are those for which "the general controls [applicable to Class I devices] are insufficient to provide reasonable assurance of the safety and effectiveness of the device." *Id.* § 360c(a)(1)(B). These devices are subject to "special controls" and include items such as "tampons, syringes, and neonatal incubators." Ausness, *supra* note 35, at 731.

88. 21 U.S.C. § 360c(a)(1)(C)(i)(I–II); *see also* Ausness, *supra* note 35, at 731.

89. *See* 21 U.S.C. § 360c(a)(1)(C); *id.* § 360e (outlining the premarket approval process).

90. *See id.* § 360e(c)(1)(A)–(C), (F)–(G). *See generally* 21 C.F.R. § 814.20(b)(6)(ii), (b)(8)(ii)–(iii) (2007) (describing the data and information the FDA requires in medical device applications). Upon completion, the FDA reviews the application, "spending an average of 1,200 hours on each submission." *See* Medtronic, Inc. v. Lohr, 518 U.S. 470, 477 (1996). The manufacturer must update the application upon receipt of new information that significantly affects the device's evaluation and must also comply with any requests for additional information made by the FDA. *See* 21 C.F.R. § 814.20(e); *see also id.* § 814.20 (b)(13).

91. *See* 21 C.F.R. § 814.44(e). The "conditions of use" in the proposed labeling serve as a basis for determining the safety of the device. 21 U.S.C. § 360e(d)(1)(A).

92. 21 U.S.C. § 360e(d)(6)(A)(i), (B)(i). However, certain incremental changes to a device's design or labeling that actually enhance its safety may be incorporated without approval. 21 C.F.R. § 814.39(d), (f).

93. *See Lohr*, 518 U.S. at 477-78.

94. *See id.* at 478.

95. *Id.*; *see also* 21 U.S.C. § 360e(b)(1)(A); 21 C.F.R. § 814.1(c)(1).

premarket notification process for devices that were sufficiently similar to those on the market.[96] The "substantially equivalent" or § 510(k) process[97] allows a device to enter the market, but it is not considered "approved" by the FDA.[98] Congress, in including these exceptions, sought to strike a balance between "the need for regulation to assure that the public is protected" and the recognized "benefits that medical research and experimentation to develop devices offers to mankind."[99]

Congress also included a preemption provision, prohibiting states from establishing "requirements" that are "different from, or in addition to" those imposed on the device by the FDA.[100] Since many states had enacted their own premarket approval systems to deal with a lack of regulation at the federal level, Congress drafted a savings clause that allowed states to apply for exemptions under certain circumstances, including when a state's regulations or requirements were "more stringent than [those] under [the MDA]."[101]

Regardless of the MDA's stated interests in regulatory consistency and fostering competition in medical device innovation, the overriding goals of the legislation were "to provide for the safety and effectiveness of medical devices intended for human use"[102] and to ensure "that the benefit of the doubt is always given to the consumer" as "it is the consumer who pays with his health and his life for medical device malfunctions."[103]

II. The Preemptive Scope of 21 U.S.C. § 360k(a) Prior to *Riegel v. Medtronic, Inc.*

The Court's evolving preemption jurisprudence converged with § 360k(a) of the MDA in the 1996 case of *Medtronic, Inc. v. Lohr*.[104] In *Lohr*, the main point

---

96. 21 U.S.C. § 360e(b)(1)(B) (2000); *see also id.* § 360c(f)(1)(A), (4)(A)–(B).

97. *Lohr*, 518 U.S. at 478 (the § 510(k) label refers to the section number from the original Act).

98. *See* Ausness, *supra* note 35, at 733 (explaining that a "substantially equivalent" device is "cleared" rather than "approved" by the FDA).

99. S. REP. NO. 94-33, at 6 (1975), *as reprinted in* 1976 U.S.C.C.A.N. 1070, 1075.

100. 21 U.S.C. § 360k(a).

101. *Id.* § 360k(b)(1). The other circumstance in which a state may apply for an exemption is when the state requirement "is required by compelling local conditions and compliance . . . would not cause the device to be in violation of any applicable requirement under this chapter. *Id.* § 360k(b)(2). Congress recognized California's "comprehensive" statutory scheme as "an example of requirements that the Secretary should authorize to be continued." H.R. REP. NO. 94-853, at 45-46 (1976).

102. Medical Device Amendments of 1976, Pub. L. No. 94-295, 90 Stat. 539 (1976), *as reprinted in* 1976 U.S.C.C.A.N. 539, 540.

103. 121 CONG. REC. 10,688 (1975) (statement of Sen. Kennedy).

104. *See* Medtronic, Inc. v. Lohr, 518 U.S. 470, 481 (1996). *Lohr* involved the failure of the plaintiff's pacemaker, which had been manufactured by Medtronic. *Id.* at 480–81. The device's lead, the part that transmits the electrical signal to the heart, was also manufactured by Medtronic and entered the market through the § 510k, or the "substantially equivalent" process. *Id.* at 480.

of contention between the Justices was whether the "requirements" in the MDA's preemption provision included state common-law tort claims.[105]

The important split in the opinion, and the portion that would later shape the Court's preemption analysis concerning state common-law claims with regard to § 360k(a), occurred during the disposition of the plaintiffs' manufacturing and labeling claims.[106] The plaintiffs argued that their claims were not preempted due to the instructive language of 21 C.F.R. § 808.1(d)(1).[107] Five Justices agreed with the plaintiffs' use of the regulation in determining the scope of § 360k(a), finding that all of the claims alleging defective manufacturing and labeling escaped preemption.[108] Although the Court did not state that general federal requirements would never preempt state requirements, the Court found it "impossible to ignore [the] overarching concern that pre-emption occur only where a particular state requirement threatens to interfere with a specific federal interest."[109]

Justice Breyer, in his concurring opinion, aligned himself with the dissent as to the meaning of "requirement" in the context of the MDA.[110] He noted that in *Cipollone*, the Court had previously determined that "similar language 'easily' encompassed tort actions because '[state] regulation can be as effectively exerted

---

The plaintiff's physician hypothesized that the failure of the device was due to "a defect in the lead," prompting the plaintiff to allege claims of negligence and strict liability against Medtronic. *Id.* at 481.

105. *See infra* notes 109–117 and accompanying text (discussing the Court's different interpretations of the term "requirements").

106. *See Lohr*, 518 U.S. at 498.

107. *Id.* at 498-99. The regulation provides that "[s]tate or local requirements of general applicability where the purpose of the requirement relates either to other products in addition to devices . . . or to unfair trade practices in which the requirements are not limited to devices" are not preempted. 21 C.F.R. § 808.1(d)(1) (2007). It also states that preemption of state requirements only occurs when the FDA has established "specific counterpart regulations" or "other specific requirements applicable to a particular device." *Id.* § 808.1(d).

108. *Lohr*, 518 U.S. at 501–02.

109. *Id.* at 500. Part VI of the opinion, in which only four Justices joined, stated that "few, if any, common-law duties have been pre-empted by this statute" and that "[i]t will be rare indeed for a court hearing a common-law cause of action to issue a decree that has 'the effect of establishing a substantive requirement for a specific device.'" *Id.* at 502–03 (plurality opinion). Justice Breyer, who wrote separately in concurrence, did not join in Part VI of the opinion because he was "not convinced that future incidents of MDA pre-emption of common-law claims will be 'few' or 'rare.'" *Id.* at 508 (Breyer, J., concurring in part and concurring in the judgment).

Part IV of the opinion, in which the same four Justices joined, rejected Medtronic's argument that all state common-law claims were "requirements" for preemptive purposes under the MDA. *Id.* at 486–87 (plurality opinion). Based on Congressional motivation in enacting the legislation, the Justices stated that if it was Congress' intent to preclude all state common-law claims, "its failure even to hint at it is spectacularly odd, particularly since Members of both Houses were acutely aware of ongoing product liability litigation." *Id.* at 491.

110. *See Lohr*, 518 U.S. at 503–08 (Breyer, J., concurring in part and concurring in the judgment).

through an award of damages as through some form of preventive relief.'"[111] Since the effects of state tort actions and state regulatory or legislative enactments are "identical," "[t]o distinguish between them for pre-emption purposes would grant greater power . . . to a single state jury than to state officials acting through state administrative or legislative lawmaking processes."[112]

In her dissent, Justice O'Connor also utilized *Cipollone's* rationale, concluding that the plain language of the MDA preempted state common law claims that were "different from, or in addition to" those required by the FDA "just as it would pre-empt a state statute or regulation that had [such an] effect."[113] Based on this analysis, the dissent found that at least some, if not all, of the plaintiffs' manufacturing and labeling claims would be preempted by § 360k(a).[114]

What seemed to be a lack of consensus among the Justices was later "characterized as a 'dual majority' case because Justice Breyer joined with the plurality to support the result, but agreed with the dissent that § 360k could preempt common-law tort claims."[115] Although the opinion did not specifically address the issue, the alliance between Justice Breyer and the *Lohr* dissent prompted many lower courts to find that § 360k completely barred state common-law claims concerning devices with premarket approval, while other lower courts continued to find no preemption.[116] As courts on "both sides of the issue have relied heavily on [*Lohr's*] analysis to support their conclusions," the time was ripe for clarification from the Court.[117]

---

111. *Id.* at 504.

112. *Id.*

113. *Id.* at 510–11 (O'Connor, J., concurring in part and dissenting in part). The dissent criticized the Court's use of FDA regulations in its interpretation, stating that the presence of expressly preemptive language in the statute rendered reliance on an Agency's interpretation "improper." *Id.* at 512.

114. *Id.* at 513–14. The dissent found that the FDA's Good Manufacturing Practice regulations and its labeling requirements were "extensive" requirements "certainly applicable to the device manufactured by Medtronic," and that "[§ 360k] require[d] no more specificity than that for pre-emption of state common-law claims." *Id.*

115. Ausness, *supra* note 35, at 744. Scholars have noted that neither Justice Breyer nor Justice O'Connor "once addressed the [MDA's] textual structure or legislative history, or the law's relationship to then-existing legal doctrine unquestionably permitting state-law tort claims against sellers of defective FDA-approved products." Leflar & Adler, *supra* note 26, at 707.

116. Ausness, *supra* note 35, at 748.

117. *Id.* Five years after *Lohr*, the Court decided a case involving the MDA but did not interpret the legislation's express preemption provision. Buckman Co. v. Plaintiffs' Legal Comm., 531 U.S. 341 (2001). However, the *Buckman* opinion indicated that § 360k might extend to state common-law claims. *Id.* at 348. The Court held that the state common-law claims brought by the plaintiffs "conflict[ed] with, and [were] therefore impliedly pre-empted by [the MDA]," as the FDA's sole statutory authority was "used by the Administration to achieve a somewhat delicate balance of statutory objectives." *Id.* at 348; *see also* Sharkey, *supra* note 41, at 464 (stating that the Court's preemption appeared to be "seeking to protect the federal regulatory scheme from being

III. *Riegel v. Medtronic, Inc.*

In *Riegel v. Medtronic, Inc.*,[118] an 8-1 decision, the United States Supreme Court held that the preemption clause in the Medical Device Amendments of 1976, 21 U.S.C. § 360k, barred state common-law tort claims challenging the safety and effectiveness of a medical device given premarket approval by the FDA.[119] The Court based its holding on the following premises: (1) premarket approval is specific to an individual device and requires little deviation from the terms of the approved application;[120] (2) common law causes of action can impose requirements for purposes of the MDA that are preempted by federal requirements specific to a medical device;[121] and (3) statutory reference to a state's "requirements," absent specification, includes common law duties.[122]

Justice Scalia, writing for the majority, began by comparing the facts of *Riegel* with *Lohr*, where the Court examined the preemptive scope of § 360k(a).[123] Justice Scalia noted that the *Lohr* majority used 21 C.F.R. § 808.1(d) to "substantially inform" its interpretation of the MDA's preemption clause, prompting the Court to hold that the plaintiffs' negligence and strict liability claims were not preempted.[124] He recognized the legitimacy of the Court's use of the regulation because the device at issue in *Lohr* was approved under the less stringent § 510(k) process and was therefore subject to generic federal labeling and manufacturing requirements applicable to most devices.[125] Although Justice Scalia opined that the "substantial-equivalence review under § 510(k) *is* device specific,"[126] he noted that the Court in *Lohr* rejected a similar argument from the defendant, finding instead that equivalency to a pre-1976 device was "an exemption rather than a requirement."[127]

Justice Scalia then contrasted the § 510(k) process of substantial equivalency with the premarket approval process at issue, stating that the latter "is in no sense an exemption from federal safety review—it *is* federal safety review."[128] He further distinguished the two processes, finding that premarket approval imposed "requirements" under the MDA as contemplated in *Lohr* because the process is

---

adulterated by the tort system's incentives"). The Court demonstrated concern not only for the stability of federal regulatory practices, but also for the interests of medical device manufacturers, observing that "complying with the FDA's detailed regulatory regime in the shadow of 50 States' tort regimes will dramatically increase the burdens facing potential applicants." *Buckman*, 531 U.S. at 350.

118. 128 S. Ct. 999 (2008).
119. *Id.* at 1007.
120. *Id.*
121. *Id.*
122. *Id.* at 1008.
123. *Id.* at 1006.
124. *Id.* at 1006; *see also supra* notes 107–109 and accompanying text.
125. *Riegel*, 128 S. Ct. at 1006–07.
126. *Id.* at 1007 (emphasis added).
127. *Id.*
128. *Id.*

tailored to each individual device.[129] Under the premarket approval process, "the FDA has determined that the approved form provides a reasonable assurance of safety and effectiveness" and such a device is required to enter the market with "almost no deviations."[130] This contrasts with the § 510(k) process in which devices are not required to "take any particular form for any particular reason."[131]

Justice Scalia then turned to the question of whether the claims "relate[] to the safety or effectiveness of the device" and whether the plaintiffs' state common-law claims were "different from, or in addition to" the catheter's federal requirements.[132] He found that "[s]afety and effectiveness [were] the very subjects of the Riegels' common-law claims," and then addressed the "critical issue" of "whether New York's tort duties constitute 'requirements' under the MDA."[133]

In determining whether the state common-law claims constituted "requirements," Justice Scalia looked to the Court's analysis in prior preemption cases, focusing again on *Lohr*.[134] He interpreted Justice O'Connor's opinion and Justice Breyer's concurrence to assert that "common-law causes of action for negligence and strict liability do impose 'requirement[s]' and would be pre-empted by federal requirements specific to a medical device."[135] Justice Scalia proclaimed that the Court "adhere[s] to that view," thus solidifying the lower courts' interpretation of *Lohr*—that Justice Breyer's concurrence constituted a fifth vote in support of Justice O'Connor's interpretation of "requirements."[136] Justice Scalia also noted that in both *Bates* and *Cipollone* the Court interpreted statutory provisions that preempted state "requirements" as preempting common-law duties.[137]

Justice Scalia reiterated the Court's newly solidified interpretation that "reference to a State's 'requirements' includes its common-law duties," stating that "Congress is entitled to know what meaning this Court will assign to terms regularly used in its enactments."[138] He went on to describe the logic in this

129. *Id.*

130. *Id.*

131. *Id.* (quoting Medtronic, Inc. v. Lohr, 518 U.S. 470, 493 (1996)).

132. *Id.* (quoting 21 U.S.C. § 360k(a) (2000)).

133. *Id.*

134. *Id.*

135. *Id.* Justice Breyer's specific language states that "the MDA will *sometimes* pre-empt a state-law tort suit" and that "[o]ne can reasonably read the word 'requirement' as including the legal requirements that grow out of the application, *in particular circumstances*, of a State's tort law." *Lohr*, 518 U.S. at 503–04 (emphasis added).

136. *Riegel*, 128 S. Ct. at 1007; *see, e.g.*, Riegel v. Medtronic, Inc., 451 F.3d 104, 116 n.14 (2d Cir. 2006) (stating that "we believe that Justice Breyer's crucial fifth vote endorsed the proposition that a state requirement could stem from a state common law tort action premised on the breach of a standard of care"); Horn v. Thoratec Corp., 376 F.3d 163, 176 (3d Cir. 2004) (stating that "on the state requirement issue, Justice Breyer joined with the four-member dissent to make a majority").

137. *Riegel*, 128 S. Ct. at 1007–08.

138. *Id.* at 1008.

construction: "State tort law that requires a manufacturer's catheters to be safer, but hence less effective, than the model the FDA has approved disrupts the federal scheme no less than state regulatory law to the same effect."[139] Justice Scalia also stated that tort law as applied by juries was "less deserving of preservation" than either a state statute or regulation because a state entity "could *at least* be expected to apply cost-benefit analysis similar to that applied by the experts at the FDA."[140]

Following this strict interpretation of the statutory term, Justice Scalia dismissed the dissent's use of legislative history to determine Congressional intent in enacting the MDA, Congress's interpretation of "requirement," and the preemptive scope of § 360k(a).[141] He stated that "[t]he operation of a law enacted by Congress need not be seconded by a committee report on pain of judicial nullification" and that it was not the Court's "job to speculate upon congressional motives."[142] Justice Scalia then observed that the text of the statute indicated that concern for injuries caused by FDA-approved devices was superseded by "Congress's . . . solicitude for those who would suffer without new medical devices if juries were allowed to apply the tort law of 50 States."[143]

Justice Scalia then dealt with another point raised by the dissent: the comparison between the approval process for medical devices and the approval process for drug, food, and color additives under the Federal Food, Drug, and Cosmetic Act ("FDCA").[144] He found "unreliable" Justice Ginsburg's conclusion that tort suits should be permitted for medical devices in much the same way that they were for drugs and additives.[145] In support of this view, he stated that (1) there was no evidence in support of the conclusion that *no* common-law tort suits were preempted by FDCA approval for drugs or additives and (2) if Congress had intended the two "regimes" to be alike, it "could have applied the pre-emption clause to the entire FDCA ."[146]

Justice Scalia next responded to the plaintiffs' argument that the state common-law claims against the defendant were general duties not specific to a particular device, and, therefore not preempted.[147] He found that "[n]othing in the statutory text suggests that the pre-empted state requirement must apply *only* to the relevant device, or only to medical devices and not to all products and all actions in general."[148] He rejected the basis of the plaintiffs' argument, which relied on 21 C.F.R. § 808.1(d)(1), in interpreting the statute's preemptive scope,

---

139. *Id.*
140. *Id.* (emphasis added).
141. *Id.* at 1008–09.
142. *Id.* at 1009.
143. *Id.*
144. *Id.*
145. *Id.*
146. *Id.*
147. *Id.*
148. *Id.* at 1010.

concluding that "[the regulation] can add nothing to our analysis but confusion."[149]

The majority stated that § 360k did not extend so far as to "prevent a State from providing a damage remedy for claims premised on a violation of FDA regulations" as such claims would "parallel" the federal requirements.[150] However, Justice Scalia concluded that the plaintiffs' claims were not of this category and affirmed the court of appeals' judgment in favor of the defendant.[151]

Justice Stevens concurred in part and in the majority's judgment that the plaintiffs' state common-law claims were preempted by the MDA.[152] Although he agreed with the dissent's analysis of the history and original purpose of the MDA's preemption provision, he noted that the statute's "text and general objective cover territory not actually envisioned by its authors."[153] Although "requirements" could include more than statutes or regulations, he stated that "common-law rules administered by judges," and not jury verdicts, were requirements for purposes of the MDA.[154]

Justice Stevens disagreed with the majority's suggestion that Congress had determined that the cost of administering judgments in favor of those injured by FDA-approved medical devices outweighed the benefits to those who would suffer from the application of each state's tort law.[155] He described this suggestion as a "policy argument advanced by the Court, not by Congress."[156] Justice Stevens agreed with Justice Ginsburg's assessment that the legislation's intent was to provide more, not less, protection to consumers.[157]

Justice Ginsburg, the sole dissenter,[158] approached the issue in diametric opposition to the majority. Whereas Justice Scalia stated that it was not the Court's "job to speculate upon congressional motives,"[159] Justice Ginsburg's analysis hinged on her assertion that "Congress . . . did not intend § 360k(a) to effect a radical curtailment of state common-law suits seeking compensation for injuries caused by defectively designed or labeled medical devices."[160] She began with a thorough recitation of precedential standards from the Court's past

149. *Id.* at 1010–11.

150. *Id.* at 1011.

151. *See id.* Justice Scalia noted that while the plaintiffs argued that their claims were parallel to the catheter's federal requirements, they did not make this assertion in the lower courts or in their petition for certiorari. *Id.* at 1011. As a result, the Court "declin[ed] to address that argument in the first instance here." *Id.*

152. *Id.* at 1011–13 (Stevens, J., concurring in part and concurring in the judgment).

153. *Id.* at 1011.

154. *Id.* at 1012, 1012 n.1.

155. *Id.* at 1012.

156. *Id.*

157. *Id.*

158. *Id.* at 1013 (Ginsburg, J., dissenting).

159. *Id.* at 1009 (majority opinion).

160. *Id.* at 1013 (Ginsburg, J., dissenting).

preemption analyses, language largely absent from the majority's opinion, all of which supported the long-held "presumption against preemption."[161]

Justice Ginsburg next discussed the history of the MDA's passage, which coincided with the well-publicized Dalkon Shield litigation.[162] She considered the absence of legislative history demonstrating any Congressional intention to preempt state common-law tort claims as an indication that Congress intended such claims to remain available to those injured by FDA-approved medical devices.[163] The absence of any federal compensatory remedy under the MDA further indicated that it was not Congress's intent to "broadly . . . preempt state common-law suits grounded on allegations independent of FDA requirements."[164] Justice Ginsburg stated that the Court's failure to consider these aspects of the MDA's history in its statutory construction of § 360k(a) "has the 'perverse effect' of granting broad immunity 'to an entire industry that, in the judgment of Congress, needed more stringent regulation.'"[165]

Justice Ginsburg also looked to the long history of drug and food additive regulation as a way to properly construe the MDA's preemption provision, arguing that this arena "informed, and in part provided the model for, its regulation of medical devices."[166] She specifically noted that drugs had received premarket clearance from the FDA since 1938 and that no evidence suggested that this precluded plaintiffs from bringing state common-law claims against drug manufacturers.[167] Justice Ginsburg pointed to the fact that when premarket approval requirements for drugs and additives were promulgated, "no state regulations required premarket approval . . . so no preemption clause was needed as a check against potentially conflicting state regulatory regimes."[168] Many

---

161. *Id.* at 1013–14.

162. *Id.* at 1014–15.

163. *Id.*; *see also id.* at 1015 n.7 ("If Congress intended [to preempt state tort claims], its failure to even hint at it is spectacularly odd, particularly since Members of both Houses were acutely aware of ongoing products liability litigation." (quoting Medtronic, Inc. v. Lohr, 518 U.S. 470, 491 (1996) (plurality opinion))).

164. *Id.* at 1015.

165. *Id.* at 1016 (quoting *Lohr*, 518 U.S. at 487 (plurality opinion)). Justice Ginsburg also used the rationale from *Silkwood* to support her view that the MDA's enactment was not synonymous with the goals of tort remedies. *Id.* ("It is 'difficult to believe that Congress would, without comment, remove all means of judicial recourse' for large numbers of consumers injured by defective medical devices." (quoting Silkwood v. Kerr-McGee Corp., 464 U.S. 238, 251 (1984))).

166. *Id.* at 1016–17.

167. *Id.* at 1017 n.10. Justice Ginsburg provided a list of citations to cases illustrating the proposition that most defendants in state tort litigation involving FDA-approved drugs did not raise preemption as a defense, and those that did were unsuccessful. *Id.* at 1017 n.11; *see also id.* at 1017 n.12 ("Surely a furor would have been aroused by the very suggestion that . . . medical devices should receive an exemption from products liability litigation while new drugs, subject to similar regulatory scrutiny from the same agency, should remain under the standard tort law regime." (quoting Leflar & Adler, *supra* note 26, at 704 n.71)).

168. *Id.* at 1018.

states, however, had implemented their own premarket regulations for medical devices, which, Justice Ginsburg concluded, was Congress's motivation for including a preemption clause in the MDA, "not any design to suppress tort suits."[169]

Justice Ginsburg also disagreed with the majority's argument that § 360k(a) preempted state common-law claims because "Congress would not have wanted state juries to second-guess the FDA's finding that a medical device is safe and effective when used as directed."[170] She again compared the premarket approval processes for medical devices with that of drugs, drawing attention to the fact that drug approval was "at least as rigorous as the . . . process for medical devices."[171] Further, Justice Ginsburg noted that "[c]ourts that have considered the question have overwhelmingly held that FDA approval of a new drug application does not preempt state tort suits."[172]

Applying her findings to the facts of the case, Justice Ginsburg, unlike the majority, did not find that § 360k(a) preempted the plaintiffs' claims.[173] She expressed apprehension at the far-reaching implications of the majority's interpretation of the statute, stating that "regardless of the strength of a plaintiff's case, [state common-law] suits will be barred *ab initio*."[174] She concluded that "[t]he constriction of state authority ordered today was not mandated by Congress and is at odds with the MDA's central purpose: to protect consumer safety."[175]

IV. CRITIQUE OF *RIEGEL*

Riegel provided the Court with an ideal platform from which to solidify its interpretation of § 360k(a) and its preemptive effect upon state common-law tort claims. The facts of *Riegel*, which indicated misuse of the catheter, rather than a product defect,[176] made it relatively easy for the Court to dispose of the case on the merits and clarify the interpretive issue at the center of the controversy. However, the Court's sweeping proclamation that "[a]bsent other indication, reference to a State's 'requirements' includes its common-law duties"[177] did far more than advise Congress on statutory drafting techniques or indicate the Court's preferred interpretation of the provision's language. In *Riegel*, the Court delivered manufacturers of premarket-approved medical devices a powerful tool with which to defend themselves from state tort liability. A scholar noted in response to the decision: "The Court goes the furthest it has to date in terms of

169. *Id.*
170. *Id.*
171. *Id.*
172. *Id.* at 1018–19.
173. *Id.* at 1020.
174. *Id.*
175. *Id.*
176. *See supra* text accompanying notes 2, 5–8.
177. *Riegel*, 128 S. Ct. at 1008.

provocatively suggesting that jury-imposed liability decisions would wreak havoc upon the federal regulatory scheme ensconced in the MDA."[178]

Riegel also issued a crushing blow to victims harmed by FDA-approved medical devices. The Court's deference assumes infallibility on the part of the FDA and a belief in the Agency's ability to police the safety of medical devices. Even today, years after the MDA increased the FDA's authority, there is evidence that a lack of funding and bureaucratic backlog has resulted in the FDA's inability to consistently perform its regulatory functions and sufficiently safeguard the public from the threat of defective drugs and medical devices.[179] By preempting state common-law claims regarding medical devices solely on the basis of FDA-approval, the Court has "remove[d] the opportunity for litigation to aid the FDA in its goal of monitoring product safety."[180] More importantly, those harmed by FDA-approved devices will be unable to obtain compensation.[181] Similarly, Senator Edward M. Kennedy, the sponsor of the MDA in the Senate, and Representative Henry A. Waxman, another staunch supporter of the legislation, noted that "design defect litigation affords an opportunity to identify . . . newly emergent risks and to consider alternatives . . . that would further consumer safety—the focus of the MDA."[182]

Although the *Riegel* majority decried such use of tort liability as "a potent method of . . . controlling policy,"[183] the Court's holding merely represents a policy it is willing to judicially promulgate, namely tort reform in favor of big business and regulatory consistency at the expense of injured consumers.[184] This is not to say that these interests are unworthy of judicial protection, but they are not the main interests Congress sought to protect by enacting the MDA.[185] The

---

178. Sharkey, *supra* note 41, at 464–65.

179. Catherine T. Struve, *The FDA and the Tort System: Postmarketing Surveillance, Compensation, and the Role of Litigation*, 5 YALE J. HEALTH POL'Y L. & ETHICS 587, 604–05 (2005) (stating that the FDA has much more work to do in improving its "postmarketing surveillance").

180. *Id.* at 591.

181. *Id.*

182. Brief of Sen. Edward M. Kennedy & Rep. Henry A. Waxman as Amici Curiae in Support of Petitioners at 20, *Riegel*, 128 S. Ct. 999 (2008) (No. 06-179).

183. *Riegel*, 128 S. Ct. at 1008 (quoting Cipollone v. Liggett Group, Inc., 505 U.S. 504, 521 (1992)).

184. *See* Leflar & Adler, *supra* note 26, at 707. The authors discussed a similar viewpoint in reference to the *Lohr* dissent, which Justice Scalia had joined. *Id.* They stated that the "Justices' failure to employ standard techniques of statutory interpretation is highly significant because it evidences a result-oriented jurisprudence—a willingness . . . to 'legislate from the bench.'" *Id.* Additionally, Leflar and Adler noted that "[i]n ascribing to Congress a fictive purpose to preempt state-law damage actions, [the dissent in *Lohr*] attempt[ed] to advance a tort reform program of [their] own. *Id.* at 708. They argued that had the dissent's "view prevailed, the result would have been the undercutting of decades of accepted tort jurisprudence that sellers of injury-causing medical products meeting minimal or general federal standards are not immunized from responsibility for their wrongful acts." *Id.*

185. *See id.* at 707–08.

Court's analysis ignores the statute's legislative history by failing to take into account that "[t]here is no suggestion anywhere . . . that Congress even considered preempting state tort suits, much less that it intended to preempt [them]."[186] Instead, the preemptive clause was meant "to reconcile the... federal regulatory scheme with device regulatory schemes that states had adopted in the absence of federal regulation."[187]

The erroneous nature of the *Riegel* majority's interpretation of § 360k's preemptive effect is further demonstrated by the context of the term "requirement."[188] Under § 360k(a), "[e]very [federal] requirement imposed on a medical device manufacturer... is imposed by... statute, by regulation or by an order applicable to a specific device," not by "a court or any other body."[189] Therefore, the reference to state requirements "different from or in addition to" federal requirements gives every indication that "Congress intended for the term state 'requirement' to apply to medical device requirements enacted by a legislature or promulgated by an administrative agency."[190] Similarly, § 360k(b) was included so that states with more exacting regulations could apply for exemptions, and "[i]t is difficult to imagine how such an exemption could apply to a products liability jury verdict, either in advance of or after such a verdict."[191]

Since "identical words used in different parts of the same act are intended to have the same meaning,"[192] it follows that the word "requirement," so clearly indicative of positive enactments in § 360k(b), "must be limited in the same way" in § 360k(a).[193] Although the majority claimed that the "statute itself speaks clearly" to the meaning of "requirements" for purposes of the MDA,[194] it seems that the Court misconstrued the term's meaning by ignoring the clear indication that Congress contemplated "requirements" to mean positive enactments.[195]

Although statutes do not exist in a vacuum and often evolve through application, it is erroneous to completely disregard not only Congressional intent in enacting a statute like the MDA, one so bound up with the goal of consumer protection, but also to ignore the clear textual meaning of the statutory language itself. The majority's suggestion, based on the plain language of the statute, was that Congress's concern for those injured by FDA-approved devices was

---

186. Brief of Sen. Kennedy & Rep. Waxman, *supra* note 182, at 8.

187. *Id.* Support for this interpretation is also found in the history leading up to the MDA's enactment, particularly regarding the Dalkon Shield intrauterine device. *See supra* notes 85–86; *see also* Brief of Sen. Kennedy & Rep. Waxman, *supra* note 182, at 15–16 (discussing Congressional interpretation of the term "requirement").

188. Brief of Sen. Kennedy & Rep. Waxman, *supra* note 182, at 12.

189. *Id.*

190. *Id.*

191. *Id.* at 13–14.

192. Comm'r v. Lundy, 516 U.S. 235, 250 (1996) (quoting Sullivan v. Stroop, 496 U.S. 478, 484 (1990)).

193. Brief of Sen. Kennedy & Rep. Waxman, *supra* note 182, at 14–15.

194. Riegel v. Medtronic, Inc., 128 S. Ct. 999, 1009 (2008).

195. *See* Brief of Sen. Kennedy & Rep. Waxman, *supra* note 182, at 14–15.

superseded by Congress's "solicitude for those who would suffer without new medical devices if juries were allowed to apply the tort law of 50 States."[196] This results in a perverted interpretation of the MDA's preemptive reach out of step with the drafters' intention that "the benefit of the doubt is always given to the consumer."[197]

The extent to which *Riegel* will affect future litigants with state common-law claims against manufacturers of FDA-approved medical devices and drugs remains to be seen. The prospects of succeeding on such claims, even if brought against a device marketed under the less stringent § 510k process, do not look promising, as Justice Scalia indicated that § 510k claims would be subject to the same stringent preemptive analysis as claims brought against a device with premarket approval.[198]

The force of the Court's decision in *Riegel* will become clearer after it decides *Levine v. Wyeth*, where the Court will consider "[w]hether the prescription drug labeling judgments imposed on manufacturers by the [FDA] preempt state law product liability claims premised on the theory that different labeling judgments were necessary to make drugs reasonably safe for use."[199] The general trend in the Court's preemption jurisprudence suggests that the Roberts Court is "very willing to let federal law trump state law when business interests are at stake,"[200] even if it means leaving plaintiffs injured by FDA-approved devices or drugs with no viable means of compensation.

V. CONCLUSION

In *Riegel*, the Court clarified two issues regarding FDA-approval under the Medical Device Amendments of 1976 and its effect on the preemption of state common-law tort claims. First, the Court determined that premarket approval under the MDA establishes federal "requirements" specific to an individual medical device.[201] Second, in adherence to the method of statutory interpretation present since its decision in *Cipollone*, the Court proclaimed that state common-law tort claims can constitute state "requirements" that are expressly preempted by § 360k(a) if different from, or in addition to, the requirements established by a device's premarket approval application.[202] However, in bypassing the statute's legislative history and the textual context of the term "requirements," the majority promulgated an interpretation of the MDA's preemption provision that

196. *Riegel*, 128 S. Ct. at 1009.

197. 121 CONG. REC. 10,688 (1975) (statement of Sen. Kennedy).

198. *See Riegel*, 128 S. Ct. at 1007 (determining that § 510k review is device specific, contrary to the finding of the Court in *Lohr*).

199. *Id.* at 1019 n.16 (Ginsburg, J., dissenting) (citing Petition for Writ of Certiorari, Wyeth v. Levine, 128 S. Ct. 1118 (No. 06-1249)) (citation omitted).

200. Erwin Chemerinsky, *A Troubling Trend in Preemption Rulings*, 44 TRIAL 62, 62 (2008).

201. *See Riegel*, 128 S. Ct. at 1007.

202. *See id.*

is inconsistent with the statute's original goals and likely to result in an application of § 360k(a) never intended by the legislation's sponsors.

FRANCES REGINA KOHO

TENNESSEE LAW REVIEW

ANNUAL INDEX
VOLUME 75
2007-2008

THE UNIVERSITY OF TENNESSEE
COLLEGE OF LAW

INDEX TO VOLUME 75

Tennessee Law Review Association, Inc.
1505 West Cumberland Avenue
Knoxville, Tennessee 37996-1810

AUTHOR INDEX

SUBJECT INDEX

TABLE OF CASES

D

E

F

T

U

V

W

Y

Z