The Mason Ladd Lecture, the premier lecture series presented at the Florida State University College of Law, honors the memory of Mason Ladd, the founding Dean of the law school. After guiding the law school at the University of Iowa for over 25 years as Dean, Ladd accepted the position of Dean at Florida State and began to plan the establishment of our new law school. When he accepted this challenge, there was no building or faculty, and were no students. However, in the fall of 1966, the law school at Florida State opened with over 100 students enrolled. Under his leadership and constant prodding, the College of Law began its steady growth. Prior to his retirement at Florida State in 1969, he obtained funding for a new College of Law building and completed the preliminary planning for its construction. For three years thereafter Mason returned each winter to Tallahassee to teach evidence and renew old acquaintances.

Mason Ladd was a scholar who was intimately involved in the creating of the modern law of evidence. He was a member of the Standing Committee on Rules of Practice and Procedure of the Judicial Conference of the United States which reviewed the work of the United States Supreme Court Advisory Committee on the Federal Rules of Evidence and forwarded the rules to the Supreme Court of the United States for their approval. He also participated in the drafting of the Model Code of Evidence and was a member of the committee of the National Conference of Commissioner on Uniform State Laws which drafted the Uniform Rules of Evidence. He authored and co-authored casebooks on evidence, federal jurisdiction, civil procedure, and Iowa probate practice. His book, Cases and Materials on Evidence, was frequently cited by courts and widely used in law schools. Among the numerous awards Mason received for his scholarly endeavors was the Fellows Research Award presented by Fellows of the American Bar Foundation for his extensive research in law and government.

Ladd was a master in the classroom. I first knew him as a student in his evidence class at Iowa. The joy and fervor that Mason brought to the law of evidence was catching and his students quickly became
immersed in the theory and rules of evidence. It was almost impossi-
ble not to learn with a teacher who was so inspiring. Some of the
cases which were discussed will never be forgotten by his students.
With his canes waving in the air, his discussion of the escapades of
“Chiggers,” the mule, during his trip down the Grand Canyon was a
classic.

Mason had a sincere interest in seeing his faculty succeed. In
1967, Mason hired me at the new law school at Florida State. I was
more than a little nervous when Mason told his three new faculty
that during the week before classes each of us would teach our first
class to him as he sat alone in a classroom equipped for over 125 stu-
dents. Although I was somewhat shaky asking Mason the hypotheti-
cals that I intended to use in my first Torts class, he played the role
of a student well and was full of praise at the completion of the hour.
After that experience, my first hour in front of real students was an-
ticlimactic.

Mason was proud of the law school he started at Florida State and
worked hard to make it a success. The charter class was encouraged,
cajoled, and made to work hard. He took a personal interest in his
students. There were a number of students having difficulty in a
course who were stopped by Mason in the hall, invited to his apart-
ment for Sunday dinner, and told to bring their class notes. After the
meal and visiting had been finished, the table would be cleared and
Mason would spend the remainder of the afternoon tutoring. His in-
terest in Florida State remained with him after he returned to Iowa.
Shortly before his death, he helped with Florida State’s application to
The Order of the Coif. He was proud and happy when our chapter
was awarded.

In his free time in Tallahassee, Mason tried to swim daily and
liked to fish for bass when he had the opportunity. He talked fre-
quently about his farm outside of Iowa City and looked forward to re-
turning. He enjoyed seeing his orchard bear fruit and driving his well
cared for tractor over the fields. Mason was close to his family. His
wife, Esther, was truly his partner in life. He was a man of strong
moral conviction and character.

The law and legal education miss Mason Ladd. His personal traits
together with his dedication, scholarship, and enthusiasm for the law
serve as an exemplar for those who knew him. He would be pleased
that this lecture series endures in his memory.

CHARLES W. EHRRHARDT
LADD PROFESSOR OF EVIDENCE
FLORIDA STATE UNIVERSITY

---

1. His scholarship has been cited almost 600 times in law reviews and journals since
his death. Tome v. United States, 513 U.S. 150 (1995), is the most recent decision of the
United States Supreme Court to cite Dean Ladd as authority.
Professionalism has an idealistic dimension and an institutional one. The idealistic dimension is the notion of voluntary commitment to both client interests and public values. The institutional dimension is the ideal of self-regulation by the bar.

The idealistic dimension remains powerful. However disappointed we are by the distance between the profession’s ideals and its members’ practices, these ideals continue to inspire valuable efforts. Various professional organizations are making admirable contributions through pro bono representation of disadvantaged people, public education, and disinterested law reform efforts in a range of areas, such as litigation procedure, prisons, and judicial selection. Moreover, the bar’s ideals of public service provide the vantage point from which the profession’s critics assess and propose improvements to its practices.

The institutional dimension is another story. It is implausible in principle and corrupt in practice. Its current manifestations are a set of strained rationalizations for tawdry self-seeking. The cynicism that the bar’s self-regulatory project induces in lay people spills over to discredit the idealistic dimension. We could strengthen the appeal of the idealistic dimension of professionalism by jettisoning the institutional one, or at least revising it substantially.

The core of the institutional dimension is private, monopolistic regulation. Traditionally, professionals have sought to exempt themselves from the suspicion that conventionally attends private monopoly. But few disinterested observers have been persuaded. It is not certain that the bar acts as a self-seeking monopolist, but in such matters as admission to practice, the marketing of legal services, and even conflict and disclosure norms, it seems unlikely that a self-seeking monopolist would have behaved any differently.

The most salient alternative to private monopolistic self-regulation is public regulation through state institutions. I think...
that some of the most promising recent developments in professional responsibility have occurred through interventions of national public institutions applying general legal norms in ways that have modified or put pressure on the profession’s own norms. What comes to mind is the Supreme Court’s striking down of various restrictions on admission and marketing under the free speech and federalism provisions of the Constitution.\(^1\) I am also thinking of the efforts of the Securities and Exchange Commission and the Office of Thrift Supervision to better align the professional commitment to confidentiality with the public values reflected in fraud and misrepresentation doctrine.\(^2\)

Nevertheless, we are especially sensitive these days to the limitations of state institutions as regulatory actors. And indeed, state institutions have been complicit in the regulatory abuses I attribute to the bar. The bar’s exercise of its monopoly power has depended on ratification by state judiciaries and legislatures. The kernel of plausibility in the idea of self-regulation is the implication that the state is too remote, inflexible, and compromised to provide the full range of institutional support for the idealistic aspects of legal professionalism. Moreover, the state is itself a monopolistic organization. To the extent that the objections to the institutional premise of traditional professionalism rest on its monopolistic rather than its private character, it seems most promising to consider non-monopolistic approaches, both public and private.

That is my plan. I first elaborate on the widely felt doubts about monopolistic self-regulation. Then I consider the possibility of non-monopolistic regimes for two of the most important areas—certification for practice and professional discipline.

I. THE TRADITIONAL STRUCTURE

I should explain what I mean by monopolistic self-regulation and, indeed, what I mean when I ask “Who Needs the Bar?” It may not be immediately obvious what one means by “the bar” in the American context.

The notion of self-regulation in the traditional conception is that practitioners organize in a single collectivity to promulgate and en-

---

1. See infra note 5.
force norms of good practice. Now, of course, no such organization has ever existed in the United States. The American model departs from the theoretical paradigm in two respects. First, the remission of primary regulatory authority over lawyers to the states means that, instead of a single over-arching association, we have more than fifty separate sets of regulatory institutions. Second, our constitutional arrangements forbid explicit delegation of regulatory power to private organizations. Thus, regulatory power at the state level resides formally, not in organizations openly controlled by practitioners, but in public institutions, most notably the judiciary.

Nevertheless, the image of the bar as a monopolistic self-regulatory organization has some descriptive power. Our arrangements are more complex, but their output looks very much like what one would expect of the private monopolistic paradigm. Most of the ethics rules adopted by our many jurisdictions are quite similar. These rules in turn resemble those promulgated by the American Bar Association (“ABA”), a private organization that is the closest we come to a national lawyers’ collective. The ABA aspires to represent the entire bar and in fact includes more than a third of it.

Moreover, within each state, organizations of practitioners have powerful influence over admissions and professional responsibility regulation. The “integrated” bars include all practitioners within the state and often have explicitly delegated powers (subject to judicial oversight). In states without integrated bars, inclusive voluntary bar associations that purport to speak for practitioners throughout the state play a strong role in both admissions and ethics rule-making and enforcement.

The political structure of the states would seem to make judges responsive to organized practitioners. Most state judges run for office in elections in which the average voter has little information or interest. Often, local practitioners will be the largest constituency with enough information or incentive to take an interest.

The legitimacy of this regulatory structure is currently under tremendous pressure. The pressure arises from two basic problems. The first problem is that this structure gives major influence to a group with a strong conflict of interest. This seems most apparent with respect to admission and marketing practices. The bar’s norms have restricted admission and inhibited price and service competition. The bar has public rationales for these norms, but since a substantial range of its members have a selfish interest in them, non-lawyers tend to be skeptical.

3. E.g., EMILE DURKHEIM, PROFESSIONAL ETHICS AND CIVIC MORALS (Cornelia Brookfield trans., 1957); A.M. Carr-Saunders, Professions, in 12 ENCYCLOPEDIA OF THE SOC. SCI. 476 (1934).
There are other grounds for this skepticism. Bar leaders are occasionally caught discussing the admission and marketing restrictions more or less openly as devices for insuring the economic welfare of incumbent practitioners. Further ground for doubt appears when we consider that, while these rules depend on controversial empirical assumptions, the bar has never shown any interest in investigating them. Would easier admission requirements—say two years of law school instead of three—really lead to lower quality practice? Would client interests be jeopardized if lawyers could practice with nonlawyer partners? Would deterrence of illegality be decreased or enhanced if confidentiality was cut back? The bar’s rules have been premised for centuries on empirical assumptions about such matters, but there is almost no research on any of them. The American Bar Association supports an excellent research institution—the American Bar Foundation—but it has never done any research on the factual premises of the profession’s core commitments.

The recent history of federal court review of state exclusionary practices in response to constitutional challenges is instructive and troubling. The challenged practices include citizenship requirements, residence requirements, in-state office requirements, requirements of association with local counsel, and a panoply of restrictions on advertising and solicitation. When the courts demand more than minimal rationality, the bar loses. The courts demand more when the practice impinges directly on an important federalism or free speech value. In such situations, the bar must offer more than fanciful speculation suggesting that there might be some legitimate policy


that under some not-totally-ridiculous but completely hypothetical scenario would be rationally served by the rule. The courts demand a policy with substantial plausibility and factual premises with some empirical support. The bar almost always fails to deliver.\(^6\)

When the values harmed by the practice are less constitutionally salient, courts apply minimum rationality. Only a failure of imagination could lead to flunking this test. When the bar justifies its requirement that out-of-state lawyers affiliate with local counsel in state court, even in cases under federal law, as a means of insuring familiarity with local procedural rules, the courts accept this as minimally rational.\(^7\) When the bar justifies its requirement that its members have in-state offices as a means of assuring availability to clients, this too will pass the test.\(^8\)

These conclusions may follow from the minimum rationality standard, but we should not ignore the layers of implausibility this standard by-passes. First, the connections between local admission and knowledge of local rules and between an in-state office and accessibility to clients are minor and wildly imprecise. It is likely that nearly all practitioners affected by the exclusions would have learned the local rules and made themselves accessible without having to undertake the expense of associating with a second lawyer or opening an in-state office.

Second, whatever benefits the exclusionary rules produce for clients have to be weighed against the costs they impose, which are ultimately borne by clients. The requirements raise the cost of lawyers and reduce client choice among them. The bar, of course, makes its

---

\(^6\) The Court did hold a speculative rationale—the likelihood of undue pressure from in-person solicitation—sufficient for more than minimal scrutiny in *Ohralki v. Ohio State Bar Ass'n*, 436 U.S. 447 (1978). The only Supreme Court case I am aware of in which the bar actually produced evidence that the Court deemed sufficient to satisfy greater-than-minimal scrutiny is *Florida Bar v. Went For It, Inc.*, 515 U.S. 618 (1995). The evidence, however, did not indicate that the advertising was harmful to clients, merely that it negatively affected “public perceptions” of the bar. The Court’s acceptance of this effect as a basis for a restriction of protected speech seems inconsistent with basic First Amendment principles. See *Am. Booksellers’ Ass’n v. Hudnut*, 771 F.2d 323 (7th Cir. 1985) (stating in the course of invalidating a pornography ordinance, that the tendency of the speech to induce negative perceptions of women is not a legitimate basis for regulation under the First Amendment).

The Court also accepted the bar’s rationale that current Communist political associations have some connection to fitness to practice law. See *Law Students Civil Rights Research Council v. Wadmond*, 401 U.S. 154 (1971); *In re Anastaplo*, 366 U.S. 82 (1961); *Konigsberg v. State Bar of Cal.*, 366 U.S. 36 (1961). But such inquiries are rarely pursued today, and at least one bar has come to regret the activities held permissible by the Court. David Rani, *30 Years Have Passed But License Still Sought*, Nat’l L.J., Aug. 22, 1983, at 8 (reporting that the Illinois State Bar Association petitioned the state supreme court to reconsider the 30-year old decision upheld in *In re Anastaplo*).

\(^7\) *Ford v. Israel*, 701 F.2d 689, 692 (7th Cir. 1983) (stating that the rule has “the air of a guild restriction” but satisfies minimum rationality).

\(^8\) *Tolchin v. Supreme Court of the State of N.J.*, 111 F.3d 1099 (3d Cir. 1997).
decisions without any information on the comparative magnitudes of the costs and benefits.

Third, in all such cases the bar has a choice of two approaches to dealing with the problem. It can deal with them prophylactically by excluding people or restricting practices it perceives have a tendency to harm clients. Or it can deal with the problem through after-the-fact sanctioning of bad practice. Lawyers who fail to comply with local court rules or to make themselves available to clients can be penalized. Thus, a plausible decision to adopt the exclusions would not only have to weigh the costs and benefits of the exclusions, but also compare them against the alternative of increasing after-the-fact enforcement. Of course, the bar does not do this, at least not explicitly and systematically. The bar’s conflict of interest is severe here, since the costs of exclusion are borne by outsiders, while after-the-fact enforcement sanctions members. And in fact, the indications are that after-the-fact enforcement tends to be indulgent and lax.9

The second source of pressure on the traditional regulatory model is the increased difficulty of formulating any principled delineation of the professional monopoly. Our version of the traditional model requires both that we distinguish law practice from non-law practice and that we assign particular instances of practice to particular states. It gets harder every day to do either. Substantively, the bar is hard pressed to explain why what lawyers can do in tax counseling is or should be different from what accountants do, or why what lawyers can do in conveyancing is different or should be different from what brokers and lay title searchers do. Even in litigation, once within the paradigmatic core of the lawyering field, we find lay mediators performing work that is hard to distinguish from that which lawyers do. In the areas that the federal agencies have opened to lay practice—patents, immigration, and tax—lay practitioners have flourished alongside lawyers.

The other aspect of the problem is the increasing extent to which practice is either federal or multi-jurisdictional. The typical practice of large firm business lawyers involves the law of numerous states and requires travel to many states where they are not licensed. It is commonly said that lawyers for large businesses routinely engage in unauthorized practice, and under the stricter definitions of unauthorized practice, this is certainly true. But a more fundamental concern is the difficulty of formulating any coherent definition of intrastate practice that embraces any significant range of lawyering.

How do we decide in what state or states a particular act of lawyering takes place? Clearly, the source of the relevant law is no

longer a plausible touchstone. It is too late in the day to say that a lawyer needs to be a member of the bar of any state whose law she gives advice about. Most experts on Delaware corporate law are not members of the Delaware bar. They are scattered throughout the country, the largest group of them being in New York. A broad range of lawyering tasks require consideration of the law of many jurisdictions. No one suggests that a national securities offering requires that fifty different lawyers give opinions on their respective states’ “blue sky” laws. Nor has there been a recommendation that a national publication needs to get its advice on libel from lawyers licensed in all of the jurisdictions where its product is distributed. And, of course, with respect to federal law, no state could claim that its lawyers are presumptively more qualified than out-of-state lawyers.

Recognizing this, the regulators are driven to focus on “presence” and “contacts.” But most physical contacts are relevant only because they implicate some relevant law. Surely there is no reason, other than that its law applies, for New Jersey to think that its own practitioners are better qualified to do conveyances of in-state real estate.

The one kind of contact or presence that might reflect a distinct state interest is the residence of the client. A state might decide that it had a special responsibility to protect its residents from bad lawyering, regardless of what law applied to the problem in question. But current doctrine is generally not consistent with such a policy. In litigation, we apply the disciplinary rule of the place where the court sits rather than the residence of the parties. In transactional work, the key factor is the state where the lawyer is licensed.

Moreover, the residence rationale fails to explain regulation of in-state practice on behalf of non-resident clients. The New Jersey bar insists, for example, that lawyers practicing in New Jersey maintain a “bona fide” office within the state. It enforces this rule zealously against lawyers based in Philadelphia who seek to represent clients in New Jersey. The principal rationale for the rule is that an in-state office assures that the lawyer will be available to clients. In fact, however, it appears that most of the clients served by the Philadelphia-based lawyers are Pennsylvania residents—banks that do lend-

Note that the two aspects of the boundary problem push the bar in opposite directions. The bar’s response to the fact that nonlawyers practice competently in particular fields is to emphasize lawyers’ competence in broad skills of complex analysis that cut across fields. Thus, it is the general analytical skills of lawyers that are said to be the comparative advantage that they have over nonlawyers. This is the general view of the ethics codes. The *Model Code of Professional Responsibility*, conceding that lay practitioners may be able to apply law in specialized fields, asserts that what lawyers have that lay practitioners do not is “professional judgment.” It defines “professional judgment” as the “educated ability to relate the general body and philosophy of the law to a specific legal problem of a client.”

And the *Model Rules* assert that it is fine for lawyers to accept employment in an area in which they are unprepared, so long as they intend to study up on it. The assumption is that lawyers are certified, not for knowledge of specific rules, but for a general aptitude that covers any legal field.

By contrast, the response to the jurisdictional aspect of the problem is to emphasize the importance of particular competence in local law. Thus, we are told that the reason out-of-state lawyers must affiliate with local counsel to appear *pro hac vice* in many state courts, and even some federal courts, is that in-state lawyers have better knowledge of local rules. A lawyer in any state is free to take on the most complicated tax or securities matter without demonstrating any prior knowledge of these bodies of law. But fear that lawyers might not be able to master a handful of unfamiliar procedural rules is said to warrant an exclusion that imposes substantial burdens and expenses.

It is difficult to see how the bar can have it both ways. If the “essence” of lawyering is, as the *Model Code of Professional Responsibility* asserts, “the ability to relate the general body and philosophy of law to a specific legal problem,” then there is no reason why lawyers

13. *Model Code of Prof’l Responsibility EC 3-5* (1981); see also *Model Rules of Prof’l Conduct* R. 1.1 cmt., para. 2 (2001): “Perhaps the most fundamental legal skill consists of determining what kind of legal problems a situation may involve, a skill that necessarily transcends any particular specialized knowledge.”
14. “A lawyer need not necessarily have special training or prior experience to handle legal problems of a type with which the lawyer is unfamiliar. . . . A lawyer can provide adequate representation in a wholly novel field through necessary study.” *Model Rules of Prof’l Conduct* R. 1.1 cmt., para. 2 (2001).
15. *E.g.*, Ford v. Israel, 701 F.2d 689, 692 (7th Cir. 1983).
should be required to demonstrate knowledge of any particular body of law. On the other hand, if law is just a series of discrete specialties, then it is hard to distinguish what lawyers do within those specialties from what experienced nonlawyer practitioners do, and hence hard to explain why they should or could have an exclusive right to a certain scope of practice.

A striking fact of recent history is the increasing homogenization of legal education and the admissions process. The trend in bar examinations for many decades has been away from testing local law to testing general principles. National law schools consider it beneath them to give systematic instruction in the law of any given jurisdiction. In most states, the Multistate Bar Exam counts for half or more of an applicant’s score. Even the essay portions test mostly general principles. For example, California, one of the most exclusive jurisdictions with a low bar pass-rate and almost no waive-in opportunities, tests virtually no local law. Review courses for the California bar typically tell the student what the “prevalent view is,” and what the “minority view” is; a star performer on the exam may have no idea what the California view is on most of the questions tested.

In this situation, it is difficult to take seriously the idea that a member of a particular state’s bar can be presumed to have a better knowledge of its law—the principal rationale for the exclusion of out-of-state lawyers.17

The recent report of the ABA Commission on Multijurisdictional Practice comes as close as possible to acknowledging the bankruptcy of monopolistic regulation without abandoning it. “[T]here is no evidence,” the report concludes, that unauthorized practice in one state by lawyers licensed in another “result[s] in the provision of incompetent representation.”18 Nevertheless, the Commission endorses the clarification and re-affirmation of unauthorized practice prohibitions for no better reason than that “a large segment of the bar” supports it.19

17. The ABA’s Commission on Multijurisdictional Practice mentions two other rationales for exclusion—the possibility that members of a state’s bar will have better knowledge of unwritten local customs that are relevant to effective practice and that members will be more likely to engage in in-state pro bono activity. See Am. Bar Ass’n, supra note 10, at 9. These rationales were held insufficient to justify residence requirements against federalism-based constitutional challenges. See, e.g., Supreme Court of N.H. v. Piper, 470 U.S. 274 (1985). Since the only remaining major entry barrier—the bar exam—is probably subject to a more deferential constitutional test, these rationales may be sufficient for that purpose, but it is doubtful that they could survive disinterested, critical examination. The bar exam has only the most speculative and attenuated relation to either goal, and if the bar were serious about either, it could pursue them more effectively at less cost by testing knowledge of local practice on the exam and by setting minimum pro bono requirements.


19. Id. The Report’s full explanation of its recommendation reads:
If the bar were a private organization, its practices would be struck down under the antitrust laws. What legitimacy they have is due to their adoption by the state courts. But the courts seem to have been a weak check on the economic self-seeking we readily impute to private monopolists. They have no accountability to out-of-state interests. Within the state, it seems at least possible that they will be more sensitive to the interests of highly organized practitioners than to the more diffuse interests of clients. And the fact that the courts have tended to more or less rubber stamp the output of the ABA and the state associations strengthens such doubts.

So it seems promising to consider competitive approaches to regulation. I focus on two of the most important areas—admissions and professional discipline. I consider an approach to admissions that involves competition among public authorities, and an approach to discipline that involves competition among private associations. But there are many possible variations on each proposal, involving different mixes of public and private institutions.

II. A COMPETITIVE ADMISSIONS REGIME

Given the large extent to which practice involves federal law and multi-state relationships, the case for federalization of lawyer regulation is a strong one. But the state-based regime might look more attractive if it were shorn of its monopolistic elements. At least, it is worth considering what a non-monopolistic regime of admissions would look like.

Some invoke the analogy of the driver’s license in discussing such a regime. Taken literally, the analogy connotes a regime in which each state accords those licensed in other states all the privileges accorded by the licensing state without imposing any additional qualifying conditions. No one proposes to go this far with lawyers, however; nearly all proposals contemplate some local requirements.20

Given the principle of state-based judicial regulation of the legal profession, the assumptions underlying that principle, and the support of a large segment of the bar for preserving it, the Commission believes that a stronger case would have to be made that national law practice is essential and that a more measured approach will not suffice to facilitate law practice and to promote the public interest.

Although this may sound like three reasons, it is only one. In the conceded absence of “evidence,” the “principle” of state monopolization and its underlying assumptions carry no more weight than the competing principle of freedom of contract and its underlying assumptions. The conclusion thus rests only on the “support” of the bar. Note also how the Commission, having failed to come up with any evidence to support the bar’s predisposition, insists without explanation that the burden of proof be placed on those who challenge it.

2003] WHO NEEDS THE BAR? 649

I find it useful to consider lawyer regulation in the light of another federalist analogy—corporate chartering. The corporate chartering regime involves a measure of reserved local power over out-of-state licensees that seems closer to what reformers contemplate for the legal profession. Moreover, a large literature on corporate chartering has sensitized us to the potential advantages of competitive state regulation in this area.21

Finally, the corporate regime is also interesting because the Supreme Court has indicated that the Constitution limits monopolistic regulation in this area.22 The Court summarily rejected suggestions that there are comparable limits on monopolistic regulation of the legal profession, but the arguments for Constitutional limitation seems at least as strong with the legal profession, and the Court may some day reconsider its casually-taken prior position.23

To an even greater extent than the legal profession, businesses that assume the corporate form have become increasingly multistate and increasingly subject to federal regulation. Yet, here, as with law and other occupations, our system has not moved toward federalization of the basic regulatory regime. The states retain primary responsibility for corporate chartering. Yet, the corporate model is one of competitive rather than monopolistic federalism.

In this model, each state permits corporations chartered by other states to conduct business within the state, subject to conditions tain-


23. The authority is sparse and ambiguous. In Norfolk & Western Railroad Co. v. Beatty, 423 U.S. 1009 (1975), aff'd mem. 400 F. Supp. 234 (S.D. Ill. 1975), the Court summarily affirmed a district court opinion suggesting there are no constitutional limits on state court decisions regarding admission pro hac vice in cases involving federal rights. In Leis v. Flynt, 439 U.S. 438 (1979), it rejected another constitutional challenge to a state court denial of admission pro hac vice in a case involving federal rights. The Court's opinion in Leis focuses on and rejects the claim that the lawyer has a federal constitutional right to state court admission pro hac vice. It addresses the more important claim of whether the claimant (client) has a right to have the attorney of his choice admitted only in a brief response to a dissenting opinion with a conclusory reference to the Norfolk & Western opinion. The dissent in Leis had invoked the contention in a Second Circuit opinion by Judge Friendly that “under the privileges and immunities clause of the Constitution no state can prohibit a citizen with a federal claim or defense from engaging an out-of-state lawyer to collaborate with an in-state lawyer and give legal advice concerning it within the state.” Spanos v. Skouras Theatres Corp., 364 F.2d 161, 170 (2d Cir. 1966).

Where the client is not asserting a federal right, the constitutional claim would be that the restriction unreasonably burdens interstate commerce, as in Edgar v. MITE. However, given the Court's deference to state regulatory interests in the Norfolk & Western and Leis cases and in occupational regulation cases in general, the prospects of such a claim seem dim.
lored to local needs. These conditions are usually minimal. The foreign corporation must pay fees and subject itself to suit in the state’s courts. Most states demand little more. In particular, while the states apply local law to the local activities of the corporation that affect outsiders to the corporation, they tend to defer to the chartering state in enforcing protections for shareholders.24

A few, however,—notably California and New York—do.25 They impose some restrictions from their own corporate law designed to protect local shareholders. These provisions are narrowly tailored, however. Only a few provisions deemed important are imposed, and they are limited to corporations whose investors and/or operations are concentrated within the state. Moreover, they typically exempt publicly traded corporations, without regard to whether their shareholders are concentrated within the state. The Supreme Court has interpreted the commerce and privileges-and-immunities clauses of the Constitution to preclude the states from going much farther than this. But most states have seen no need to approach the limits of constitutional power in this matter. Thus, within the typical state, the protection accorded in-state corporate shareholders in a foreign corporation is usually the law of the chartering state.

Although many would prefer a regime of federal chartering, there is an argument that a competitive federal regime has an advantage over a national one. If corporate promoters or managers have a choice of regimes and if shareholders are sensitive to the variations among them, then promoters or managers have a reason to seek out regimes that shareholders regard as superior. A solid legal regime increases the value of the corporation to shareholders and hence the price they are willing to pay for shares. State governments desiring to attract incorporations, in order to get revenues or prestige, will have reasons to make their protections of shareholders effective, and the competitive process will penalize those who fail. The extent to which the process functions this way is controversial, but the corporate chartering regime clearly enjoys considerably more legitimacy and respect than the very different state-based regime of lawyer licensing.

There is one element of the current regulatory regime for lawyers that corresponds to the corporate regime. In professional responsibility cases, states adopt a choice-of-law approach that looks to the licensing state’s norms for most purposes.26 However, since states ex-

25. Id. at 189-99.
26. MODEL RULES OF PROF’L CONDUCT R. 8.5 (2001). The main exception is for litigation practice, where the jurisdiction in which the court sits controls. Id. Note that, to the extent that the choice of law rule looks to the licensing state’s norms with respect, not just
tensively prohibit in-state practice by foreign lawyers, the choice-of-law norm is of limited importance. We can easily imagine, however, a lawyer regime developed along the corporate chartering model.

In it, each state would be obliged to permit lawyers licensed by other states to practice within the state subject to narrowly tailored restrictions designed to protect specifically identified local interests. It would, for example, be appropriate to require foreign lawyers, like foreign corporations, to consent to the adjudication of claims by local citizens before local tribunals. It would be appropriate to require minimum levels of liability insurance coverage. And a state might require special qualification by way of study or examination, but only with respect to genuinely local practice.

Thus, for example, it might in theory be appropriate to require, as a condition of practicing conveyancing or divorce within the state, an exam focused on the particular subject in question. I say “in theory” because, in fact, I doubt if most states, acting in good faith, would find any need for such protections, just as most do not find any need for specific protections for local shareholders of foreign corporations. After all, under the current regime, states do not condition the right to practice in most areas on specific testing in these areas. Many local practice areas are not tested at all on the bar exams, and none are tested more than cursorily. Nevertheless, restricting conditions to those focused on local matters in this manner would focus regulation on the areas of at least potentially legitimate state interest. Moreover, it is possible that testing focused on specific areas of practice in which the particular lawyers are about to engage would induce more valuable preparation than the once-over-lightly approach of the current exams.

In the competitive federalist regime, consumers within each state would face a choice of lawyers licensed in many different states for most services. Lawyers would be required to make clear at the outset where they had been licensed. The market might develop so that consumers could usefully take account of variations in state certification regimes. Some states might acquire a reputation for especially high general standards; some might acquire reputations for effective emphasis on certain areas of practice. Some states might compete to become the premier national brand, like Delaware in corporate law. Other states might focus on perceived local needs. A rural state like North Dakota, for example, might focus on small business skills. One could imagine that Florida might focus on estate planning, among other subjects.

to the protection of clients, but to the protection of third parties, it goes beyond the corresponding corporate norm, and is, I would submit, unjustifiable.
To the extent that state certifications send useful but varying signals about lawyer quality and preparation, there is every reason to think that sophisticated clients would be able to take account of such information. Even unsophisticated clients might, with the help of consumer rating services, be able to make meaningful distinctions among different regimes. The legal needs of unsophisticated clients tend to fall predominantly in a small number of categories, such as personal injury, divorce, and estate planning. Cross-state comparisons in quality of training might yield information that would be accessible to consumers.

No doubt some will fear a “race to the bottom” in which some states try to earn revenues by certifying poorly qualified candidates under easy standards and then export them to prey on out-of-staters. But reputation would seem a strong obstacle to this course in two respects. First, consumers would shy away from lawyers certified in the low-standards jurisdictions. Second, the better-qualified local practitioners in the low-standards state would feel degraded by the poor reputation of the local bar, and would likely push to change it.

If a state found that its citizens were victimized by low-quality foreign practitioners, it could, as I have suggested, act by establishing local standards with respect to genuinely local issues. But the restrictions ought to be narrowly tailored to clearly identified local interests.

### III. A COMPETITIVE DISCIPLINARY REGIME

Turn now to the ethical rules that protect third party and public interests. Critics within the profession, and most lay people, consider that the bar’s rules on confidentiality and zealous advocacy excessively sacrifice third party and public interests to client interests. They require or permit lawyers to withhold material information in situations where withholding may contribute to substantial injustice, and they even require or permit lawyers to actively obfuscate in some circumstances, for example, attempting to discredit witnesses they know are testifying truthfully.

While the general public seems not to share the bar’s commitment to strict confidentiality and aggressive advocacy, these rules are less often seen as an expression of economic self-interest than the rules specifically focused on admission and marketing. Critics are as likely to explain the bar’s ethical orientation in terms of ideological commitments as in terms of economic self-interest.

---

And in fact, the bar’s economic interests are ambiguous. On the one hand, strong confidentiality and aggressive advocacy enable lawyers to promise prospective clients that they will pursue their interests aggressively, even at the cost of injustice to others. This undoubtedly appeals to many clients. On the other hand, these norms may impair lawyers’ ability to serve other clients. Much of what lawyers do involves efforts to induce either state officials or private third parties to rely on their clients—that is to credit their claims and assertions or to trust them enough to enter into relations of mutual dependence with them. Strong confidentiality and advocacy norms limit such efforts. The client of the lawyer who adopts an ethic of minimal third party and public responsibility will pay a price in terms of diminished trust by third parties and officials. The client gets the benefit of zealous advocacy relatively unconstrained by commitment to anything but the unambiguous commands of the positive law. But she may experience more wariness on the part of the people she is dealing with. These people may be inclined to insist on more substantiation of factual representation or more specification of contractual terms to protect against the higher danger of opportunism from the client who opts for low commitment ethics. In marginal cases, private parties may refuse to enter into relationships that would enter with a higher degree of trust, and public officials may exercise their discretion against the client simply because they feel unable to credit the lawyer’s representations.

So, it is less clear whether the ethics rules reflect economic self-interest. Nevertheless, as with the admission and marketing rules, the ethics rules rest on debatable empirical assumptions that the bar tends not to analyze rigorously and has never sought to investigate factually. For example, does strong confidentiality promote compliance with law by inducing more disclosure to lawyers, thus enabling lawyers to give advice that encourages compliance? Or does it undermine compliance by reducing disclosure by lawyers, thus depriving lawyers of leverage to induce compliance and precluding enforcers from acting on the information? Obviously, the rule must have both effects. It increases compliance in some cases by increasing disclosure to lawyers, and it reduces compliance in others by reducing disclosure by lawyers. Whether strong confidentiality is legitimate depends on which effect is greater. But the bar’s premise that the pro-compliance effect dominates is based on nothing more than faith. It has no evidence on the matter and has never sought to acquire any.

One of the virtues of a competitive regime is that it would enable us to observe and compare the effects of alternative regimes. Regulators would have a range of information on effects that they do not now have.
Let us try to imagine what a competitive regime would look like with respect to such rules. Where the lawyer negotiates on behalf of clients with third parties, the parties might agree contractually to adopt a particular set of ethical rules for the transaction. When the lawyer appears before a tribunal or agency, she might simply indicate to the tribunal or agency what level of ethical responsibilities to the tribunal or agency she was committing herself to. Some lawyers might opt for the current regime of the ABA rules, with its low level of commitment to third-party and public interests. Others might opt for ethics reflecting a high level of commitment to third party and public interests, one requiring them to volunteer material information, for example.

In this way, individual lawyers could tailor their ethical commitments to particular clients or clienteles, and third parties could adjust accordingly. Moreover, one can imagine tribunals and agencies might adjust their conduct in accordance with an advocate’s level of commitment. Judges would be less apt to accept the informal assurances of low-commitment lawyers, and more wary about drawing inferences from their presentations at trial. Agencies might focus more enforcement resources on verifying the presentations of low-commitment lawyers. The IRS, for example, might consider the level of ethical commitment of a taxpayer’s preparer or advisor in allocating its auditing resources.

If it could work, such a regime would be both efficient and fair. It would be efficient because it would enable those with whom the lawyer deals to allocate their efforts to counter deception and opportunism more rationally. They would have better information about the degree to which they could plausibly trust a lawyer or correspondingly the degree to which they should devote resources toward protection against opportunism. It would be fair because it would facilitate protective and enforcement activity in ways that would better vindicate the substantive law. Deception and opportunism would succeed less often, which would enhance fairness. Clients of low-commitment lawyers would pay the price of increased wariness, but this seems entirely fair. The current situation in which lawyers are presumed to commit only to a low level of third-party obligation is unfair to clients who would be willing to bind themselves to a higher level. They currently pay a penalty because it is harder to distinguish them from low-commitment clients. A competitive regime would remedy this unfairness. Each client would receive the level of trust appropriate to the ethical commitment she was willing to make.

If such a regime would be both efficient and fair, you might ask why we do not have it already. The current regime does not preclude private parties from contracting for a higher level of ethical commitment than the minimum one required by the ABA rules, and many
tribunals and agencies would have the power to enact rules forcing lawyers who practice before them to declare the level of commitment they would commit to. In fact, we do occasionally see private contracts over such matters as disclosures. Think, for example, of a “10b-5 opinion” in a securities dealing, in which the lawyer warrants that, as far as she knows, no material information has been withheld. Nevertheless, we see little of this kind of activity.

There are a variety of reasons why such a regime has not arisen, even though it might be efficient and fair. Let me mention three problems, each of which might be remediable through public-spirited intervention.

First, issues of ethical commitment may be part of the category of issues research suggests that psychological inhibitions impede people from recognizing and raising. Summarizing some of this literature, Melvin Eisenberg suggests that people may not negotiate contractual safeguards against opportunism because they tend to be overly optimistic and insensitive to remote, hypothetical contingencies. Ethical issues also have an emotional charge that may generate anxiety. Because to doubt someone’s ethics is sometimes taken as a negative judgment on her character, a person may feel that simply raising the issue of ethical commitment will be perceived by the other as offensive.

The second problem arises from the costs of communicating about, negotiating, and drafting contracts. In order for a lawyer to find out a client’s preference on the matter, the lawyer has to explain to the client what a high commitment ethic would mean. In order to negotiate over whether to adopt a high commitment ethic, two parties have to arrive at some understanding of what it would mean. If they want to agree that each of them will act in accordance with the highest ethical standards, they can’t simply write a covenant promising to adhere to “the highest ethical standards.” They cannot be sure that they both agree about what the “highest ethical standards” are, or even if they do agree, they cannot be sure that in the event of a later dispute, an enforcement authority would know what their understanding was. So, first with their clients and then with opposing lawyers, lawyers would have to spend a lot of time discussing what these standards were and how they applied across the range of situations that might arise in their relationship between the parties. Once they arrived at an understanding, they would have to write out this understanding in a form intelligible to a court or other enforcement authority. This process would be costly and time-consuming, too costly and time-consuming for most situations.

In many areas, the legal system responds to this problem by providing default terms or optional terms that parties can use to fill in the gaps in their contracts. Private associations also provide standard form contracts designed to obviate the costs of elaborating terms for each deal. In the professional responsibility field, however, there is really only one standard form contract in each jurisdiction—the ABA Model Rules or Code. The problem is not that deviation is not permitted from the Rules or Code. Although many of the rules are mandatory, most of them simply provide a floor; they don’t preclude contracting for a higher level of commitment. But they do not do anything to reduce the costs of elaboration of these terms. Parties can have the low-commitment default terms of the ABA rules for free without doing anything. If they want higher standards, they have to assume the costs of explaining and drafting themselves.

My colleague Michael Klausner has suggested the preference of public corporations for Delaware’s corporation law might reflect, not the substantive superiority of Delaware law, but simply the fact that people are widely familiar with it and it is more extensively elaborated than other states’ laws. Because opting into Delaware’s rules saves communication, negotiation, and drafting costs, people may choose it even though they don’t especially like its substance. It’s possible that the same thing accounts for the failure to draft out of the ABA rules.

And third, we have the problem of enforcement. People will negotiate for a particular level of commitment only if they expect compliance with the negotiated standard, and compliance usually requires some enforcement apparatus. Tribunals and agencies will usually have some enforcement powers that they can devote to punishing defections from disciplinary commitments. Moreover, private actors who encounter each repeatedly can informally sanction defectors by refusing to trust them in later encounters. But where people encounter each other only once or sporadically, enforcement will be a problem. It would be easy enough to make some right of action in court available in contract or tort. But the high costs of enforcement in conventional damage suits would often not be warranted by the provable and recoverable damages. This, of course, is one of the reasons we supplement common law enforcement of ethical duties with disciplinary enforcement by the agencies of the bar. But these agencies, at present, are only available to enforce the low-commitment ethics of the ABA rules.

The project of competitive ethical regimes could be advanced by reforms designed to mitigate each of these difficulties. First, we need

rules that force the parties to focus on and make commitments regarding ethical standards. It would be easy enough for tribunals and agencies to require those who practice before them to indicate their type or level of commitment, either categorically or on a case-by-case basis. The Rules currently require lawyers to force the client to address the level of fees at the outset of the relation. In a similar manner, they could require lawyers to raise the issue of ethical commitment with both client and opposing counsel.

Second, to mitigate the costs of communication, negotiation, and specification, we need publicly subsidized alternative ethical codes. These codes need to be sufficiently elaborated so that, once chosen, they provide an array of relatively clear answers to a broad range of contingencies. These alternative codes would provide sets of norms that lawyers and clients could adopt without having to assume the costs of communication and drafting. They could, for example, take the form of a Restatement. Unfortunately, the American Legal Institute’s Restatement of the Law Governing Lawyers does not fill the bill at all. Like most of the Restatements, this one is largely concerned with summarizing the dominant tendencies of established low-commitment doctrine. It thus reinforces the tendency to opt for this doctrine simply because it is easiest to do so. What we really need are comparably clear and developed but strongly differentiated codes that increase the range of ethical options.

Here there may be a promising role for specialized bar associations, and it is heartening to see some rising to the challenge. To date, the outstanding example appears to be the American Academy of Matrimonial Lawyers. The Academy is a national association of great prestige to which some of the most distinguished practitioners belong. One of its many projects is the promulgation of a code of advocacy that explicitly “aspires to a level of practice above the minimum established in the [Rules of Professional Conduct].” The Academy’s rules take a familiar code form, with principles followed by elaborative comments and illustrative cases. Many of its precepts seem to depart notably from the ABA rules. For example, one provides, “[a]n attorney should not permit a client to contest child custody . . . for . . . financial leverage.” Another condemns “avoidance of compliance with discovery through overly narrow construction of in-
terrogatories or requests for production.”33 More activity of this kind could significantly expand the range of choice and the potential for ethical competition.

The Academy, however, does not directly respond to the enforcement problem. It does not sanction members for violation of its norms. There is no practical obstacle to its doing so. Private associations have the power of expulsion, as well as fines and reprimands, that can be effective deterrents. It would also be possible for the existing public enforcement agencies to enforce private codes to the extent that lawyers have made particular commitments to them.

IV. CONCLUSION

Lest my suggestion that we may not need “the bar” be taken as an aspersion on lawyer associational activity in general, I want to emphasize that I think there is a lot of valuable associational activity going in many areas, including that of professional responsibility. The work of the American Academy of Matrimonial Lawyers ethics code is just one example. We need more of this kind of work, as well as other kinds fostering pro bono representation, public education, peer support, and law reform. Such activities could be, and to a large extent are, undertaken by voluntary bar associations without the direct assistance of monopolistic state power. The “bar,” the need for which is now in doubt, is the monopolistic bar that seeks to occupy a broad field exclusively with a single set of mandatory standards.

It is truly an honor to be asked to Comment on the work of William Simon, one of the scholars who has done the most to contribute to the reputation of legal ethics as a field with intellectual rigor and depth, as well as one with significant implications for legal theory generally. The power of his critical faculties is unmatched: the platitudes offered by the organized bar in defense of the dominant view of legal ethics lie in tatters after the sustained assault in the first three chapters of _The Practice of Justice_. In fact, it can be difficult to find objections to the dominant view that Simon has not already articulated more forcefully. But his project is not merely critical, as his construction of the alternative contextual view of ethics shows. His Mason Ladd Lecture is a welcome extension of the contextual view, moving from the micro-evaluation of the ethics of individual lawyers into the macro level of institutional analysis and questions of regulatory regime design. Section I of this Comment is a brief review of this proposal.

Simon’s work has been a tremendous influence on my own thinking about legal ethics, so I have good reason to fear the ignominious fate of commentators who end up agreeing with the subject of their evaluation. Indeed there is a great deal in this Lecture to agree with. Some of his suggestions for reform are so far-reaching, however, that one is bound to have a few reservations and questions. Section II of this Comment contains some questions about the details of using a market-based approach and a diversity of ethical norms to regulate lawyers. Simon has anticipated many of these objections, none of which are likely fatal to his project, but some of which seem to be a bit more problematic than he acknowledges. For example, even if a
sufficient number of clients desire to hire high commitment lawyers, the transaction costs involved in matching up high commitment lawyers and clients may be sufficiently high to thwart the operation of the reputational market Simon envisions. Section III takes issue with the argument that nonlegal methods of regulation can avoid the corruption of regulation by self-interested professionals. Although nonlegal regulation offers many advantages over a formal scheme of legally enforceable rules, it is no less susceptible to capture by powerful actors than a system of legal regulation.

I.

There are two interrelated proposals in Simon’s Lecture: The first is to consider additional sources of legal regulation of lawyers, such as courts and legislatures in addition to bar associations, which tend to become corrupted by the economic self-interest of lawyers. Institutions not under the control of lawyers have a better record of looking out for the interests of non-clients, and once these other institutions have articulated different norms of lawyering, there is a chance they may catch on more widely. For example, the response of the Office of Thrift Supervision (OTS) to the law firms that assisted Lincoln Savings focused the attention of many in the profession and the academy on the justification for strenuously partisan norms of lawyering in the context of counseling and advising clients. Even if the OTS’s approach did not ultimately carry the day, the debate was useful. Thus, there is an intrinsic value in a diversity of normative approaches to lawyering, provided there is some way for regulators or lawyers themselves to select among them.

Another way to diversify the legal norms that may potentially be applicable to lawyers is to free states from the American Bar Association’s (ABA) monopoly on the drafting of disciplinary rules. The hope is that one or more states may seek to establish a reputation for regulatory norms that are more protective of third-party interests than the current crop of ABA models, including those recently proposed by the Ethics 2000 Commission. But why would a state do this, if the lawyers who control the organized bar have a selfish interest in limiting the protections offered by the rules to third parties? Here is where Simon’s second proposal comes in: namely, to rely on informal, decentralized mechanisms of social control to do some of the work currently entrusted to formal legal regimes. There is a certain irony

4. Id. at 640.
5. Id. at 648.
6. Id. at 652-54. The contrast between “formal” and “informal” methods of regulation and “legal” and “nonlegal” regulation is intended to track the usage of these terms in
here, because the talk of markets, “reputation bonds,” and mechanisms for reducing information costs that inevitably attends this kind of proposal relies, at its root, on the self-interest of the contracting parties. It is almost as though Simon’s idealistic side hopes that states, freed from the ABA’s regulatory monopoly, will aspire to rules that reflect a greater commitment to protecting third party interests, while his realistic side acknowledges the pervasiveness of market rhetoric and economic self-interest among lawyers,7 and has therefore tailored a proposal for reform that takes these realities into account. In the end, however, Simon’s market-based mechanism may be just as susceptible to corruption by the self-interest of lawyers as the formal, profession-based scheme of regulation that is the target of the critique in this Lecture.

It is important to clarify the usage of the term “professional monopoly” in the Lecture. Three different senses of the term crop up, and as a result Simon’s arguments against the professional monopoly can occasionally be inconsistent. One is that the profession seeks a regulatory monopoly vis-à-vis anything that can plausibly be characterized as “providing legal services.” Simon is right that there is no principled distinction between, say, tax planning as performed by a lawyer and by an accountant, so we have reason to be suspicious of this aspect of the professional monopoly.8 A second sense of monopoly, though, exists within the profession, because each state maintains a monopoly over the practice of law within its borders, enforced through prohibitions on the unauthorized practice of law.9 The notorious Birbrower10 case showed the expansive definition of practicing law within a state that may be articulated by an ambitious state aiming to protect local practitioners from out-of-state competition. Simon is again correct that it is almost impossible to draw principled lines here.11 Corporate lawyers in New York interpret Delaware law all the time, so the source of law cannot be dispositive. Physical presence is not only an illegitimate proxy for the source of law, as Simon notes,12 but an anachronistic standard given the prevalence of tech-

ROBERT C. ELICKSON, ORDER WITHOUT LAW: HOW NEIGHBORS SETTLE DISPUTES (1991). Ellickson’s study of cattle ranchers and farmers in Shasta County, California, showed that residents of this rural county relied on informal norms of neighborliness, not formal legal entitlements, as the starting point for bargaining to settle issues such as cattle trespasses and boundary fence disputes.

7. Simon, supra note 3, at 5, at 653; see also Jonathan R. Macey, Professor Simon on the Kaye Scholer Affair: Shock at the Gambling at Rick’s Place in Casablanca, 23 LAW & SOC. INQUIRY 323, 324 (1998).
8. Simon, supra note 3, at 644.
9. Id. at 641-44.
12. Id. at 644-45.
ology that facilitates interactions between lawyers and clients. And finally, there is no reason to suspect that resident lawyers are more competent that out-of-state lawyers, particularly since the law has such a marked tendency toward national uniformity, with local quirks being just that: quirks that can easily be mastered by an out-of-state lawyer. This leads to the third sense of professional monopoly as used in the Lecture, “monopolistic federalism,” which is another way of referring to the hegemonic effect of disciplinary rules drafted by the ABA and the ALI’s Restatement of the Law Governing Lawyers. Those two sources of regulatory norms exert so much influence that it is difficult to find much variation among states in the legal restrictions they place on lawyers.

Actually, one might quibble with the last claim. Some rules vary quite a bit among the states, and one of the provisions of the ABA Model Rules that most directly implicates the protection of non-clients, the scope of exceptions to the duty of confidentiality, is the one on which the diversity of state variation is the most pronounced. The rules supplement I use in my professional responsibility course even contains an elaborate chart, prepared by an insurer of large law firms, which details each state’s rules on the confidentiality rule and its exceptions. Nitpicking aside, though, Simon’s “competitive federalism” proposal is worth thinking carefully about. The heart of the programmatic portion of the Lecture is a celebration of polycentric regulation, from a diversity of state-based licensing approaches to a range of private mechanisms through which lawyers can signal their levels of ethical commitment, letting the market take over some of the task of post-admissions monitoring of lawyers’ conduct. I will concentrate on the second (and least developed) of these proposals, which is the most radical in its challenge to the regulatory monopoly of the organized bar.

In Simon’s view, the current disciplinary rules, at least in the ABA’s model version, articulate a low commitment standard of ethics in the sense of sacrificing legitimate interests of third parties, and the substantive justice of a matter, to the interests of lawyers and clients. The classic example is the strict rule against disclosing any

13. Id. at 649.
14. The inconsistency in the attack on the professional monopoly is between the objections to the second and third definitions of monopoly. If the Birbrower decision is objectionable because the practice of law is essentially national in scope anyway, because of the increasing homogenization of the law itself, as well as the nationwide uniformity in legal education and bar examinations, then it seems to make sense to regulate lawyers on a national basis as well.
information relating to the representation of a client even if its disclosure would avert a major financial catastrophe to a third party;\textsuperscript{17} but one could cite many others. Adversarial discovery practice, the norms governing deception in negotiations, the permissibility of confidentiality agreements in settlements (particularly in products liability cases), and settlement patterns in class actions are all areas in which the current law governing lawyers protects the self-interest of lawyers while imposing substantial externalities on non-clients. As a result of the indifference of the organized bar to the costs to third parties of the strongly role-differentiated adversary system, lawyers have forfeited the trust and respect of the public.\textsuperscript{18} Hence the double meaning inherent in the title of this Comment: the bar has bankrupted its reserves of public trust, so it is now time to play Teddy Roosevelt and break up the power of this group to articulate the sole authoritative body of rules governing lawyers. One method of accomplishing this objective might be to supplant legal regulation altogether—or certain domains of lawyering—with non-legal regulation.

II.

If the existing structure of rules expresses a default position of low commitment ethics, and if that structure is resistant to change because of the political power of the organized bar, giving individual lawyers the power to opt into a system of high commitment ethics would permit them to make an end-run around the regulatory monopoly of the bar. As Simon uses these terms, high commitment refers to a principled reluctance or refusal by the lawyer to take an action that is unjust, even if legal, and solicitude for the rights of third parties.\textsuperscript{19} The old-fashioned term for high commitment lawyers is “officers of the court.”\textsuperscript{20} Low commitment refers to acceptance of the dominant view of lawyering, in which the lawyer will take any arguably legal action on behalf of a client.\textsuperscript{21} The colloquial term for low commitment lawyers might be “hired guns.”\textsuperscript{22} An obvious, but misplaced, objection to this proposal might be that self-interested clients would never select a high commitment lawyer. Lawyers often claim that clients are seeking “attack dog” lawyers, and that market pressures nudge them in the direction of low ethical commitment and a

\textsuperscript{17} Model Rules of Prof’l Conduct R. 1.6 (1983).
\textsuperscript{18} Macey, supra note 7, at 324-25.
\textsuperscript{19} Simon, supra note 1, at 204-05; Simon, supra note 3, at 654.
\textsuperscript{21} Simon, supra note 3, at 654.
\textsuperscript{22} See, e.g., W. William Hodes, Rethinking the Way Law Is Taught: Can We Improve Lawyer Professionalism by Teaching Hired Guns to Aim Better?, 87 Ky. L.J. 1019, 1929 (1999).
“take no prisoners” style of practice. Clients don’t care about costs inflicted upon third parties, but they do care about prevailing in litigation or obtaining a good deal in a transaction. Why in the world would they ever opt for a high commitment lawyer?

The response to this objection is twofold. In litigated disputes, there are institutional players, such as courts and other tribunals, who look after the interests of third parties. These decision makers are likely to give more weight to the representations of high commitment than low commitment lawyers, because in their experience, high commitment lawyers engage in less tendentious readings of the factual record and the law, do not file motions merely to drive up their opponents’ costs, and generally urge the court to adopt a position that is not grossly inconsistent with the merits of the dispute. There are numerous close calls in pretrial litigation and at trial: motions that could go either way, discovery disputes in which the ethical conduct of the parties is an issue that must be decided by the judge, evidentiary issues decided on the fly by the judge. Although it would be almost impossible to gather empirical proof of this proposition, it stands to reason that a lawyer who is known for honesty and probity in dealing with the tribunal would do better in some of these close cases than a low commitment lawyer. In business transactions, if the assumption is that everyone is behaving in a self-interested manner, affected third parties will take costly steps to protect themselves. If another party in a transaction is dealing with a low commitment lawyer, she is likely to demand a thorough “due diligence” process or seek other kinds of assurances that the factual representations made by the lawyer are accurate, such as an “ex-
press 10(b)-5 warranty" that no material information has been withheld.26 A party dealing with a high commitment lawyer, on the other hand, may be more willing to accept that lawyer's representations at face value, without a lot of duplicative investigation or formal warranty documents.27 In both the litigation and transactional cases, a rational client would prefer the high commitment lawyer because that lawyer will either maximize the client’s chance of prevailing in the litigation or minimize the transaction costs associated with a deal.

In order for the market to reward high commitment lawyers, there must be a means for clients to discover lawyers’ commitment levels. The most straightforward method is for the client to have a history of dealings with the lawyer, in which they can observe both the lawyer’s level of commitment and the response of institutional decision makers or transaction partners to that lawyer. In a repeat-dealing relationship, if the client prefers a high commitment lawyer, the lawyer has an incentive to stick to her principles of working for her client within the constraints of justice, because she will lose the benefit of future employment by the client if she abandons them.28

Almost as good is a small community of lawyers and clients in which all the relevant actors know one another and have ample opportunity to verify the commitment levels of others through face-to-face interaction. Thus, even if a client has not employed a given lawyer in the past, she may have been opposite that lawyer in a transaction or litigated matter, and had a chance to deal extensively with the lawyer. Even where the two lawyers have not dealt with one another in the past, the length of their professional dealings with one another over the course of a discrete matter still provides each lawyer an opportunity to learn about the other’s tendencies and adjust her behavior accordingly.

Lawyers deal with one another in discrete but lengthy interactions that may be modeled as repeated games.29 They face numerous opportunities within a single deal or litigated matter to cooperate or defect from the cooperative solution, as various small conflicts arise

26. SIMON, supra note 1, at 205.
28. In the jargon of law and economics, the lawyer has a “relationship-specific prospective advantage” as a result of a history of dealings with the client, but risks forfeiting that advantage by ignoring the client’s wishes for a high commitment lawyer. David Charny, Nonlegal Sanctions in Commercial Relationships, 104 Harv. L. Rev. 373, 392-97 (1990).
29. Ronald J. Gilson & Robert H. Mnookin, Disputing Through Agents: Cooperation and Conflict Between Lawyers in Litigation, 94 Colum. L. Rev. 509, 520 (1994); see also SIMON, supra note 1, at 66-67, for a recognition of the arms race or prisoner’s dilemma structure of adversarial litigation.
over issues like scheduling and the logistics of discovery or due diligence. In a typical two-party situation, both lawyers would be better off adopting cooperative approaches at the outset, agreeing to reasonable requests for schedule changes, not filing unduly burdensome discovery requests or unfairly evading requests, behaving themselves at depositions, and so forth. Each risks that the other side will exploit her cooperativeness, however. One response is the familiar game theory strategy of “tit-for-tat”: cooperate on the first move, and then continue cooperating unless the other player defects, and then respond by defecting. Litigators report that they sometimes feel out their adversaries early in the process, to see if they are inclined to be cooperative, for example by “experiment[ing] with minor agreements with opposing counsel early in a case to assess whether they could be trusted throughout the proceedings.” A lawyer’s acquiescence in her opponent’s request for an accommodation is in effect a “reputational bond,” which secures her right to reciprocal cooperation from the opponent. As long as both lawyers cooperate, things operate much more smoothly and costs are kept down. If a lawyer gets burned, she forfeits the adversary’s bond by refusing to extend reciprocal courtesies in exchange, and the litigation becomes more expensive.

The information about lawyers’ tendencies to cooperate or defect may be pooled within the community of lawyers, through the legal press, judicial opinions, social interactions at bar association events and CLEs, and other informal gossip networks. A lawyer who has earned a reputation for flexibility, reasonableness, and honesty obtains the benefits of reciprocity from adversaries. Conversely, ac-

30. Not all lawyers may perceive this benefit from cooperation and, indeed, some lawyers may prefer pretrial litigation that is bogged down in extensive motions practice if they are billing by the hour. This qualification is supported by the report of some in-house lawyers, who observe “the economic interest of the hourly fee lawyer to ‘churn’ cases and to use discovery disputes to run up fees.” Sarat, supra note 23, at 830; see also Gilson & Mnookin, supra note 29, at 516-17. Simon’s reliance on the market to encourage high commitment lawyers is thus undermined by the way in which most lawyers bill for their services. I am grateful to Andrew Perlman for this point. Note, however, that it is still in the interests of each client for the lawyers to cooperate. The problem then becomes designing a mechanism so that clients can learn about the propensities of lawyers to cooperate or stir up disputes in order to churn fees and giving them effective and inexpensive sanctions to use against lawyers who run up bills. The difficulty in providing this information to potential clients is discussed infra notes 42-46 and accompanying text.


33. Charny, supra note 28, at 393.

34. See Lisa Bernstein, Private Commercial Law in the Cotton Industry: Creating Cooperation Through Rules, Norms, and Institutions, 99 Mich. L. Rev. 1724, 1756 (2001); Gilson & Mnookin, supra note 29, at 522-23. Similarly, some firms have acquired reputations for forthrightness and intolerance of gamesmanship. One example is Wilmer Cutler & Pickering, and in particular partner William McLucas, the former head of the SEC’s en-
quiring a bad reputation may mean that a lawyer’s opponents do not play tit-for-tat, but instead start out by assuming that the lawyer will be uncooperative.35 This drives up the cost for the client, one reason why Simon is right to argue that clients would prefer a high commitment lawyer in many cases.36 Unfortunately for these clients, information about a lawyer’s level of ethical commitment may be well known by other lawyers, but as a practical matter not accessible to clients. It is well understood that individual clients, who are generally one-shot players within the legal system, have difficulty assessing the competence of lawyers and bargaining over the structure of the lawyer-client relationship, as would be required in order to give effect to Simon’s proposal to permit clients to select from a menu of levels of commitment to which they could hold lawyers.37 Even highly sophisticated clients may not be able to tap into the information networks used by lawyers, however, particularly if they are not repeat players with respect to a particular geographic legal community or one defined by a specialized kind of practice. A large manufacturer of consumer products, for instance, may get sued in any state, and it is unlikely that it could keep adequately informed about the reputation of lawyers in each of these locations. Such clients also may have only episodic need for lawyers in some subspecialty, like environmental or securities law, while making more frequent use of product liability or patent lawyers. Perhaps the problem can be addressed by Simon’s proposal that lawyers be required to disclose their commitment level up front in negotiations with a prospective client,38 but it seems unlikely that the organized bar (which is, in Simon’s view, not interested in ceding any of its authority over professional regulation) will require an explicit discussion of commitment levels between lawyers and prospective clients. For this reason, highly visible actions that function as signals of commitment become even more important.

Signaling is a concept from game theory, which assumes that two strangers are interested in cooperating in order to achieve gains that

36. There may, of course, be cases in which the client is specifically looking for a low commitment lawyer. As Andrew Perlman pointed out, the tobacco industry relied for many years on law firms with a reputation for scorched-earth litigation to deter prospective plaintiffs from filing lawsuits. Cf. Gilson & Mnookin, supra note 29, at 516-17. Simon’s market-based response to the professional monopoly would do nothing to ameliorate this sort of unethical conduct by lawyers.
38. Simon, supra note 3, at 656-57.
they could not realize acting alone. Each party would like to know whether the other is inclined to cooperate, or to defect, if the possibility exists of realizing a greater gain by cheating than by maintaining cooperation. This disposition, however, is private information, not subject to verification by third parties. To put it another way, talk is cheap, so a low commitment lawyer might induce a client who is interested in the benefits of a high commitment lawyer to hire him, merely by professing to be a high commitment lawyer. Although talk is cheap, actions are costly, so a high commitment lawyer can take the kind of costly action that would be avoided by a low commitment lawyer. This action is the signal. An example of a signal might be joining a law firm with a reputation for probity or an organization whose admissions procedures are designed to screen out unethical applicants. The purpose of signaling is to bring high commitment lawyers together with clients who are seeking that type of lawyer (call them high commitment clients). High commitment lawyers will be able to gain a competitive advantage over their low commitment colleagues if the signals are effective and, crucially, if there is a sufficient number of high commitment clients to make it worthwhile to invest resources in actions that function as signals.

In order to function as a signal, though, an action must be clear and unambiguous, as well as difficult to mimic by low commitment lawyers. Belonging to Firm X is only a signal of being a high commitment lawyer if Firm X effectively screens out low commitment lawyers through its hiring process. Some research suggests that firm affiliation is not useful as a signal, because the conduct of lawyers within a firm is variable, as well as the actions of the firm in different cases. It is entirely plausible that Firm X may have two powerful partners, H.C. and L.C., who vary in their commitment to lawyering within Simon’s contextual view, paying due regard to the interests of third parties. In fact, in the ABA litigation section’s study of ethics of large firm litigators, plaintiffs’ lawyers in two cities were asked to name the three most and least ethical firms. The result was “only weak convergence,” with a wide disparity of votes, and some firms even appearing on both lists. As one of the study authors concludes, firms may have reputations for litigation style, but
these reputations can vary across cases. The same may be true for voluntary associations such as the American Academy of Matrimonial Lawyers, particularly if they do not sanction members for violating their high commitment standards. A member of the Academy may be a high commitment lawyer or a low commitment lawyer, but the fact of membership alone does not provide the necessary information. Contrast organizations like the cotton shippers’ associations, who expel members for refusing to comply with arbitration awards, or the trade associations who employ voting procedures and membership criteria designed to weed out untrustworthy applicants. Continuing membership in one of these associations is a signal at least of the willingness of the member to abide by the organization’s dispute-resolution procedures, which in effect means assent to a substantial body of trade norms and practices of fair dealing. The organizations Simon mentions, however, do not screen out low commitment applicants or sanction low commitment members, so low commitment lawyers may join in the hopes of duping high commitment clients into retaining them. The value of membership as a signal is considerably reduced if it does not enable the creation of a separating equilibrium in the market for legal services, in which high commitment lawyers and clients match up only with one another, and do not inadvertently match up with those of low commitment.

The discussion of signaling points in the direction of a significant complexity elided in the Lecture, namely the definition of a “high commitment” to ethics. This is certainly not the place to re-enter the
debate over the extent to which the lawyer’s role should take account
of the justice of the client’s position, the merits of the case, or the ef-
fect of the representation on third parties. Although I have a some-
what different view about the nature of the lawyer’s obligation to the
law and am not as enthusiastic a proponent of nullification as
Simon,51 I do agree that the lawyer’s moral agency requires in some
cases that she either seek to persuade the client to change her posi-
tion, withdraw, or accept moral responsibility for harms caused by
the client. The problem is that this approach, at best, serves only one
of several competing views within the academic legal ethics commu-
nity,52 and is decidedly a minority view among practicing lawyers.
Simon does not call his rhetorical adversary the “dominant” view for
nothing. Given the prevalence of the dominant view, it would not be
surprising to learn that many lawyers and law firms proclaim their
commitment to ethics in the sense of zealous client service. Perhaps
this is only a problem of labels, which could be solved by clarifying
our terms, but I fear that making ethical commitment the basis of
reputational markets and signaling behavior will only lead to confu-
sion. For example, what is the “high commitment” response to the
following cases?

(1) The defendant’s lawyer files a request for an extension of
time to answer a complaint, but through a clerical error it is misdi-
rected. The plaintiff’s lawyer, knowing of the mistake, has the op-
tportunity to file a motion for a default judgment.53 Does she do it?

(2) In a product liability case, the defendant’s lawyers define
the terms in a discovery request narrowly, to avoid producing an
inculpatory document. The plaintiffs do not object or file a motion
to compel.54 Is the defense lawyer’s conduct proper?

(3) A defendant is charged with armed robbery and has admit-
ted the crime to his lawyer. At the preliminary hearing the victim
testified that the crime took place at midnight, when the defen-
dant was (truthfully) playing cards with three friends, all of whom
have a good reputation in the community and will probably be be-

52. For powerful moral arguments that lawyers are permitted or required to take all
actions on behalf of their clients within the framework of law without regard to the justice
of their clients’ ends, see, for example, Monroe H. Freedman, Professional Responsibility of
the Criminal Defense Lawyer: The Three Hardest Questions, 64 MICH. L. REV. 1469 (1966);
W. William Hodes, Accepting and Rejecting Clients: The Moral Autonomy of the Second-to-
the-Last Lawyer in Town, 48 U. KAN. L. REV. 977 (2000); W. William Hodes, Lord
Brougham, The Dream Team, and Jury Nullification of the Third Kind, 67 U. COLO. L.
REV. 1075 (1996); Stephen L. Pepper, The Lawyer’s Amoral Ethical Role: A Defense, a Prob-
54. A slight modification on the facts of Washington State Physicians Insurance
lieved by the jury.55 Unfortunately, the victim was mixed up on the
time, probably because the defendant had hit him on the head in
the course of the robbery. (The crime actually occurred at 2:00
a.m.) The defendant is not going to take the stand. May the lawyer
call the friends to testify about the card game?

(4) A man whose lifelong dream has been to run a
restaurant persuades a millionaire cousin to lend him $100,000;
the man signs a demand note to the cousin. The restaurant is a
spectacular success, so the conniving cousin immediately calls the
note and brings an action on it, intending to acquire the restaurant
in a foreclosure sale. The man goes to a lawyer who, seeing little in
the way of defense on the merits, files a series of dilatory motions
to give the cousin time to pay back the note from revenues from
the restaurant. The motions are not frivolous in the Rule 11 sense,
but they are extremely unlikely to affect the result on the merits.
Is this a high commitment or low commitment stance with respect
to ethics?

Does the answer change if the client is a construction company
whose falling debris injured a single mother of three children who
worked as a housekeeper and is now permanently disabled, and
where the delay will put pressure on the plaintiff to settle early
and cheaply?56

Because the boundaries of the relevant context are contestable, I
suspect that there would be little convergence on a single answer to
each of these cases, even among lawyers who accept the basic out-
lines of the contextual view.57 Thus, high commitment lawyers
would have to figure out a way not only to signal that they “take ethics seri-
ously,” but that their understanding of ethics has a particular result
in a given case, which matches up with their prospective clients’ un-
derstanding.58 Again, I do not think it is implausible that a lawyer or
law firm might find this approach desirable, but to the extent
Simon’s proposals depend on information sharing, reputation, signal-
ing, and decentralized mechanisms of enforcement, clarity is of the
essence. Transaction costs would become prohibitive if lawyers and
prospective clients had to introduce all the relevant contextual fac-
tors into a discussion of ethical commitment levels.

http://www.michbar.org (last visited Jan. 12, 2003) (on file with Florida State University
Law Review).

56. These hypotheticals are taken from Stephen Gillers, Can a Good Lawyer Be a Bad

57. Robert W. Gordon, The Ethical Worlds of Large-Firm Litigators: Preliminary Ob-
servations, 67 FORDHAM L. REV. 709, 712-13 (1998) (stating that because ethical evaluation
calls for complex discretionary judgment, it is difficult to generalize).

58. Simon does acknowledge that the lawyer and prospective client would need to
flesh out what it means for the lawyer to promise to adhere to the “highest ethical stan-
dards.” Simon, supra note 3, at 655. I think he is correct to worry about the effect of this in-
formational asymmetry on decentralized enforcement mechanisms, but too optimistic
about the institutional means for alleviating the problem.
A committed contextualist client with sufficient bargaining power could always structure the attorney-client relationship contractually, so that the lawyer would be obligated to act in the interests of justice. Because of the difficulty in specifying what it means to practice contextually, or in the interests of justice, the transaction costs involved in this kind of approach appear daunting. However, Simon responds, in his book and in the Lecture, that this obstacle could be surmounted if lawyers and clients had a diverse stock of ethics rules from which to choose.\textsuperscript{59} Often, nonlegal mechanisms of regulation offer transaction-cost savings as compared with the legal system,\textsuperscript{60} but Simon is right to point out that institutional actors can nevertheless play a role in the process of decentralized regulation by, for example, providing a menu of contract provisions that parties may incorporate into their agreements.\textsuperscript{61} There is no particular reason to suppose that this menu must take the form of state disciplinary rules developed under a regime of competitive federalism—voluntary associations could perform the drafting function—although authoritative groups like the organized bar do have the salutary effect of collectivizing the costs of this project, preventing free-riding.\textsuperscript{62} The problem with this proposal is still disseminating the relevant information throughout the pool of prospective clients. Even assuming there is a substantial population of high commitment clients who are not seeking “attack dog” lawyers, they still must be educated that they can demand a high commitment lawyer, and that there are standardized provisions governing the lawyer’s conduct that they can insist be inserted in retainer agreements.\textsuperscript{63} This education campaign would be expensive, so only organized groups would be able to fund it, again because of free-rider problems that would make collective action by individual high commitment lawyers difficult. The dilemma for Simon, therefore, is that the only actors in the market for legal services who are capable of solving the problem of the organized bar’s corruption are exactly those actors who are alleged to be corrupt. The grassroots activism of high commitment lawyers and clients risks being swamped by the “noise” in the market created by the dominance of low commitment actors.

Finally, the contexts in which reputational markets and informal sanctions have shown promise are those in which there are a large number of potential transactors and the interactions between them

\footnotesize{\textsuperscript{59} Simon, supra note 1, at 207-08; Simon, supra note 3, at 657.  
\textsuperscript{60} See, e.g., Charny, supra note 28, at 403-08; Macaulay, supra note 35, at 63-65.  
\textsuperscript{61} Simon, supra note 3, at 656.  
\textsuperscript{62} Bernstein, supra note 34, at 1742-43; Charny, supra note 28, at 412.  
\textsuperscript{63} One way of doing this would be to adopt “commitment-forcing rules,” which require lawyers to inform clients that they have a choice of commitment levels for their lawyer. SIMON, supra note 1, at 211.}
are entirely voluntary. Lisa Bernstein has produced richly detailed studies of nonlegal regulation in the diamond and cotton industries, and in both cases, there are enough buyers and sellers that it is relatively costless to avoid doing business with someone in the industry who has a bad reputation. There are significant differences in the market for legal services. Most lawyer-lawyer relationships are not voluntary at all. A plaintiff’s lawyer cannot avoid suing the entity responsible for her client’s injury just because it is represented by the Low Commitment Law Firm. The only thing the lawyer can do in that case is protect herself by not attempting to cooperate early in the relationship. But this only protects the client from the costs of having her lawyer’s cooperation exploited at some later time in the lawsuit, not from the greater costs of the lawyers’ inability to cooperate from the outset. A larger number of client-lawyer relationships are voluntary, at least in some areas of practice, like commercial litigation and criminal defense, although there are some highly specialized practices in which an oligopolistic market for lawyers’ services exists. With respect to one-shot litigants, however, the episodic nature of their interaction with lawyers and the social differentiation of these populations of lawyers and clients make it extremely difficult to acquire the information necessary to seek out high commitment lawyers (if that is their desire) and to structure the legal relationship with their lawyer accordingly.

III.

Simon worries very much about the corruption of formal entities such as the organized bar, which is beholden to its economically self-interested members. The dismal record of the state bar associations in controlling unethical behavior by lawyers is almost taken for granted, and there seems to be no reason to quarrel with Simon’s assessment on this point. Despite powerful critiques by Simon and others, lawyers as a whole do not seem terribly interested in finding an alternative to the dominant view, in which they can sell zealous advocacy to clients and rest easy that they will not be held morally accountable for their clients’ ends. Nevertheless, the hope expressed

66. See also Richard L. Abel, American Lawyers 156-57 (1989); Deborah L. Rhode, In the Interests of Justice 158-65 (2000).
in Simon’s Lecture is that nonlegal regulation will be less susceptible to capture by lawyers who are not committed to practicing law in the interests of justice. At the risk of sounding an unduly pessimistic note on this hopeful occasion, I wonder whether the informal, non-state-centered mechanisms he favors are any more insulated from the pressures of lawyer self-interest than the formal, monopolistic, profession-based regulatory regime he attacks.

The reason for this pessimism is not related only to the persistence of the dominant view, the hired gun ethic, the ideology of advocacy, or whatever one wants to call it, among practicing lawyers. Rather, Simon’s alternative is vulnerable to capture by self-interested lawyers by reason of the structure of its enforcement mechanism. As noted above, reputational markets, gossip, informal means of retaliation, and decentralized control mechanisms work better among parties who are repeat players with respect to each other. In a small town, all the lawyers may have incentives to cooperate with one another; in larger cities there may at least be specialized bars in which lawyers follow norms of reciprocal fair dealing with each other. The situation becomes more complicated when we introduce the lawyer-client relationship. A lawyer may be a repeat player with respect to judges and other lawyers, but only a one-shot player with respect to the client. In addition, the client may herself be only a one-shot player with respect to the litigation system. This asymmetry gives rise to a “confidence game” structure, in which the lawyer’s primary loyalty is to other institutional players, who assist her in duping the client into believing that the client is receiving zealous representation. As Marc Galanter puts it, the “real” clientele of a lawyer who represents individuals who have only episodic contact with the legal system are the other institutional actors. Good relationships among, for example, criminal defense attorneys, prosecutors, judges, and court personnel such as clerks and bailiffs, are essential in order to facilitate the resolution of cases through plea-bargaining. Furthermore, in order to do well in her own career, a defense lawyer must have access to courthouse personnel who can cut her client breaks. By obtaining this access, the defense lawyer

70. See Gilson & Mnookin, supra note 29, at 512-13, 551.
72. Galanter, supra note 37, at 117.
73. Blumberg, supra note 71, at 18-22.
becomes a “fixer” who has a valuable commodity—access—to sell to clients.\textsuperscript{74}

A profound tension thus exists between the procedural entitlements of the client and the systemic interest in the efficient processing of cases. To put it unkindly, criminal defense lawyers sell out some of their clients in the interests of procedural justice and of long-term economic success. To make the same point more sympathetically, if each client received what he was entitled to under law—a full, adversarial trial on the merits—the machinery of criminal justice would grind to a halt. As officers of the court, lawyers have some obligation to facilitate the fair and efficient processing of disputes. The alternative to accepting some of this responsibility is the ethic of zealous advocacy within the bounds of the law that Simon decries. After all, the reason he criticizes lawyers who take formally non-frivolous but substantively unwarranted legal positions is that they are neglecting their responsibility to facilitate the just resolution of disputes.\textsuperscript{75} It may be objected, of course, that a lawyer in Simon’s contextual view is not permitted to “sell out” her client’s interests merely to enhance the lawyer’s reputation with other courthouse regulars. We need not take this uncharitable view of settlement in all cases, however. Some settlements may be just, if they are calibrated to the value of the plaintiff’s claims, discounted for the probability of a defense verdict at trial.\textsuperscript{76} Moreover, Simon explicitly includes the effective functioning of the system within his definition of justice, and charges contextualist lawyers with making the procedural apparatus as efficient as possible, consistent with deciding cases on the legal merits.\textsuperscript{77}

To the extent that this Lecture can be understood as a brief for shifting some of the responsibility for regulating the profession onto nonlegal mechanisms, the problem noticed by Blumberg becomes more acute. Simon argues that reputational markets should encourage lawyers to increase their level of ethical commitment, because clients would seek a high commitment lawyer over an attack dog in order to benefit from the lawyer’s good reputation.\textsuperscript{78} In litigation, one way to acquire a good reputation is to acquiesce to reasonable requests by opposing counsel, not annoy judges unduly, and generally refrain from rocking the boat. “[A] tribunal might reward lawyers who appear to adhere to high-commitment ethics with procedural accommodation or more ready acceptance of their representations.”\textsuperscript{79}

\textsuperscript{74} Id. at 25-26.
\textsuperscript{75} Simon, supra note 1, at 140-41.
\textsuperscript{76} Id. at 141.
\textsuperscript{77} Id. at 143.
\textsuperscript{78} Simon, supra note 3, at 651-52.
\textsuperscript{79} Simon, supra note 1, at 205.
Everything works beautifully if the judge, other lawyers, and assorted courthouse regulars, like clerks, are interested in doing justice, in the sense of deciding cases on their merits. The trouble is that the tribunal and all the supporting players may be no more interested in justice than the dominant-view lawyers Simon criticizes. Or, to put the point less harshly, other institutional actors may act under a Weberian, bureaucratic conception of justice, emphasizing docket-clearing, expeditious handling of cases through settlements and plea bargains, and efficiency rather than resolution of cases after careful consideration of the parties’ factual and legal contentions. In such a system, lawyers’ incentives would be largely as Blumberg described them—to contribute to the realization of administrative values such as the saving of cost, time, and labor.

In The Practice of Justice, Simon recognizes the alienating effect that the rationalization of professional practice through categorical norms of ethics has on lawyers, and argues for the contextual view as a way of returning meaning to lawyers’ work. By vesting lawyers with the autonomy to make flexible, substantive decisions regarding the application of law to their clients’ cases, the contextual view returns artisanal values like personalization, judgment, and craft (i.e. non-fungibility of one’s work product) to the practice of law, as well as connecting lawyers’ lives more directly with the pursuit of social justice. His diagnosis of alienation is perceptive, and he may be right to look to the idealism of law students and young lawyers as a source of hope for the renewal of the profession’s aspirations. As noted at the outset, however, this idealism makes an uncomfortable fit with enforcement mechanisms that depend so critically on the reputation of lawyers within the professional community. If professional norms become differentiated into low commitment and high commitment, the only way to avoid a race to the bottom is to make high commitment strategies more beneficial to clients, which in turn is plausible only if the consequences of adopting a low level of ethical commitment are visited on the clients. Unfortunately, the most likely mechanism for creating this feedback effect creates divided loyalties on the part of lawyers, who must please other institutional actors whose commitments may not be to substantive justice. This means that the lawyer may have an incentive to sell out the short-run inter-
ests of one client in favor of her own long-run interests in maintaining good relationships with the other actors.

Perhaps I am exaggerating the problem because I have unwittingly subscribed to the dominant view position that the lawyer ought to be primarily loyal to her client, not to the system of justice generally.84 One person’s “double agent”85 is another’s Brandeisian lawyer for the situation.86 Nothing in Simon’s Lecture is inconsistent with the ideal of loyal client service. His point is rather that a lawyer provides better client service by realizing that opposing counsel and courts are interested in achieving a cooperative solution to various problems, such as discovery abuse and misrepresentations in negotiation, and assuming a cooperative posture with respect to these other institutional players. I do think, though, that the tension between procedural and substantive justice is not likely to go away if the professional monopoly is dissolved. The market, no less than the organized bar, is a rough instrument for regulating lawyers’ behavior. Simon’s categorical view is a response to the tension between procedural and substantive justice, and he envisions trade-offs between the two conceptions of justice as the merits of a particular representation require.87 Precisely because legal ethics is so contextual, these subtle distinctions cannot be translated into clear, unambiguous signals of a lawyer’s level of ethical commitment, which can be used by participants in the market for legal services. In the end, Simon may not be able to have it both ways, with a flexible, contextual, case-by-case approach to lawyers’ ethics and an enforcement mechanism that depends on clarity and the absence of ambiguity.

Depending on one’s attitude toward the rhetoric of professionalism, whether one regards it as self-serving window dressing or as genuine ideals to which lawyers ought to respond, Simon’s embrace of nonlegal regulation is either regrettable or welcome. Authentically high commitment lawyers presumably adopt this level of commitment because they believe it is morally obligatory, not because they believe it will help them attract clients. On the other hand, even morally ambitious lawyers have to eat and pay the mortgage, and it would be comforting for them to expect that their ethical commitments will not result in losing a race to the bottom. There is a long tradition in moral philosophy of trying to show that doing the right thing is ultimately conducive to one’s happiness. This Lecture attempts to show that the marketplace will encourage a race to the top,

84. See generally William H. Simon, The Ideology of Advocacy: Procedural Justice and Professional Ethics, 1978 Wis. L. Rev. 29 (arguing that even critics of the adversary system tend to accept the ideology of advocacy without question).
85. Blumberg, supra note 71, at 39.
86. Simon, supra note 1, at 127-32.
87. Id. at 139-40.
if only it could be freed from the strictures of the organized bar's low commitment rules. It is an extremely ambitious and hopeful, but not utopian, claim. Simon does not think we are all angels, only that our self-interested natures may drive us in the direction of beneficial cooperation.88 A reader looking for more idealism will be disappointed in this stance, but what we know about lawyers' ethics in practice suggests that any serious proposals for reform will have to contend with the pervasiveness of self-interest and the erosion of professional ideals.89 If morality is all about striking the appropriate balance between values, then Simon is to be commended for recognizing the balancing act inherent in professional regulation, and offering a proposal for reform that attempts to get it right.

89. "As market forces take on greater and greater salience and the competitiveness of the legal market increases, the very meaning and viability of professionalism as an ideal guiding law practice is being called into question." Sarat, supra note 23, at 811.
RATIONALIZING DRUG POLICY
UNDER FEDERALISM

DAVID W. RASMUSSEN* & BRUCE L. BENSON†

I. INTRODUCTION

Legal systems and economic organizations have a common purpose because they exist to create incentives and constraints that modify human behavior. The field of law and economics is an intel-

* James H. Gapinski Professor of Economics and Director of the DeVoe L. Moore Center at Florida State University. A.B., Earlham College, 1964; M.A., Washington University, St. Louis, 1967; Ph.D., Washington University, St. Louis, 1969.
† DeVoe L. Moore and Distinguished Research Professor of Economics at Florida State University. B.A., University of Montana, 1973; M.A., University of Montana, 1975; Ph.D., Texas A&M University, 1978.
lectual niche that rests on the assumption that, in the common law at least, economic efficiency is the principle that guides legal reasoning. But some legal rules appear very inefficient, and economic analysis can also be employed to see why the incentives and constraints emerging from laws do not promote their stated purpose. Such analysis often ferrets out unintended consequences of rules and regulations.

This Article uses economic analysis to examine the consequences of legal incentives and constraints that are designed to curb the use of illicit drugs. The analysis indicates that U.S. drug policy, to the extent that it shrinks these illicit markets, does so at an enormous financial burden and also generates many other unintended costs. We examine why such a policy persists and argue that, in a federalist system, devolution of policy making from the national government to local jurisdictions is required for a more rational drug policy. In this introduction, we identify the areas of economics that are especially pertinent to this analysis and describe how the subsequent argument was developed.

Following the publication of Gary Becker’s economic theory of crime and Issac Ehrlich’s controversial empirical tests of that theory, a large amount of theoretical and empirical literature developed very rapidly, testing, extending, and criticizing the use of rational decision-making models to predict criminal behavior. Similarly, fol-

The authors wish to acknowledge Daniel Maier-Katkin, Iljoong Kim, Kathy Makinen, Will H. Moore, Andrew Morris, and Matthew Schultz for helpful comments on previous versions of this paper.

lowing the publication of books by James Buchanan and Gordon Tullock, Mancur Olson, and William Niskanen that present economic models of government institutions, there has been an explosion in the literature using or criticizing the use of the economic, or “public choice,” model to analyze decision-making in the public sector. With few exceptions, developments in the economics of crime and public choice have not overlapped. However, they are now being brought together in the rapidly growing economics literature that analyzes the causes and consequences of illicit drug policy.


9. This “public choice” literature is very large as the approach has been adopted by a large portion of political science scholars as well as economists who focus on the determinants of public policy. A number of journals are devoted to the approach (e.g., PUBLIC CHOICE, CONSTITUTIONAL POLITICAL ECONOMY, and THE JOURNAL OF PUBLIC FINANCE AND PUBLIC CHOICE), so the literature continues to grow rapidly. For useful reviews, see DENNIS C. MUeller, PUBLIC CHOICE II: A REVISED EDITION OF PUBLIC CHOICE (1989); Bruce L. Benson, Understanding Bureaucratic Behavior: Implications from the Public Choice Literature, 13 J. PUB. FIN. & PUB. CHOICE 89 (1995); Gary J. Miller, The Impact of Economics on Contemporary Political Science, 35 J. ECON. LITERATURE 1173 (1997); and the papers in THE ELGAR COMPANION TO PUBLIC CHOICE (William F. Shughart II & Laura Razzolini eds., 2001). A similar “rational choice” perspective has nascent roots in sociology as reviewed in Michael Hechter & Satoshi Kanazawa, Sociological Rational Choice Theory, 23 ANN. REV. SOC. 191 (1997).


The economic analysis of illicit drug policy considers the incentives and constraints that affect the behavior of drug users, drug suppliers, employees of the criminal justice system, and treatment providers in order to understand how alternative policies affect drug use in particular and drug markets in general. Furthermore, drug


policies, like all public interventions that affect incentives and constraints, can have unintended consequences that potentially offset their intended purposes.\textsuperscript{13} Much of the drug policy debate focuses on extreme arguments of legalization and punitive prohibition commonly referred to as the “drug war,” but there is a long continuum between these alternatives that has not been adequately examined from the analytical perspective offered here.\textsuperscript{14} This Article analyzes the current policy of drug prohibition from an economics perspective based on measurable costs and benefits, rather than entering into

\textsuperscript{13} Examples abound in the economics literature. See, for example, JAMES D. GWARTNEY ET AL., ECONOMICS: PRIVATE AND PUBLIC CHOICE 101 (9th ed. 2000), for classic examples involving minimum wage legislation causing unemployment and changing the nature of compensation to low wage employees. \textit{Id.} at 101-03. This source also provides an example of a wide variety of public insurance schemes that increase the prevalence of risky behavior due to moral hazard problems. \textit{Id.} at 346.

\textsuperscript{14} Much of this middle ground can be claimed by people who advocate “harm reduction.” Harm reduction is viewed by some drug war advocates as an ill-disguised argument for the legalization of drugs. Although some advocates of harm reduction may, in fact, have such views, here we accept that the case for legalization of drugs can be made without the prior condition that it also minimizes the harms of drug use. Advocates of legalization, such as THOMAS SZasz, \textit{Our Right to Drugs: The Case for a Free Market} 8-11, 13-20 (1992), place an \textit{a priori} value on the right of individuals to consume any substances when they are not harming others. While this view is obviously held to be consistent with harm reduction in the minds of its advocates, in this discussion, harm reduction explicitly excludes the benefits accruing to people when they are allowed to exercise what some may consider their natural rights to consume what are now illicit substances. The arguments presented here focus on observable consequences of drug policy in order to conceptually clarify the benefits and costs of illicit drug policy.

There is no consensus about the role of prohibition in a harm reduction strategy. Diane Riley et al., \textit{Harm Reduction: Concepts and Practice. A Policy Discussion Paper}, 34 \textbf{SUBSTANCE USE & MISUSE} 9, 12 (1999), define harm reduction as a pragmatic program that balances benefits and costs and which also respects “[t]he dignity and rights of the drug user.” \textit{Id.} Since drug users in this view have standing in the benefit calculation, they argue that “harm-reduction strategies would \textit{not} include strategies such as abstinence-oriented treatment programs or the criminalization of illicit drug use.” \textit{Id.} at 11. In contrast, consider the position of the UN International Drug Control Programme (UNDCP): “The principles of harm reduction . . . are sometimes confused with those of legalization . . . . Many of the arguments put forward on behalf of harm reduction may also be applied through a flexible interpretation of prohibition.” See UNDCP, \texti{WORLD DRUG REPORT} 188 (1997). The latter interpretation is adopted for the purposes of this Article.

R.J. MACCOUN & PETER REUTER, \textit{Drug War Heresies: Learning from Other Vices Times, and Places}, 318 (2001), point to an important ambiguity in the notion of harm reduction by making a distinction between “macro harm” and “micro harm.” Macro harm is defined to be the product of the number of users, average drug consumption, and the average harm per unit of consumption (micro harm). Believing that any drug consumption generates significant harms to users will generate very high estimates of macro harms even when the average user consumes very little, thereby logically suggesting a case for prohibition. A common perspective among advocates of drug policy reform, on the other hand, is that few user harms are generated by casual use; thereby they view macro harms as being comparatively small. The ensuing analysis suggests that these distinctions are systematically distorted in policy discussions due to the institutional context in which drug policy is conducted.
the normative debate that focuses on other important but intangible values.\textsuperscript{15}

The argument for a decentralized drug policy is developed in the following way. Section II considers the intended benefits of criminalizing drug use and enforcement activity. It shows how enforcement can increase the price of illicit drugs and therefore reduce some potential harms by lowering the incidence and frequency of drug use. This approach is also costly, however. Thus, its actual effectiveness must be considered and compared to its costs, many of which are unintended consequences attributable to the enforcement process. These issues are also investigated in this Section. The focus on the consequences of illicit drug enforcement continues in Sections III and IV. Since enforcement may be more effective against some drugs than others, Section III shows that the resulting changes in relative prices may shift drug consumption patterns in ways that may or may not be consistent with a goal of reducing the harms of drug abuse. In Section IV, we explore the impact of enforcement on the supply of illicit substances and examine additional unintended consequences of drug enforcement. The relationships among drug use, drug enforcement policy, and public safety are explored in Section V. The implication of this examination of drug enforcement is that the current level of criminal justice involvement is clearly excessive. Therefore Section VI, entitled “Suppliers of Drug Policy: The Role of Bureaucratic Self-Interest,” explains how the federal government creates incentives for the excessive enforcement at the state and local level. The implications for drug policy reform, which are rooted in the proposition that we should decentralize drug policy in our federalist system, are explored in Section VII, which is followed by our conclusion that such institutional reform is a prerequisite for a rational drug policy.

II. THE INTENDED BENEFITS OF CRIMINALIZATION AND ENFORCEMENT: ARE THEY EFFECTIVELY ACHIEVED?

Enforcement policy can discourage drug use by raising the “full” price of drugs, and in doing so, it is assumed to reduce the harms of drug use.\textsuperscript{16} Full price refers to all costs borne by consumers of illicit

\textsuperscript{15} Issues of civil and economic liberties are undeniably important in this debate, for instance, but the current Article analyzes the efficiency of drug prohibition policies without consideration of the more fundamental issues of individual rights with respect to drug use. These issues are discussed in Douglas N. Husak, Drugs and Rights (1992). Objection rather than moral harms will be the focus of our drug policy analysis because any morals-based policy debate inevitably involves competing and highly subjective values and rights, such as sobriety, social order, freedom, and responsibility. A classic debate about social values competing with individual rights was conducted by Patrick Devlin, in The Enforcement of Morals (1965), and H.L.A. Hart, in Law, Liberty, and Morality (1963).

\textsuperscript{16} If users get utility from consuming drugs, then, properly construed, the net harm reduction generated by enforcement would be calculated by subtracting the consumer
drugs: the money price paid for the product as well as the increased costs of searching for the product and risks of toxicity, arrest, incarceration, and violent altercations in these illegal markets.\textsuperscript{17} This Section considers the anticipated benefits of drug enforcement and discusses some of the costs, both anticipated and unanticipated, that are likely to be associated with these efforts.

A. Reducing Drug Consumption

First, consider the monetary component of full price. Given the demand for illicit substances, drug enforcement reduces drug use if targeting suppliers raises the money price of illicit drugs, and drug consumption falls as a consequence of these higher prices. Drug enforcement targeted against drug suppliers most assuredly raises the money price of what are otherwise quite ordinary agricultural products,\textsuperscript{18} so the effectiveness of enforcement designed to reduce the harms of drug use depends, at least in part, on the responsiveness of drug users to monetary prices.

At times, in the public debate, demand for drugs has been characterized as being perfectly inelastic, meaning that drug users purchase the same amount of drugs irrespective of changes in the money price. The scientific evidence overwhelmingly demonstrates that the quantity of drugs purchased falls at higher drug prices, so enforcement directed against suppliers of illicit drugs will almost certainly reduce drug use in the aggregate, all else being equal.\textsuperscript{19} For instance, the case for claiming that drug consumption does not vary with changes in money price would seem to be strongest among addicts, who will supposedly do anything to avoid the pains of withdrawal. Addicts can be highly motivated to avoid withdrawal pain and still be

\begin{small}
benefits from the harms eliminated by the enforcement effort. For the sake of this argument, it is assumed that any user benefits are not pertinent to the analysis—that is, consumers do not have standing in this benefit-cost analysis. See William N. Trumbull, \textit{Who Has Standing in Cost-Benefit Analysis?}, 9 J. POL'Y ANALYSIS & MGMT. 201-02 (1990). This obviously raises important questions about the appropriate role of government paternalism, and, more immediately, clearly biases this analysis in favor of enforcement in drug policy. Not surprisingly, denying users standing in the benefit-cost analysis of drug policy is clearly implicit in the arguments made in support of a drug war.

\textsuperscript{17} Moore, \textit{Policies}, supra note 12, at 270, introduced this notion as “effective price.”

\textsuperscript{18} Jonathan P. Caulkins & Peter Reuter, \textit{What Price Data Tell Us About Drug Markets}, 28 J. DRUG ISSUES 593, 594-95 (1998), indicate the success of enforcement against suppliers when they report that marijuana prices per ounce vary from $140 to $1,000, depending on quality. This puts marijuana on par with gold, which sells for about $300 per ounce. MACCOUN & REUTER, supra note 14, at 344 (estimating that a highly taxed cigarette costs about 10 cents, a small fraction of the cost of a typical $4 marijuana joint); Moore, \textit{Supply Reduction}, supra note 12, at 124 (estimating that the cost of illegal heroin is about 70 times its estimated price under legalization).

\textsuperscript{19} This claim that drug consumption does not respond to price has been thoroughly discredited by the literature reviewed in RASMUSSEN & BENSON, \textit{Economic Anatomy}, supra note 11, at 45-49, and Caulkins & Reuter, supra note 18, at 604-05.
\end{small}
responsive to price, of course, if “average daily consumption . . . significantly exceeds the amount needed to avoid withdrawal pain,” and this is often the case. Kaplan further undermines the argument that addicts are insensitive to changes in the price of heroin even when withdrawal pain is involved because they voluntarily abstain from use for substantial periods for a variety of reasons, one of which is that they cannot afford to buy the drug. Thus, striking evidence suggests that enforcement lowers drug use by increasing illicit drug prices, and therefore it may reduce the harms associated with drug use. This effect does not necessarily justify enforcement, however, because imposing heavy taxes on legalized drugs might accomplish the same purpose.

Policy initiatives also can reduce demand, which means that consumers will buy fewer drugs at any price offered in the drug market, perhaps by means of education or through the criminal justice

20. RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 46-47. See generally MOORE, BUY AND BUST, supra note 12; Roumasset & Hadreas, supra note 12.


22. A policy of imposing very high taxes after legalizing a drug also would have negative consequences, of course. Such a policy would stimulate black markets to avoid the taxes, for example, which in turn would require increasing enforcement or repeal of the taxes. For instance, after Quebec imposed high taxes on cigarettes, smuggling became so rampant that the black market was estimated to account for almost half of all cigarettes consumed. In 1994, taxes were rolled back to end the smuggling. Benson & Rasmussen, Predatory Public Finance, supra note 11, at 197-98.

23. The discussion to this point has assumed that drug enforcement has only affected the supply of drugs but that the demand for drugs is unaffected by these policies. In economic terms, we have assumed that the supply curve shifted to the left and caused a decline in the quantity of illicit drugs demanded as the supply curve moved along a stable demand curve.

24. RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 71-72 (arguing that drug markets are best analyzed using the convention that the demand for drugs is captured as a relationship between money price and quantity demanded and that other elements of full price shift this relationship). Thus, an increased probability of being arrested for the possession of drugs causes a downward shift in the demand relationship between money price and quantity demanded.

25. Education programs that increase awareness of the dangers of drug use are an obvious mechanism to reduce demand. If young people could be induced to "just say no," the entire problem of drug abuse would be moot. Experience suggests that this is unlikely. Project DARE (Drug Abuse Resistance Education), which is used in schools throughout the nation, has few long-term effects on its participants. See Joel H. Brown, Youth, Drugs and Resilience Education, 31 J. DRUG EDUC. 83 (2001); Donald R. Lynam et al., Project DARE: No Effects at 10-year Follow-Up, 67 J. CONSULTING & CLINICAL PSYCHOL. 590 (1999); Dennis P. Rosenbaum & Gordon S. Hanson, Assessing the Effects of School-Based Drug Education: A Six-Year Multilevel Analysis of Project D.A.R.E, 35 J. RES. CRIME & DELINQ. 381 (1998). There is some evidence that more carefully crafted school-based programs with adequate "booster" sessions can be effective. See Gilbert J. Botvin et al., Long-Term Follow-Up Results of a Randomized Drug Abuse Prevention Trial in a White Middle-Class Population, 273 JAMA 1106 (1995); Richard Midford, Does Drug Education Work?, 19 DRUG & ALCOHOL REV. 441 (2000).
system by punishing possession, increasing the risks of consuming adulterated drugs or other health dangers such as hepatitis and AIDS, increasing the time spent searching for drugs, or increasing the probability of drug consumers being robbed when trading in illegal drugs.

B. The Elusive Benefit of Lowering Drug Use via Enforcement

Since drug policies are directed at reducing consumption either by reducing the quantity demanded through higher prices or lowering demand by raising other costs associated with consumption, it is appropriate to consider the potential benefits of reduced consumption from an economic perspective. There actually are two economic perspectives on this question, however. The standard treatment of this issue in modern economics is to presume that individual decision-makers are rational maximizers of subjective utility, so public policy should intervene only if their behavior has an adverse impact on others. From this perspective, drug consumption should be discouraged only when it generates what economists call “negative externalities.” In contrast, an emerging literature in behavioral economics tests the underlying assumptions of traditional economics and finds that people systematically make mistakes that could lead them to consume

26. The probability of arrest and conviction is usually regarded as a more effective deterrent than the severity of punishment in the economics of crime literature, although this implies that the marginal offender prefers more risk to less. See Becker, supra note 3. The empirical evidence regarding violent and property crimes suggests that apprehension is a more effective deterrent than longer sentences. See Isaac Ehrlich, Crime, Punishment, and the Market for Offenses, J. ECON. PERSP., Winter 1996, at 43, 46; Kim et al., supra note 12 (finding similar evidence regarding deterrent effects among drug offenders).

27. The fact that drugs are bought and sold in illegal markets means, of course, that consumers have no legal recourse when sold unsafe goods. Reputation can be valuable to suppliers, even in illegal markets, when repeat business is desired, but within this context greater enforcement against suppliers can disrupt established trading relations between buyers and sellers with the result that consumers face a greater risk of purchasing adulterated drugs.

28. Mark A.R. Kleiman, Against Excess: Drug Policy for Results 137-38 (1992), argued that longer search time is a particularly desirable aspect of enforcement because it reduces consumption without raising prices, which might be associated with higher crime and greater burdens on relatives. Rasmussen and Benson dispute this contention because enforcement will most effectively disrupt the markets catering to casual users; habitual users, on the other hand, are less likely to experience supply disruptions. RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 72-73. Moreover, if enforcement makes it more difficult to find alternative suppliers, drug suppliers will have more market power to raise prices.

29. Neighborhoods where drugs are regularly sold experience a relatively high rate of robbery because drug market participants carry either cash or drugs, and, when victimized, are not prone to report the crime to the police. See Paul J. Goldstein, Drugs and Violent Crime, in PATHWAYS TO CRIMINAL VIOLENCE 16, 35 (Neil Alan Weiner & Marvin E. Wolfgang eds., 1989). Such robberies only become official crime statistics when they involve sufficient violence to require medical treatment.
drugs even though they later regret that behavior. This perspective can be used to argue that drug policy should discourage consumption even if drug use does not generate any negative externalities. We consider the traditional justification for drug policy first.

Raising the price of drugs by more stringent enforcement can have unintended consequences that cause some of the negative externalities that are correlated with drug use. If the demand for illicit drugs is relatively inelastic among persons who abuse these substances, for instance, total expenditures rise with a price increase because the percentage decline in quantity consumed is lower than the percentage change in price. Thus, a drug addict chooses to sacrifice other goods (such as food or health care) in order to pay the higher drug price. Heavy drug users may choose to get more money to spend on drugs rather than sacrifice consumption of other goods, however. It is frequently alleged that the unintended consequence of higher enforcement is more crime and family deprivation if relatives willingly or unwillingly help pay for a drug addict’s purchases. Whether drug enforcement reduces the harms of drug use by raising prices will,

30. Chris Starmer, *Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice Under Risk*, 38 J. Econ. Literature 332, 377 (2000), claims that “the data we have suggests that choice behavior displays complex patterns in even very simple contexts.” This literature suggests that for many aspects of market behavior the traditional behavioral assumptions in economics are appropriate, but some economists observe that social context is probably a very important factor in many choices regarding school attendance, criminality, and labor force participation. See, e.g., George A. Akerlof, *Social Distance and Social Decisions*, 65 Economica 1005 (1997); Edward L. Glaeser et al., *Crime and Social Interactions*, 111 Q.J. Econ. 507 (1996).

31. The price elasticity of demand is the percentage change in quantity demanded divided by the percentage change in price, and in theory, it measures the consumer’s response to a price change holding all other relevant factors constant, particularly income and prices of related products. Therefore inelasticity of demand does not imply that addicts must increase property crime as is often claimed. It does imply, however, that they are willing to sacrifice other goods to get the drug. Caulkins & Reuter’s, supra note 18, at 604-06, review of the empirical literature reveals that estimates of the price elasticity of demand are much higher in absolute value than generally thought since several studies report elasticities exceeding -.10 for heroin and cocaine. Henry Saffer & Frank Chaloupka, *The Demand for Illicit Drugs*, 37 Econ. Inquiry 401, 408 (1999), estimate price elasticities of -.28 for cocaine and -.94 for heroin for survey respondents likely to be representative of occasional drug users. A price elasticity of -.94 means that a 1% increase in price will reduce quantity demanded by .94%. All estimates must be viewed with caution given inevitable uncertainties about the quality of the data in these studies, but they nevertheless suggest that the quantity of drugs consumed responds to price.

32. The relationships between drug use, enforcement efforts, and crime are discussed in Section V.

33. A survey of jail inmates reporting sources of income by drug use history revealed that among persons who had used drugs daily during the month prior to their incarceration, almost two-thirds had wage and salary income, 21% received legal benefits, 22% got money from family and friends, and 29% engaged in illegal activity. CAROLINE WOLF HARLOW, U.S. DEPT OF JUSTICE, BUREAU OF JUSTICE STATISTICS SPECIAL REPORT: DRUGS AND JAIL INMATES, 1989 (Aug. 1991). Families of drug users can be directly affected by providing money to buy drugs and indirectly by having less financial support from the drug user who is now dedicating more income to pay for higher-priced drugs.
therefore, depend on how the behavior of drug users is altered by price changes.

Although research on the price elasticity of demand for even hard drugs suggests that the quantity demanded is quite responsive to changes in money price, this market elasticity may not reflect the behavior of those who most seriously abuse these substances. It is at least plausible that some of the most severely addicted individuals will try to minimize changes in their consumption when enforcement increases drug prices, so this policy for these persons might actually increase expenditures on drugs.34 Raising drug prices in such circumstances could result in more property crime to finance drug purchases and/or increased deprivation among families of addicts. Thus, when drug policymakers justify enforcement because drug users are hopelessly addicted, the policy they advocate may, in fact, be aggravating, rather than mitigating, some of the negative externalities they are seeking to remedy. As we will see in Section V below, the preponderance of evidence suggests that excessive drug enforcement is more likely to foster crime than increase public safety.

If drug policy is to seriously consider the impact of drug use on the relatives of users, the interests of drug users cannot be totally ignored in the cost-benefit assessment of drug policy. Policies that incarcerate casual drug users, foster increased toxicity of drugs, increase the chances that users will be victims of crime, and make it difficult for drug market participants to hold legitimate jobs obviously impose costs on both the users and their families.35 Thus, while drug policy might choose to ignore the benefits users derive from drug consumption, the harms drug policy causes drug users and their families should be included in the benefit-cost calculus.36

Some economists have explored the possibility that drug users are rational consumers who will recognize that they might not be able to constrain their consumption in the future, so if the price is high due to enforcement, they will refrain from initiating the use of such goods.37 If addiction makes people more short-sighted, a change in

34. While this is a possible outcome, we show in the following sections that people who spend a large portion of their resources on a particular commodity are expected to be among the most, not least, responsive to price changes.

35. Punishment supposedly is designed to lower the harm of drug use, but it may also have the socially undesirable effect of causing drug users to be taken out of the labor force during their incarceration. To the extent that this loss of experience compromises future earnings, there are long-term private and social costs that should probably not be ignored when analyzing these policies. But see Jeffrey Grogger, The Effect of Arrests on the Employment and Earnings of Young Men, 110 Q.J. ECON. 51 (1995) (suggesting that these long-term effects may be modest).

36. Advocates of unconditional war on drugs appear to want to count the benefits of saving some people from drug use but are unwilling to consider the costs these policies impose on those they fail to save.

37. See, e.g., Becker & Murphy, supra note 12, at 675-76.
preferences that will tend to make them more likely to use these drugs, \(^{38}\) the totally rational consumer will consider this fact prior to initiating any drug use. Recognizing the power of addictive goods, the theory of rational addiction holds that individuals will refrain from using them, or engage only in limited experimentation, in order to avoid the long-term costs of being addicted. This approach suggests that consumption of potentially addictive goods will be more sensitive to price in the long run than in the short run, an implication that has been verified in the case of cigarettes.\(^ {39}\) While many scholars with expertise in addiction have been highly critical of assuming rational behavior among drug users,\(^ {40}\) it is nevertheless the case that a reduced incidence of experimental drug use may be another potential benefit of higher prices caused by enforcement.

Research in what has been labeled “behavioral economics” questions the assumption of rational behavior that underlies all of traditional economics. “Rational behavior,” as used by mainstream economists, means that behavior is based on stable time and risk preferences, an assumption that is not supported in various experimental settings.\(^ {41}\) Contrary to the rational addiction hypothesis, these unstable preferences, together with limited knowledge and imperfect cognitive ability, suggest that some individuals are likely to use drugs and later regret this decision.\(^ {42}\) This offers an economic ra-

---


39. The short run price elasticity of demand for cigarettes is typically about -.4 and the long run elasticity is about -.75. See Michael Grossman et al., *A Survey of Economic Models of Addictive Behavior*, 28 J. Drug Issues 631, 636 (1998). It is interesting to note that cigarette demand responds much less to price than heroin and cocaine. Studies also report that youth respond more to cigarette prices than adults, no doubt because cigarette consumption accounts for a larger portion of their income than it does among adults.

40. See MacCoun & Reuter, *supra* note 14, at 64 n.5 (stating that “Becker’s model [of rational addiction] is an intellectual tour de force of unknown relevance to the phenomenon of real-world addiction”).


Rationalale for government policy that increases the costs of drug use to offset such “irrational” behavior. Providing a cogent case against the strict rationality assumption of traditional economic analysis, the behavioral perspective provides a potential reason to reject the standard economic argument for legalization even if there are no externalities associated with drug use, but provides little additional insight into how drug policy should be implemented. After all, implementing a drug policy requires the use of scarce resources, so even if criminalization does save some people from their own irrationality, it imposes costs on other people. Indeed, the trade-offs even arise within the drug-using population, where the appropriate weights must be determined for the alleged benefits and costs that drug enforcement generates for both potential users who are “saved” and actual users who may be destroyed by this policy.

Enforcement efforts that lower demand may reduce some harms of drug use, but these policies also can have the effect of harming users and others. Some of these consequences, such as punishments, are obviously intended while others, such as increasing toxicity of drugs and violence, are probably unintended, although some drug-war zealots would no doubt consider them just deserts. Diligent enforcement of illicit drug laws creates other social harms because it is a blunt instrument that cannot distinguish between what might be called normal experimentation by youth and problem drug use. To the extent that casual use of mind-altering substances is a normal part of the adolescent experience, higher levels of enforcement may generate serious harms by increasing the probability of criminal sanctions for relatively benign behavior.

people undergo the experiences of life. See Karen I. Vaughn, Austrian Economics in America: The Migration of a Tradition 80-81 (1994). Thus, a decision or action that may be rational at the time it is made, given the decision-maker’s limited knowledge, can lead to regret as the individual accumulates additional experience and knowledge.

43. This would seem particularly pertinent for youth, who are likely to be more prone to immediate gratification and to take risks. See Ted O’Donoghue & Matthew Rabin, Risky Behavior Among Youths: Some Issues from Behavioral Economics, in Risky Behavior Among Youths: An Economic Analysis 29-68 (Jonathan Gruber ed., 2001).

44. The relationship between psychological characteristics and drug use suggests that adolescents who engage in some drug experimentation, primarily with marijuana, are better adjusted than individuals who entirely abstain from use. Jonathan Shedler & Jack Block, Adolescent Drug Use and Psychological Health: A Longitudinal Inquiry, 45 AM. PSYCHOLOGIST 612 (1990), followed their subjects from preschool through age eighteen and conclude that heavy users of drugs are maladjusted, with poor impulse control. Youth who never experiment with any drug are less well-adjusted than those who do, and are described as anxious, emotionally constricted, and having impaired social skills. A review of the medical effects of extended daily use of cannabis reveals uncertain but potentially serious consequences, with the implication that occasional use does not constitute a health hazard. See Wayne Hall & Nadia Solowij, Adverse Effects of Cannabis, 352 LANCET 1611 (1998).
III. ENFORCEMENT CAN CHANGE RELATIVE PRICES AND CONSUMPTION PATTERNS

In the previous Section we saw that enforcement might reduce the quantity of illicit drugs demanded, but this potential benefit is much more elusive than our analysis of the direct impact of raising price (and even full price) suggests. Economic theory is based on the proposition that individuals respond to incentives and that, on the margin, prices play an important role in guiding individual choice. Advocates of the drug war often seem to believe that drug users will only respond to the rising price by reducing or stopping use, but in this Section we show that there is no reliable evidence that this earnest hope is warranted.

A. Persistence of Demand

The demand for mind-altering substances seems to be persistent; in fact, it is alleged to be common among many species. If individuals consume drugs to achieve an altered mental state then anything that raises the full price of the drug of choice will give consumers an incentive to seek alternative, relatively low-priced drugs that provide a similar effect. This tendency can also explain the life cycle of individual drugs. New drugs are often attractive to users because their intoxication effects are immediately apparent while the consequences of using them are not. As experience of adverse consequences accumulates, the full price of use becomes more apparent and the drug loses popularity. Similarly, if enforcement increases the price of an

45. See RONALD K. SIEGEL, INTOXICATION: LIFE IN PURSUIT OF ARTIFICIAL PARADISE 207-27 (1989) (asserting that there is a powerful natural force, which he calls the “fourth drive,” that motivates the pursuit of intoxication). For a discussion of the idea that the “fourth drive” is common to many species, see id. at 10.

46. Microeconomic theory suggests that people purchase items with particular characteristics rather than a specific product, and therefore goods with similar characteristics are substitutes for one another. See Kelvin J. Lancaster, A New Approach to Consumer Theory, 74 J. POL. ECON. 132 (1966). This suggests that drug users may not demand a particular drug, say cocaine; instead, they can have a generic demand for an altered mental state that can be satisfied to varying degrees by alternative substances.

47. A closely related issue deals with attempts to reduce beer consumption among youth. There has been a lively debate in the economics literature on whether raising the excise tax on beer is an effective policy to reduce alcohol-related traffic fatalities. Until recently, the evidence suggested that taxes were the most effective policy to reduce these deaths, but recent evidence has challenged the prevailing view. One reason why taxes are not expected to influence beer consumption that leads to traffic fatalities is that beer drinkers can buy a very close substitute without increasing their expenditures, to wit, they can buy a cheaper brand of beer and thereby offset the impact of the tax by this product substitution. For a discussion of this literature, see Brent D. Mast, Bruce L. Benson & David W. Rasmussen, Beer Taxation and Alcohol-Related Traffic Fatalities, 66 S. ECON. J. 214 (1999).

illicit drug, consumers often can shift to alternative illegal substances or to new products that have not yet been declared illegal.

Persistence of demand for mind-altering substances can be responsible for substantial unintended consequences of enforcement that undermine efforts to reduce the harms of drug use. When a specific drug is viewed as a particular problem, a policy of increased enforcement to combat its use may appear attractive because the resulting higher price will curtail its use. However, users of this drug are likely to adjust their consumption patterns by looking for alternative psychoactive substances, and there can be no presumption that the alternative is less harmful than the substance being targeted for increased enforcement.

B. Persistence Leads to Substitution

Critical to understanding the impact of enforcement on drug use is determining the extent to which various drugs are substitutes for one another. Unfortunately, the literature on this point is not definitive due to data limitations and the fact that these relationships almost surely vary by type of drug and characteristics of users.

49. The appeal of this policy is rooted in the mistaken assumption that users will not change their behavior and that they are limited in their response to the first order effect, i.e., lower consumption of the drug targeted by the policy. Thus, an unsophisticated approach to drug enforcement suggests that when users of marijuana face a higher full price they will lower consumption and use the savings to buy legal products, such as caffeinated soft drinks. Those committed to a war on drugs all too often believe the only conceivable response by drug consumers is the one desired by the policy-maker. In fact, consumers have a myriad of options, including not responding to the policy, reducing the frequency of consumption, switching to another illegal substance that now has a relatively lower full price, as well as the desired response of complete abstinence from the targeted drug and all its close substitutes.

50. Such a case is suggested by one study where it is reported that lower marijuana prices, which reduces consumption of a substitute, alcohol, leads to a significant drop in the probability of a non-fatal automobile accident. See Frank J. Chaloupka & Adit Laixuthai, Do Youths Substitute Alcohol and Marijuana? Some Econometric Evidence, 23 E. ECON. J. 253, 265-66 (1997). Assuming that the consequences of marijuana use are not as severe as the expected losses associated with an automobile accident, it appears that increasing enforcement against marijuana is harm enhancing. Additional discussion of the alcohol-marijuana relationship is provided below.

51. If two drugs are substitutes, increasing the price of one will cause an increase in the demand for the other. Some drugs could also be complements, meaning that a reduced price and rising consumption of one drug will be accompanied by more consumption of the other.

52. Most studies exploring the relationship between marijuana and other drugs are compromised because they do not have reliable measures of the money or full prices of any drugs. For a brief review of the literature on the demand for marijuana among youth, see Rosalie Liccardo Pacula et al., Marijuana and Youth, in RISKY BEHAVIOR AMONG YOUTHS, supra note 43, at 283-88.

53. There is some evidence that the demand for drugs among persons under 21 is different from demand among young adults aged 21-30, in that the latter are more responsive to punishments for marijuana possession and less responsive to beer taxes. See Matthew C. Farrelly et al., THE EFFECTS OF PRICES AND POLICIES ON THE DEMAND FOR
Given that the absolute value of the price elasticity of demand is inversely related to the percent of income spent on a good, we would expect heavy drug users who spend most of their income on drugs to be very sensitive to changes in relative drug prices while young, infrequent experimenters’ drug of choice may not be very sensitive to such changes. For most youth, the relevant choice among drugs is between alcohol, marijuana, and hashish. Recognizing that it is not realistic to expect social policy to ever get youth to “just say no” to all mind-altering substances, it follows that a crucial dimension of this policy debate rests in how we assess the relative costs of using and abusing alcohol and cannabis. Current policy obviously favors the former over the latter, but it is not clear that an objective evaluation would support policies that increase the full price of cannabis relative to alcohol.

C. Alcohol and Cannabis

According to a leading think-tank on substance abuse, “research has not established a direct causal relationship between substance abuse and . . . social problems” related to criminal justice, social service expenditures, and business. An estimate of the economic costs of alcohol in terms of direct healthcare costs and the indirect burden of productivity losses shows that in Canada they are about six times...

---


55. Over 23% of high school seniors in 1999 reported using marijuana in the last 30 days, compared to 51% reporting use of alcohol and only 2.6% using cocaine. Bureau of Justice Statistics, U.S. Dept. of Justice, SOURCEBOOK OF CRIMINAL JUSTICE STATISTICS–2000 245-46 tbls. 3.71-3.72 (Kathleen Maguire & Ann L. Pastore eds., 2001) [hereinafter CJ Statistics 2000].

56. SIEGEL, supra note 45, at 207-27.

57. CTR. ON ADDICTION & SUBSTANCE ABUSE (CASA), SUBSTANCE ABUSE AND URBAN AMERICA: ITS IMPACT ON AN AMERICAN CITY, NEW YORK 84 (1996). Reflecting a frequent practice, the lack of evidence did not deter the Center on Addiction and Substance Abuse, CASA, from estimating substance abuse related costs by analyzing budgets and expenditures related to substance abuse. Thus, in principle, any anti-drug program that is ineffective and wasteful is counted as a “cost” of substance abuse.
greater than the corresponding costs of illicit drugs. The same study estimates that almost 98% of direct costs due to traffic accidents arising from substance abuse are associated with alcohol with the remainder attributed to illicit drugs. Since alcohol tends to impair drivers more that marijuana, the most frequently used illicit drug, this result is not surprising. Adding to this evidence is the fact that there are no recorded overdose deaths associated with marijuana, while acute alcohol poisoning is a relatively common occurrence. Thus substantial evidence suggests that using enforcement to raise the full price of marijuana, relative to alcohol, may enhance the harms of substance abuse.

Advocates of marijuana prohibition reject such arguments, primarily by evoking the gateway hypothesis: the proposition that marijuana use precedes and leads to hard drug use. Scientific evidence of such a causal relationship is sparse; this is not surprising since the gateway hypothesis is based on a fundamental error of logic. Early delinquent behavior, including use of tobacco, alcohol and marijuana, does appear to be correlated with subsequent use of hard drugs, but not by way of causation. Frequent drug users have a multitude of maladies, reflecting a broad array of problems that are not related to drug use per se.

58. See Eric Single et al., The Economic Costs of Alcohol, Tobacco and Illicit Drugs in Canada, 1992, 93 ADDICTION 991, 1000 (1998). Use rates of alcohol and illicit drugs in the U.S. in 1992 are consistent with these data. The National Household Survey on Drug Abuse reported that, in 1992, 34.2% of young adults had engaged in binge drinking in the last two weeks. By comparison, only 1.8% of this population reported having used cocaine in the last thirty days. By any calculus, the frequency of binge drinking is many times the rate of non-cannabis drug use. CJ STATISTICS 2000, supra note 55, at 254-55 tbls. 3.80-3.81.

59. Drivers under the influence of marijuana make mistakes, but are more cautious than sober drivers in that they keep a greater distance from the car in front of them and drive more slowly. Hendrik W.J. Robbe & James F. O’Hanlon, Nat’l Highway Traffic Safety Admin., Marijuana, Alcohol and Actual Driving Performance (1999); Alison Smiley, Marijuana: On-Road and Driving Simulator Studies, 2 ALCOHOL, DRUGS, AND DRIVING: ABSTRACTS AND REVIEWS 121 (1986). In 1999, alcohol was implicated in 38% of traffic fatalities. CJ STATISTICS 2000, supra note 55, at 275 tbl. 3.117.

60. See Michalis P. Charalambous, Alcohol and the Accident and Emergency Department: A Current Review, 37 ALCOHOL & ALCOHOLISM 307 (2002); Hall & Solowij, supra note 44, at 1612.

61. Most users of hard drugs have prior experience with tobacco, alcohol, and marijuana, a relationship that appears to support the gateway hypothesis. See Robert J. Kane & George S. Yacoubian, Jr., Patterns of Drug Escalation Among Philadelphia Arrestees: An Assessment of the Gateway Theory, 29 J. DRUG ISSUES 107 (1999).

62. See MacCoun & Reuter, supra note 14, at 351, for the most sympathetic yet rigorous interpretation of the gateway hypothesis, claiming that while one can present a coherent argument for the gateway hypothesis (but not one that necessarily implies the criminalization of marijuana), “[i]n the absence of better causal evidence, a strong allegiance to any particular gateway theory would seem to reflect ideology or politics rather than science.”
Relative to experimenters, frequent users are described as not dependable or responsible, not productive or able to get things done, guileful and deceitful, opportunistic, unpredictable and changeable in attitudes and behavior, unable to delay gratification, rebellious and nonconforming, prone to push and stretch limits, self-indulgent, not ethically consistent, not having high aspirations, and prone to express hostile feeling directly.63

These characteristics, like those of adolescents who abstain or experiment with drugs, are traced to the early years of childhood. Thus, the timing and extent of drug use are a symptom of these underlying conditions and probably not a consequence of earlier marijuana use.64 Indeed, many more young people experiment with alcohol and marijuana but not with hard drugs, than those who experiment with marijuana and then become regular users of hard drugs.65 The bankruptcy of the gateway hypothesis is also reflected in the evidence that youth who just experiment with drugs are better adjusted than both abstainers and frequent users.66

IV. IMPACT OF ENFORCEMENT ON SUPPLY

That drug enforcement is effective in lowering the harms of drug use can be characterized as a “faith-based” proposition, i.e., in the absence of substantive evidence it is believed that the substance abuse “problem” will decline when a specific drug price rises due to enforcement. The previous Section suggests there is no evidence to support this belief, and the present Section shows that enforcement efforts can also lead to unintended adverse consequences if dealers and suppliers respond to relative prices as economic theory suggests.

A. Enforcement Can Reduce the Supply of a Drug

Effective enforcement that raises the probability of arrest and the severity of expected punishment of drug suppliers increases the cost

63. Shedler & Block, supra note 44, at 617.
64. Id. at 626.
65. Despite the relatively high rate of marijuana use reported among high school seniors, supra note 55 and accompanying text, only 1.7% of young adults in 2000 reported cocaine use in the last thirty days. CJ STATISTICS 2000, supra note 55, at 254 tbl. 3.80.
66. Studies by economists implicitly support this finding in that there seems to be a mild positive effect of marijuana use on labor market outcomes that erodes with age. Hard drug use is associated with higher unemployment. These studies, however, inevitably suffer from measurement error since the indicator of most intense use is only use in the past month. In any event, existing evidence suggests that casual marijuana use is not associated with diminished labor market success. See Robert Kaestner, New Estimates of the Effect of Marijuana and Cocaine Use on Wages, 47 INDUS. & LAB. REL. REV. 454, 454-55 (1994); Ziggy MacDonald & Stephen Pudney, Illicit Drug Use and Labour Market Achievement: Evidence from the UK, 33 APPLIED ECON. 1655 (2001).
of doing business, and therefore reduces supply and increases price.\textsuperscript{67} Current enforcement directed against suppliers is obviously effective, since the street prices of these agricultural products are very high relative to what they would be if they were legal.\textsuperscript{68}

Enforcement increases the cost of supplying drugs primarily because workers engaged in the production and distribution of drugs demand more pay to offset the greater risks of arrest and punishment.\textsuperscript{69} Many advocates of enforcement earnestly hope that suppliers will reduce or cease activity in an illicit drug market when faced with rising costs due to enforcement. But for many relatively unskilled persons supplying drugs, the opportunity cost of their time is probably minimum-wage employment.\textsuperscript{70} Even if some drop out of the business due to rising enforcement, there is no scarcity of people prepared to enter the drug business to replenish the personnel needs of suppliers.\textsuperscript{71} But the impact of rising enforcement will unquestionably increase production costs and raise the street price of the drugs targeted by police agencies, all other things constant, thereby reducing the quantity demanded of these substances.\textsuperscript{72} The net impact of rising enforcement on drug consumption, however, will be determined by the degree to which drug suppliers can successfully counteract enforcement efforts by changing the methods of operation and by altering the composition of drugs supplied.

\textsuperscript{67} Economic analysis is designed to discern how economic agents respond to marginal changes and our discussion here is in this tradition. The impacts of draconian changes in enforcement, such as imposing the death penalty for relatively minor drug offenses, cannot be considered in the framework employed here. Since such a policy clearly violates the tenet that penalties be proportional to harms, this limitation of our analysis seems to be appropriate.

\textsuperscript{68} See Caulkins & Reuter, supra note 18, at 593-95.

\textsuperscript{69} See RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 78; REUTER ET AL., MONEY FROM CRIME, supra note 12, at 104-05 (reporting significant annual risks of death, injury and incarceration).

\textsuperscript{70} REUTER ET AL., MONEY FROM CRIME, supra note 12, at viii (noting that the gross return per hour for drug dealing in the Washington D.C. area is about $30, slightly over four times the average legal hourly wage among this population). Drug dealing is complementary to legal employment because most trading is conducted in the evening and on weekends. As Reuter et al. also note, dealers with legal jobs report making the most money from dealing. Id. at 67. Their greater earnings could be the result of having better connections because their co-workers are potential customers with a steady flow of income. A not mutually-exclusive alternative explanation for their higher earnings is that legally employed dealers are simply more ambitious and/or skilled in the pursuit of both legal and illegal earnings.

\textsuperscript{71} See Moore, Supply Reduction, supra note 12, at 137-38.

\textsuperscript{72} This statement assumes that the drug suppliers are operating efficiently prior to the increase in enforcement in that they maximize profits by using the most productive set of inputs. If this assumption is rejected, a rise in enforcement that lowered profits could stimulate drug suppliers to choose a more optimal set of inputs. Under this scenario, a rise in enforcement does not necessarily result in a reduction in the amount of drugs supplied. Similarly, as noted in the following Section, increased enforcement can lead suppliers to innovate in ways that, in theory, might actually lower costs and prices.
B. Suppliers Act to Offset the Effects of Enforcement

A fundamental premise in the economic theory of the firm is that employers will change the combination of inputs when their relative prices change.73 For example, California's use of aerial surveillance to thwart marijuana growers stimulated indoor production that was highly capital intensive, thus substituting capital for land, but was nonetheless profitable because the technology could annually produce four crops of a more potent strain.74 The mix of labor resources used in drug selling is also responsive to enforcement policy. Suppose a change in policy increases the expected punishment for all drug dealers but that the expected impact on adults is more severe than for juveniles, because punishment of juveniles is less severe than punishment of adults.75 To reduce their own risk of arrest in the face of increased threats from enforcement, drug entrepreneurs have an incentive to lengthen the distribution chain, thereby personally dealing directly with a smaller number of individuals. Furthermore, since the change in enforcement raises the price of adult workers relative to juveniles, suppliers have an incentive to substitute youth for adults in the distribution chain. Rising enforcement in this instance is likely to have the intended effect of raising the cost of production, and therefore discouraging use since the street price will be higher. However, the unintended consequences of the rising enforcement are that more people are engaged in supplying a smaller amount of drugs and more juveniles have been lured into the drug trade.76

Geographic substitution effects are also likely. As drug enforcement efforts become effective against producers in one geographic area, production will shift to other areas. In the international arena, successful control of Turkish heroin in 1973-1974, discussed in the famous "French Connection" case,

resulted in a significant reduction in the flow of heroin into the United States. Heroin prices rose sharply in this country, reducing use but also giving other suppliers an incentive to enter the mar-

---

73. See Rasmussen & Benson, Economic Anatomy, supra note 11, at 77-81 (discussing input substitution in this context).
75. Juveniles face less severe and shorter sentences for most crimes than adults do, but even if the sentences were identical in length, juveniles are likely to perceive the potential sentence as less severe. After all, a two-year prison term may appear very severe if an individual is sixty-five since it is a substantial part of his expected remaining life, while it is much less significant to someone who is fifteen. In addition, juveniles are likely to be more myopic (i.e., put less weight on future possible costs and benefits) than adults, and/or less risk averse (i.e., they put less weight on the potential negative consequences of decisions).
76. This suggests that increased enforcement can cause one measure of the drug problem, the number of people engaged in the drug trade, to rise while another measure, the quantity sold, falls.
ket. The [supply-reducing effects] of breaking the French Connection lasted for two to three years before there was an expansion in heroin supply from Mexico and Southeast Asia.\textsuperscript{77}

This is an instructive example because it demonstrates that the effect of law enforcement focused in one direction can be completely mitigated by drug market entrepreneurs within a relatively short period of time and, of course, even the short-term benefits of the lower supply of heroin can be at least partially offset by consumers shifting to other psychoactive substances. Further, the very success of the attack on Turkish heroin resulted in a more diversified supply system that made future control of this drug even more difficult.\textsuperscript{78}

Efforts to stop the importation of cocaine into the U.S. that were directed against the Medellin cartel in Colombia reveal impacts that are similar to the French Connection effort. In damaging the cartel, the enforcement efforts led to a much more dispersed processing and shipping network for cocaine. Suppliers created labs processing cocaine in many Latin American countries, and greatly increased the number of transshipment points, effectively emasculating enforcement by making subsequent interdiction efforts much more costly and ineffective.\textsuperscript{79} A long history of drug enforcement efforts suggests that elimination of supplies coming from one area will soon lead to increased cultivation elsewhere.\textsuperscript{80}

Drug suppliers can also shift to the production and distribution of other drugs when faced with effective enforcement, a response called “output substitution.”\textsuperscript{81} Such changes in output can increase the harms associated with drug use rather than reduce them. Efforts to intercept drugs in the Miami area in 1984 were highly successful against the importation of marijuana, no doubt because this product is bulky and relatively difficult to conceal. Smugglers did not change their occupation. Instead they simply changed the product being smuggled, shifting to a lower risk commodity, cocaine.\textsuperscript{82} Successful

\textsuperscript{77} RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 80. In Moore, \textit{Supply Reduction}, supra note 12, at 136, it is argued that such a temporary increase in price significantly lowers the harms of drug use by reducing the prevalence of use and intensity of use among a number of age cohorts. This is plausible, but the net harm reduction is likely to be less than the decline in heroin use if the discouraged users shift to other mind altering substances, as discussed in Section III.

\textsuperscript{78} RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 80.

\textsuperscript{79} Id. at 79-81.

\textsuperscript{80} See Reuter, \textit{Eternal Hope}, supra note 12, at 87; see also KLEIMAN, supra note 28, at 284 (providing an interesting domestic example of a shift in marijuana production from Northern California to the mountains of Kentucky as a result of increasing enforcement). An unintended consequence of this shift, aside from circumventing enforcement, was to involve people whose cultural roots include moonshining and a history of violence, making the trade rougher than it had been before.

\textsuperscript{81} RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 81-82.

\textsuperscript{82} Id.; see THORNTON, supra note 11, at 109.
interdiction of marijuana thus increased the supply of cocaine, and youth correspondingly reported that this drug was more readily available. The higher price of marijuana relative to cocaine probably increased the use of cocaine, and this effect was reinforced as the increased availability of cocaine lowered its price and increased the quantity demanded. Higher demand for cocaine will tend to push its price up, resulting in higher profits in the short run that further encourage suppliers to increase the amount of cocaine brought to market. The U.S. drug war of 1984-1989 was more successful against marijuana than cocaine, with the expected result of increased supplies of cocaine.

Entrepreneurs in all industries face strong incentives to find ways to produce or distribute existing products at lower costs, and to offer new products that will attract consumer demand. Broadly described as technological change and product development, there is no reason to believe that entrepreneurs in illicit drug markets are any less likely than those in legal markets to engage in these efforts. In fact, drug entrepreneurs may have added incentives to increase revenues since they must offset the higher costs associated with the risk of arrest and punishment. If a drug entrepreneur can find a way to either lower production or distribution costs or to lower the probability of arrest, the business will be more profitable. Synthetic drugs (such as LSD and Ecstasy), product “improvements” such as crack cocaine, and introduction on the street of long known drugs such as MDMA, are representative of a long history of entrepreneurship in the illicit drug industry.

C. Enforcement Can Increase Potency

Increasing enforcement against the illicit drug industry also has a tendency to increase the potency of mind-altering substances. This first became apparent during America’s alcohol prohibition experiment. During Prohibition, consumption of high-alcohol content spirits rose sharply relative to beer because spirits were relatively easy to conceal and transport, thereby making them more attractive to

---

83. Survey data indicate that the percentage of high school seniors reporting that it was “fairly easy” or “very easy” to get cocaine rose continuously from 1983 to 1989, rising from 43.1% to 58.7%. BUREAU OF JUSTICE STATISTICS, U.S. DEPT OF JUSTICE, SOURCEBOOK OF CRIMINAL JUSTICE STATISTICS–1991, at 230 tbl. 3.65 (Timothy J. Flanagan & Kathleen Maguire eds., 1992) [hereinafter CJ STATISTICS 1991].
84. Since marijuana can be grown throughout the U.S., successful interdiction of marijuana can be offset within a relatively short time by increasing domestic production. See RALPH A. WEISHEIT, DOMESTIC MARIJUANA: A NEGLECTED INDUSTRY 36 (1992).
86. CJ STATISTICS 1991, supra note 83, at 230 tbl. 3.65.
consumers and producers alike.\textsuperscript{87} Prior to Prohibition (from 1911 to 1916), the ratio of expenditures on spirits to expenditures on beer was fairly stable, ranging from 0.70 to 0.78,\textsuperscript{88} but there was a dramatic shift in these spending patterns with Prohibition. It is estimated that by 1925 consumers spent about seven times as much on spirits as they did on beer.\textsuperscript{89} Furthermore, some of the bootleg whiskey contained as much as twice the alcohol found in commercial brands.\textsuperscript{90}

The incentives to produce and consume more powerful alcohol during Prohibition are also inherent in our current drug laws. Low dosage drugs such as marijuana are bulkier and more easily detected than harder drugs, providing the incentives already noted to increase the supply of more potent drugs. Furthermore, because the penalties for possessing and selling drugs are related to weight rather than strength, the legal system provides a strong incentive to avoid handling low-dose products that have been cut.\textsuperscript{91} But beyond this, drug entrepreneurs can produce stronger drugs just as more potent alcohol emerged during Prohibition. This possibility has long been recognized, as reflected in a 1967 Report to the President’s Commission on Law Enforcement and Administration of Justice:

> If United States law-enforcement policies become so efficient as to prevent altogether the smuggling of heroin, the black market can readily convert to narcotic concentrates that are a thousand or even ten thousand times more potent, milligram for milligram. . . . A few pounds of these concentrates might supply the entire United States addict market for a year. . . . The skills required are not beyond those possessed by the clandestine chemists who now extract morphine from opium and convert the morphine to heroin, or of better chemists who might be recruited.\textsuperscript{92}

\textsuperscript{87} As a consequence of the increased supply of spirits relative to beer, beer prices rose sharply relative to the price of spirits between 1916 and 1928: 700\% compared to 310\% for rye whiskey. IRVING FISHER, PROHIBITION STILL AT ITS WORST 91 (1928).


\textsuperscript{89} By 1930 beer consumption was rising and Warburton estimates that expenditure on spirits had fallen to about three times the amount spent on beer. Id. at 170 tbl. 83. After Prohibition, expenditure on spirits declined, accounting for about half of alcohol expenditures between 1939-1960. See THORNTON, supra note 11, at 103.

\textsuperscript{90} See THORNTON, supra note 11, at 103.

\textsuperscript{91} As in the case of heroin, a product containing any detectable amount is illegal, but the weight of the entire product is used to determine the penalty. Thus, at all levels of production and even among consumers, there is an incentive to use a purer product. Smaller bundles are also easier to conceal, giving smugglers and sellers a similar incentive to sell uncut substances.

\textsuperscript{92} EDWARD M. BRECHER, Licit and Illicit Drugs 96 (1972) (quoting ARTHUR D. LITTLE, INC., DRUG ABUSE AND LAW ENFORCEMENT, A REPORT TO THE PRESIDENT’S COMMISSION ON LAW ENFORCEMENT AND ADMINISTRATION OF JUSTICE (Jan. 18, 1967)).
Trends in the potency of marijuana are consistent with the proposition that greater enforcement efforts lead to increased potency. Thornton reports that the average potency of marijuana increased by a factor of eight between 1974 and 1984, and he provides evidence suggesting that marijuana potency is positively correlated with law enforcement expenditures.\(^9\) The U.S. Drug Enforcement Administration (DEA) also provides evidence that increasing potency can be an unintended consequence of rising enforcement. For instance, a program designed to eradicate marijuana production was targeted at indoor and outdoor cultivation, but since it is easier to detect large outdoor growers, relatively more marijuana reaching the market came from indoor growers.\(^9\) The DEA speculated that since indoor production is more capital intensive and allows for a more controlled growing environment, more potent marijuana was the consequence of this attempt to control supply.\(^9\)

Clearly, unintended negative consequences of drug enforcement are common as both drug users and suppliers respond to relative prices in hopes of limiting its impact on their activities. Furthermore, the magnitude of the alleged harms of some drug use is probably substantially less than is popularly perceived, while the hidden costs of criminal justice drug control efforts are substantially greater than a simple summation of the budgetary outlays for drug enforcement suggests.\(^9\) We now turn to a discussion of the impact of drug enforcement on public safety.

\(^9\) Thornton, supra note 11, at 105-08, regressed marijuana potency data on a measure of enforcement and estimated that a $1 million (1972 dollars) increase in federal drug law enforcement expenditures results in a 0.01\% increase in potency. This simple regression does not provide conclusive evidence, but this time-series analysis is consistent with the expectation that drug suppliers will tend to increase potency when they face rising enforcement efforts.

\(^9\) The Domestic Cannabis Eradication and Suppression Program began in 1979, but it was not until 1985 that all states received funding for the program. See Pacula et al., supra note 52, at 294.

\(^9\) See The Pothouse Effect, supra note 74.

\(^9\) The actual monetary expenditures on drug control are also largely hidden since they are dispersed across so many federal, state, and local agencies that generally have many other functions as well. It might be possible to determine the budget allocated to the vice squad of a local police department, for instance, but even the vice squad deals with more than just drugs, and furthermore, many other units in a police department are also involved in drug control efforts (e.g., the uniformed officers on patrol, those involved in administration, dispatch, and so on). It may be possible to estimate the cost that a state corrections department spends on controlling and supervising individuals convicted for drug crimes (although published estimates of average per inmate costs are often questionable and vary considerably across prisons and probation systems), but there are also a wide range of criminal justice system “diversion programs” supervised by other agencies (e.g., drug courts, local “boot camps”) that would have to be counted. And of course, agencies such as the Coast Guard, State Police, the United States Army and Air Force (with planes and other equipment along with advisers, pilots, and other personnel in places like Columbia and Peru), the FBI, the Border Patrol, and numerous other federal, state, and local agencies allocate substantial resources to drug enforcement. Accounting of these
V. DRUG ENFORCEMENT AND PUBLIC SAFETY

A. Do Drugs Cause Crime?

Drug enforcement is often defended as a crime-fighting weapon because drug users allegedly commit most of the property crimes in order to support their habits. Crimes such as selling drugs and prostitution are commonly committed by drug users, but these are offenses against social norms and morals; thus, they are not reported in usual crime rate statistics. Hence, these crimes are more directly related to issues of personal freedom rather than social order, while drug enforcement is sometimes justified by its capacity to reduce crimes against persons and property.

The fact that many criminals convicted for property and violent offenses are also drug users is well documented, and this fact has budgetary costs is likely to be inaccurate due to the myriad of departments involved in drug control. Estimates of federal expenditures are $18 billion in 2000, *CJ Statistics 2000, supra* note 55, at 15 tbl. 1.12, a figure that is no doubt a fraction of the total given that most drug offenders are arrested and prosecuted in state and local jurisdictions.

97. Blaming drug users for crimes against property occasionally takes on bizarre proportions, as in the case where users were blamed for more crime in New York City than was being reported. Peter Reuter, *The (Continued) Vitality of Mythical Numbers*, Pub. Int., Spring 1984, at 135-36.

98. Bruce D. Johnson et al., *Careers in Crack, Drug Use, Drug Distribution, and Nondrug Criminality*, 41 CRIME & DELINQ. 275, 281 (1995) (reporting that the crack epidemic in New York City did not substantially increase non-drug criminality with the exception of prostitution). *Reuter et al., Money from Crime, supra* note 12, at 65, indicate that drug selling dominates non-drug crime as a source of income. Studies of the heroin market also indicate that users are likely to turn to dealing for income. See *Leroy C. Gould et al., Connections: Notes from the Heroin World* 49 (1974); *Moore, Buy and Bust, supra* note 12, at 52.

99. Crime rates are calculated for Index I crimes, the crimes against persons and property that are routinely reported to the police. Drug transactions and prostitution are not usually reported to the police and are therefore often called “victimless” due to their consensual nature.

100. David F. Musto, *The American Disease: Origins of Narcotic Control* 273 (3d ed. 1999), claims that there was a parents’ movement that clamored for rising enforcement because such a policy would assist parents in discouraging middle and upper class children from using drugs. While such a movement could rationally support continued illegality of drugs, advocating zero tolerance puts the most well balanced of youth at risk of punishment for an activity that is described as normal youthful experimentation without long-term consequences. Shedler & Block, *supra* note 44, at 625. We will examine the impact of drug enforcement on public safety and not address the social norms and moral questions here.

101. During 1999, for instance, between 49.5 and 76.7% of male arrestees in thirty-four U.S. cities tested positive for the use of an illicit drug. *CJ Statistics 2000, supra* note 55, at 394 tbl. 4.30. A Bureau of Justice survey of 12,000 inmates indicated that over 75% had used drugs, 56% had used drugs in the month prior to their incarceration, and one-third admitted to being under the influence of drugs at the time of their offenses. Harry K. Wexler et al., *Outcome Evaluation of a Prison Therapeutic Community for Substance Abuse Treatment*, 17 CRIM. JUST. & BEHAV. 71-72 (1990). Similarly, a survey of jail inmates found that 77.7% of the inmates admitted using some illicit drug and that 55.4% had used a major drug. *Harel, supra* note 35, at 4 tbl. 6. Furthermore, 43.5% had used some drug in the month prior to the offense for which they were admitted, and 27.7% had used a major drug.
contributed to the claim that drug use is a primary cause of crime, which, in turn, has led to increasing emphasis on the control of illicit drugs as a means of general crime prevention. Despite the high rate of drug use among persons arrested for other criminal activity, however, most research suggests only a loose connection between drug use and criminal activity. Chaiken and Chaiken summarize such research and conclude that “[t]here appears to be no simple general relation between high rates of drug use and high rates of crime.” 102 A detailed study of the arrest history of persons having at least one misdemeanor or felony drug arrest in Florida indicates that there is only a modest link between drugs and other crime, suggesting that most drug offenders have no violent criminal record and that many have few previous arrests for nonviolent crimes. 103 The vast majority of persons arrested for sale or possession of drugs also had no prior arrests for property crimes. Among persons arrested for the sale of drugs, a group more inclined to property crime than persons arrested for possession, 61.9% had no previous arrest for a property crime. Among the 45,906 persons arrested in 1987 for possession, over

---


103. See RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 61 tbl. 3.2. Consider the distribution of violent crime arrests among 1987 drug arrestees. The 45,906 persons who had been arrested for at least one drug possession offense in 1987 had a history of 19,436 violent crime arrests, an average of 0.42 violent crimes per arrestee, but 76% had no prior arrests for violent crimes. Among those arrested for drug possession, 2.3% of the arrested population accounted for 6,687 violent felony arrests—34.4% of the total. This group included 1,066 offenders who averaged 6.27 violent arrests in their past. The raw data is found in FLA. DEPT LAW ENFORCEMENT, FLORIDA DRUG OFFENDER PROFILE: ANALYSIS OF ALL OFFENDERS HAVING AT LEAST ONE MISDEMEANOR AND/OR FELONY DRUG ARREST IN CALENDAR YEAR 1987, at 2-5, 16-22 (1989).
80% had never been arrested for burglary, and almost 71% had never been arrested for any property offenses.\textsuperscript{104}

Thus, the evidence suggests that most drug offenders are not active participants in non-drug related crime. And surprisingly, given popular and political perceptions, drug consumers appear to be relatively less likely to be involved in property crime than drug sellers. These statistics and others that corroborate them,\textsuperscript{105} combined with survey data on drug use among persons arrested and convicted of Index I crimes, suggest that two distinct types of drug users exist. First, a substantial portion of drug offenders apparently do not commit property or violent crimes. Second, many offenders arrested for violent and property crimes also use drugs. The Florida data presented above suggests that there are relatively few habitual offenders who are heavily involved in both drugs and other crime, so the overall population of arrested and convicted drug offenders is probably not a population of hardened criminals, whose immersion in lives of crime has left them unresponsive to incentives. Thus, an expectation that a successful drug arrest and prosecution will simultaneously take a non-drug criminal out of circulation is likely to be disappointed. A war on drugs is not synonymous with a war on property or violent crimes.

There is some evidence that periods of heavy use of hard drugs is positively correlated with reported crime,\textsuperscript{106} a correlation that supports the idea that enforcement reduces the harms of drug use. Correlation does not necessarily imply causation, however. Studies of the temporal sequencing of drug abuse and non-drug crime (e.g., property crime) suggest that non-drug related criminal activities generally precede drug use. Indeed, the evidence seems consistent with the hypothesis that delinquency is not caused by drug abuse.\textsuperscript{107} If any-

\textsuperscript{104} Data summarized in \textit{Rasmussen \& Benson, Economic Anatomy}, supra note 11, at 62 tbl. 3.3.

\textsuperscript{105} A Bureau of Justice Statistics report on recidivism of felons on probation suggests that the characteristics of the Florida drug-using criminal population just described also apply to the nation as a whole. This report found that drug offenders are far more likely to recidivate for a drug offense than for a violent or property offense. Furthermore, violent offenders who are rearrested tend to recidivate most often for a new violent crime, and property offenders are most likely to recidivate for another property crime. See \textit{Patrick A. Langan \& Mark A. Cunniff, U.S. Dept of Justice, Bureau of Justice Statistics Special Report: Recidivism of Felons on Probation, 1986-89 (1992)}.

\textsuperscript{106} See John C. Ball et al., \textit{Lifetime Criminality of Heroin Addicts in the United States}, 3 J. Drug Issues 225 (1982); Silverman & Spruill, supra note 12, at 101.

\textsuperscript{107} \textit{Isidor Chein et al., The Road to H: Narcotics, Delinquency, and Social Policy} 64-65 (1964), argue that “the varieties of delinquency tend to change to those most functional for drug use; the total amount of delinquency is independent of the drug use.” Shedler & Block, supra note 44, strongly reaffirm the view that substance abuse is a symptom of more deeply rooted psychological problems. For a review of the temporal sequencing of drug use and crime, see \textit{Rasmussen \& Benson, Economic Anatomy}, supra note 11, at 57-58.
thing, it is more likely that crime leads to drug use, although both drug use and delinquency are likely to be caused by other factors.108 Once individuals turn to crime as a source of income, they may find that drugs are more easily obtained in the criminal subculture. Under this scenario, crime leads to drug use, but once an individual gets addicted preferences may change and the drugs-crime relationship may then become salient.109 But if this is the case, it appears to apply to only a small portion of the population engaged in drug market activity, as suggested above.

B. Drug Enforcement Can Cause Violence

Even if drug use is related to crime for only a portion of the drug consuming population, the suggestion that drug enforcement enhances public safety remains plausible. The evidence is not compelling, however, because a substantial part of the causation merely reflects the illegality of drugs, rather than their use per se. Violence in drug markets is often used as an argument for enforcement,110 but the fact that drug markets are illegal means that commercial disputes must be resolved outside the courts, with threatened or actual violence being the principal means for resolution.111 In fact, the use of violence to settle disputes in these markets appears to be an inevitable consequence of any policy except legalization.112 This effect of pro-

108. See RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 57. Also, CHRISTOPHER INNES, U.S. DEP’T OF JUSTICE, BUREAU OF JUSTICE STATISTICS SPECIAL REPORT: STATE PRISON INMATE SURVEY, 1986: DRUG USE AND CRIME 1-3 (1988), reports that about half of the prison inmates who had ever used a major drug, and roughly three-fifths of those who used a major drug regularly, did not do so until after their first arrest for some non-drug crime—that is “after their criminal careers had already started.”

109. Orphanides & Zervos, supra note 38, present a model in which addiction makes people more myopic, a change in preferences that would likely increase criminal involvement.


111. See Office of the Attorney Gen., U.S. Justice Dep’t, Drug Trafficking: A Report to the President of the United States 16 (1989) (noting that “the normal commercial concept of contracts, in which disputes are adjudicated by an impartial judiciary and restitution is almost always of a financial nature, is twisted, in the world of drug trafficking, into a system where the rule of law is replaced by the threat of violence.”).

112. Goldstein, supra note 29, at 24-36, argues that the illegal status of drugs generates three types of violence: (1) disputes among sellers, (2) robbery of drug market participants, and (3) disputes among users over drugs. Goldstein is skeptical of the notion that the pharmacological attributes of illegal drugs causes violence. This suspicion is confirmed by data presented by Paul J. Goldstein et al., Drug-Related Homicide in New York: 1984 and 1988, 38 CRIME & DELINQ. 459 (1992). Resignato also confirms Goldstein’s doubts after employing a data sample drawn from the National Institute of Justice’s “Drug Use Forecasting” cities to test Goldstein’s hypothesis. Resignato, supra note 11, at 685. The National Institute of Justice has been testing arrestees for drug use in twenty-four cities for several years, so a measure of drug use (at least among the criminal population) is avail-
Prohibition is exacerbated when enforcement is more effective against relatively benign and inexperienced drug dealers, leaving the trade with better organized and more violent organizations. Furthermore, the tools of competition that are available to firms in legal markets as they attempt to increase market share and profits (e.g., media advertising and investments in reputation by, for instance, building brand names protected by trademark laws), are not available in illicit drug markets. The use or threat of violence, however, is also a “competitive” tool that can be used to increase market share. Finally, drug users and dealers make attractive targets for robbery since they are generally carrying cash or drugs and are not likely to report their victimization. When such robberies result in physical harm, they come to the attention of police as “drug-related” assault or murder, which many observers inaccurately interpret as being caused by drug use per se.

Prohibition itself breeds violence, but there also is substantial evidence that the intensity of law enforcement influences the level of both violent and property crime. When a relative increase in enforcement disrupts local drug markets, dealers in the affected areas seek new market niches where they can ply their trade. When they

able for this sample. See id. These data allowed Resignato to test the relationships between drug use, the intensity of drug enforcement efforts, and violent crime. Id. at 685-86. He found that drug use itself is not strongly related with violence when other determinants, including the intensity of drug enforcement efforts, are controlled for, and concluded that neither psychopharmacological nor economic compulsive hypotheses were supported. Id. at 687-88. Instead, more intense drug-law enforcement efforts appear to cause violent crime, presumably by disrupting drug markets; leading dealers to relocate and engage in conflict over turf; perhaps generating contract disputes due to interrupted deliveries; and causing buyers to search for drugs in unfamiliar places, making them more vulnerable targets for robbery (participants in illegal drug markets are attractive targets for robbery because they generally carry cash or drugs and they are not as likely to report the crime as victims who are engaged in legal activities). See Goldstein, supra note 29, at 30-31. The U.S. experience with the prohibition of alcohol also confirms the notion that illegal markets breed violence. See THORNTON, supra note 11, at 120-26; Miron & Zwiebel, supra note 11, at 178-79. Also recall the black market in Canadian cigarettes that developed following the imposition of high taxes. See Benson & Rasmussen, Predatory Public Finance, supra note 11, at 197-98. Smuggling was accompanied by violent confrontations between police and smugglers and between rival smuggling organizations. Id. Thus, even legalization will not eliminate all violence if high taxes are imposed.

KLEIMAN, supra note 28, at 20.

See Goldstein, supra note 29, at 30.

115. It needs to be emphasized that there are other important consequences of rising enforcement that are not discussed here aside from those rights issues that we explicitly ignore. An obvious example is the constitutional issues relating to the relaxation of the Fourth Amendment standards for reasonable search and seizure. See, e.g., Paul Finkelman, The Second Casualty of War: Civil Liberties and the War on Drugs, 66 S. CAL. L. REV. 1389 (1993); Stephen A. Saltzburg, Another Victim of Illegal Narcotics: The Fourth Amendment (As Illustrated by the Open Fields Doctrine), 48 U. PITT. L. REV. 1 (1986); Silas J. Wasserstrom, The Incredible Shrinking Fourth Amendment, 21 AM. CRIM. L. REV. 257 (1984); Steven Wisotsky, Crackdown: The Emerging "Drug Exception" to the Bill of Rights, 38 HASTINGS L.J. 889 (1987).
enter markets that are already served by other dealers, the resulting competition will involve violence, as disputes over market share and service areas, or “turf,” arise. Buyers who must search for new suppliers in areas that they are not familiar with also may be more vulnerable to robbery.

C. Do Drugs or Drug Enforcement Cause Property Crime?

Rising property crime can also accompany increased drug enforcement. Because police resources are limited, increased drug enforcement implies reduced attention to other police responsibilities that represent what economists call the “opportunity cost” of drug enforcement. Scarce police resources must be diverted from the solution of violent crimes, combating property crimes, and/or the myriad of other activities that occupy law enforcement officers. The evidence from several studies in a variety of jurisdictions suggests that drug enforcement tends to draw police resources away from the solution of property crimes, thereby reducing deterrence and increasing property crime rates.

116. See Rasmussen et al., Spatial Competition, supra note 11, at 228-30 (providing statistical evidence of a direct law enforcement effect on such violence). As law enforcement efforts against drug markets increase in a Florida jurisdiction, the level of violent crime in neighboring jurisdictions rises. See also Resignato, supra note 11, at 683-88.

117. Only about 17% of all arrests by police are for crimes against persons or property. See CJ STATISTICS 2000, supra note 55, at 362 tbl. 4.7. For a catalog of the myriad of police activities that are not related to the solution of these crimes, see Benson et al., Estimating Deterrence Effects, supra note 10, at 164 n.6. Jonathan P. Caulkins et al., Price Raising Drug Enforcement and Property Crime: A Dynamic Model, 71 J. ECON. 593 (2000), point out that jurisdictions can avoid this trade-off by raising police budgets. This can only be accomplished by lowering expenditures on other government functions and/or raising taxes. The opportunity cost of increasing drug enforcement under this scenario is, for example, lower educational spending or less private consumption. RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 21, note that voters tend to resist tax increases and public bureaucracies resist budget cuts, so the opportunity cost of increasing drug arrests is likely, at least in part, to be in the form of a reduction of other police services.

118. Support for this trade-off hypothesis is found in several studies using data on Florida jurisdictions for various time periods. See, e.g., Benson et al., Deterrence and Public Policy, supra note 11; Bruce. L. Benson et al., Is Property Crime Caused by Drug Use or Drug Enforcement Policy? 24 APPLIED ECON. 679 (1992) [hereinafter Benson et al., Property Crime]; Bruce L. Benson, Ian S. Leburn, & David W. Rasmussen, The Impact of Drug Enforcement on Crime: An Investigation of the Opportunity Cost of Police Resources, 31 J. DRUG ISSUES 987 (2001); Hope Corman & H. Naci Mocan, A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City, 90 AM. ECON. REV., 584 (2000) (reporting corroborative evidence in a study using New York City data); Silvia M. Mendes, Property Crime and Drug Enforcement in Portugal, 11 CRIM. JUST. POLY REV. 195 (2000) (replicating the Sollars et al., Drug Enforcement and Deterrence, infra, results using data on jurisdictions in Portugal); David L. Sollars, Bruce L. Benson, & David W. Rasmussen, Drug Enforcement and the Deterrence of Property Crime Among Local Jurisdictions, 22 PUB. FIN. Q. 22 (1994) [hereinafter Sollars et al., Drug Enforcement and Deterrence]. Given scarcity, some trade-off is inevitable, although police decision-makers may choose the policy objective to sacrifice. Thus, for instance, BRUCE L. BENSON & DAVID W. RASMUSSEN, ILLINOIS’ WAR ON DRUGS: SOME UNINTENDED CONSEQUENCES 1, 12 (Heartland Policy Study No. 48, 1992), find that scarce police resources were diverted from traffic enforcement rather than
D. Corruption

Handicapped by its limited effectiveness and its tendency to reduce public safety by reallocating police resources from other endeavors, enforcement aimed against high-profit, cash-laden drug enterprises also leads to another crime: corruption of law enforcement officials. Government can be viewed as an entity that assigns and enforces property rights. In this light, one avenue for corruption is the illegal (or black market) “sale” of property rights. The illegality of drug markets gives police a valuable asset that can be sold—that is, agreeing to selective law enforcement where police clients are allowed to operate illegally, while the police harass other potential drug dealers, discouraging them from entering the market. Corruption is likely to be selective, serving to reinforce the most profitable drug suppliers who bribe police to focus their enforcement efforts on potential competitors entering the market. Under this scenario, it is expected that there will be a large number of drug arrests that have little impact on the market. As Moore and Kleiman note:

property crime enforcement, leading to a dramatic rise in traffic fatalities. Rising crime against persons and property could also accompany increasing drug arrests if it causes prison overcrowding and the subsequent release of non-drug felons. This was the case in Florida during the rapid escalation of drug arrests from 1984 to 1989. Since Florida was under a federal court order to reduce prison overcrowding, increasing drug arrests and convictions meant early release for prisoners. By the end of 1989, the average portion of sentence served among released Florida prisoners fell to 33%, down from the norm of about 50%. See RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 22-23. Both property and violent felons were released early. Perhaps the most famous case of early release was that of Charles H. Street, who had been sentenced to fifteen years for attempted murder. After serving half his sentence he was released in November 1988. Ten days after his release he murdered two Metro-Dade (Miami) police officers. See David Dahl, Why Was He Freed? Prisons Too Full to Hold Man Now Accused of Killing 2 Police, ST. PETERSBURG TIMES, Dec. 1, 1988, at 1A. Thus, prison overcrowding due to more drug arrests put criminals back on the street earlier, and since expected punishments fell, deterrence of crime fell to the extent that the severity of punishment discourages criminal behavior. See RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 22-23.

119. See Bruce L. Benson & John Baden, The Political Economy of Governmental Corruption: The Logic of Underground Government, 14 J. LEGAL STUD. 391, 392 (1985); Benson, Corruption in Law Enforcement, supra note 10. 120. Benson & Baden, supra note 119; Benson, Corruption in Law Enforcement, supra note 10. 121. In effect, the police can sell monopoly rights to a private sector underground market and then enforce that rights allocation. Thus, organized crime and corruption tend to go hand in hand. Indeed, Thomas Schelling, What is the Business of Organized Crime?, 40 AM. SCHOLAR 643 (1971), argues that organized crime is really monopolized crime, and both Paul H. Rubin, The Economic Theory of the Criminal Firm, in THE ECONOMICS OF CRIME AND PUNISHMENT 155 (Simon Rottenberg ed., 1973), and ANNELISE G. ANDERSON, THE BUSINESS OF ORGANIZED CRIME 58 (1979), contend that such criminal firms possess monopoly power because there are economies of scale in buying corruption from police and other government officials. CAVE & REUTER, INTERDICTOR'S LOT, supra note 12, present a model that raises the possibility of selective enforcement against novice rather than experienced drug smugglers that is also consistent with systematic corrupt enforcement practices.
The police executive knows from bitter experience that in committing his force to attack drug trafficking and drug use, he risks corruption and abuses of authority. Informants and undercover operations—so essential to effective drug enforcement—invariably draw police officers into close, potentially corrupting relationships with the offenders they are pledged to control.122

Prohibition policies and intense law enforcement efforts breed crime, so even if there is a criminogenic consequence of drug use per se, such policies may not reduce overall levels of crime. Indeed, the increased crime that arises when activities are carried out in black markets, and when criminal justice resources are reallocated to control that market, could easily outweigh any crime reducing impact of limiting drug consumption.123

VI. SUPPLIERS OF DRUG POLICY: THE ROLE OF BUREAUCRATIC SELF-INTEREST

Our story so far shows that the effectiveness of enforcement policy is undermined by the fact that drug users and suppliers will adjust their behavior to offset the effectiveness of enforcement, and that increasing enforcement appears to compromise public safety and foster corruption. Furthermore, growing evidence suggests a more cost-effective approach is available.124 A widely cited RAND study by Rydell and Everingham estimates that another dollar spent on drug treatment is seven times more cost-effective than another dollar spent on drug enforcement—if the objective is to reduce cocaine

122. Mark H. Moore & Mark A.R. Kleiman, U.S. DEP’T OF JUSTICE, THE POLICE AND DRUGS 2 (Perspectives on Policing No. 11, Sept. 1989). It would be useful to add the adjective “low-wage” to police officers in this quotation. In 1990 the mark-up on one kilogram of cocaine at the wholesale level was about five times the entry-level salary in large police departments and about 1.5 times the police chief’s salary. For a discussion of police salary structures and drug mark-ups, see Rasmussen & Benson, Economic Anatomy, supra note 11, at 117.

123. Resignato, supra note 11, at 686-87, finds that the primary cause of drug-related violence is that drugs must be sold in illegal markets, as explained above, supra note 112. His results show no important relationships between drug use and violence, but as drug enforcement increases, violent crime also increases. Furthermore, the study in Benson et al., Property Crime, supra note 118, includes a control for the size of the drug market in their trade-off model of property crime. The authors find that there is, in fact, a significant relationship between the size of the drug market and property crime, but that a reallocation of policing resources to reduce drug crime still leads to an increase in property crime. The reduction in property crime due to a reduction in drug market size is more than offset by the effects of reduced deterrence for property crime due to the reallocation of policing effort.

124. Using an innovative dynamic control model, Jonathan P. Caulkins et al., Price-Raising Drug Enforcement and Property Crime: A Dynamic Model, 71 J. ECON. 227 (2000), ignore the opportunity costs of police resources; but their estimates based on U.S. cocaine use suggest that, contrary to U.S. policy, “as use grows toward a steady state, enforcement intensity should decline.” Id. at 248. Nevertheless, they indicate that the optimum level of enforcement is “fairly sensitive” to what parameters are chosen for the model.
use.125 They conclude that criminal justice expenditures could be reduced by twenty-five percent, which would allow for a doubling of expenditures on treatment and a reduction of total expenditures on drug control of approximately $2 billion.126

From 1960 to 1998, drug arrests per capita rose 22-fold, from 26 per 100,000 population to 615 per 100,000.127 Given the apparent ineffectiveness of the war on drugs, its unintended consequences, and the availability of alternative strategies, what accounts for the substantial enforcement bias in U.S. drug policy?

A. Bureaucratic and Political Interests

In a representative democracy there is a tendency to expect that public opinion drives drug policy. This is not the case, as “every detailed study of the emergence of legal norms has consistently shown the immense importance of interest-group activity, not the ‘public interest,’ as the critical variable.”128 Drug war, the excessive application of enforcement that aggravates rather than mitigates the social consequences of drug use, is waged because it is in the interests of particular politically influential groups, including law enforcement bureaucracies and public officials.129 According to this view, legislators can act as moral entrepreneurs, but they are more generally “middlemen” whose actions are largely determined by interest groups, in-

126. Id. at xviii. Treatment is not a panacea, however, as Rydell and Everingham estimate that this reallocation of resources would change the total social costs (including enforcement) by about $2 billion, to about $36.7 billion per year, a 12.6% decline. Id. An alternative policy of providing treatment for all heavy users reduces the total social costs to $31.7 billion from an estimated $42 billion under the current policy, a 24.5% decline. Id.
127. This was not a slow, steady shift toward more drug arrests: the periods from 1965-70 and 1984-89 account for 68% of the total increase. See Rasmussen & Benson, Economic Anatomy, supra note 11, at 7.
128. William J. Chambliss & Robert B. Seidman, Law, Order, and Power 73 (1971). Robert P. Rhodes, The Insoluble Problems of Crime 13 (1977) argues that “as far as crime policy and legislation are concerned, public opinion and attitudes are generally irrelevant.” This contention is confirmed in Rasmussen & Benson, Economic Anatomy, supra note 11, at 122-27, for the 1984-1989 drug war. Drug arrests rose from 312 per 100,000 in 1984 to 538 in 1989. Id. at 6. Only 2% of Gallup Poll respondents thought drug abuse was the nation’s most important problem in January 1985. Id. at 123-24. By September 1988, this figure was 11%. After the Drug Czar position was created by the Anti-Drug Abuse Act of 1988, Pub. L. No. 100-690, 102 Stat. 4181 (codified as amended in scattered sections of 18 U.S.C.), polls reported a rapid rise in concern over drugs that peaked in November 1989 at 38%. Id. at 124. The level of concern eroded quickly and was 11% by March 1991. Id. at 125 fig. 6.1.
129. There are many models of bureaucratic behavior based on self-interest assumptions that have been developed by economists and political scientists. For a review of these models in the context of drug policy, see Rasmussen & Benson, Economic Anatomy, supra note 11, at 127-32.
cluding those engaged in the law enforcement process—police chiefs, sheriffs, and prosecutors.130

A number of motivations for demanding drug legislation have actually been identified for both bureaucratic and non-bureaucratic interest groups. Some studies have noted the incentives of professional organizations such as the American Pharmaceutical Association to create legal limits on the distribution of drugs (historically there was significant competition between pharmacists and physicians for the legal right to dispense drugs, for example),131 while others have focused on the disparate racial impacts of illicit drug laws and the desire by some groups to control racial minorities through the enforcement of such laws.132 More importantly, from the perspective stressed here, still other studies have emphasized that law enforcement bureaucrats have been a major source of demand for the criminalization of narcotics. After the Harrison Act of 1914,133 these same groups lobbied for passage of the Marihuana Tax Act of 1937134 and played an important role in the subsequent criminalization of this illicit drug.135

130. Countervailing interests include civil libertarian groups, defense attorneys, and groups such as the National Organization for the Reform of Marijuana Laws (NORML) and Families Against Mandatory Minimums (FAMM) that lobby for policy reform. The continuity of enforcement policy suggests that these forces do not prevail, in part because support of these groups would naturally expose lawmakers to competition claiming they were soft on drugs.

131. See Musto, supra note 100, at 13-14, 21-23; Thornton, supra note 11, at 56-60; Dorie Klein, Ill and Against the Law: The Social and Medical Control of Heroin Users, 13 J. DRUG ISSUES 31 (1983).


135. See Howard S. Becker, Outsiders: Studies in the Sociology of Deviance 135-45 (1963); Bonnie & Whitebread II, supra note 132; Jerome L. Himmelstein, The Strange Career of Marihuana: Politics and Ideology of Drug Control in America (1983); Alfred R. Lindesmith, The Addict and the Law (1965); Craig Reinarman, Constraint, Autonomy, and State Policy: Notes Toward a Theory of Controls on Consciousness Alteration, 13 J. DRUG ISSUES 9 (1983). In fact, as Thornton, supra note 11, at 62-66, and Patricia A. Morgan, The Political Economy of Drugs and Alcohol: An Introduction, 13 J. DRUG ISSUES 1 (1983), have stressed, all of the various self-interests mentioned above (bureaucrats, professionals from the American Medical Association and American Pharmaceutical Association, and groups attempting to suppress certain races or classes) interacted with still more groups (e.g., temperance groups and religious groups) to produce policies against drug use. Interest groups and bureaucratic entrepreneurs continue to dominate modern drug policy as well. These groups include “civil rights, welfare rights, bureaucratic and professional interests, health, law and order, etc.” Id. at 3. For instance, the pharmaceutical industry had a significant impact on the Comprehensive Drug Abuse Prevention and Control Act of 1970, 21 U.S.C. §§ 801-971 (2002). According to Reinarman, supra, at 19, “In that case as in most of the others, the state’s policy-makers were buffeted by law enforcement interests and professional interests . . . .”
Breton and Wintrobe explained that one bureaucratic strategy to compete for resources is to “generate” demand for a bureau’s own services through direct lobbying, policy manipulation, and the selective release of information and misinformation to other interest groups and the media. These strategies are followed because bureaus must compete with one another for the support and attention of sponsors (and individual bureaucrats must compete with other bureaucrats for benefits within a bureau) and because the control of resources is necessary before most of the subjective goals of bureaucrats can be achieved.

Breton and Wintrobe emphasized that bureaucratic release of both true and false information, or “selective distortion,” can play significant roles in bureaucratic policy advocacy. This has clearly been the case in the evolution of drug policy. For example, the bureaucratic campaigns leading up to the 1937 marijuana legislation “included remarkable distortions of the evidence of harm caused by marijuana, ignoring the findings of empirical inquiries.” The “reefer madness” scare traces to the misinformation propagated by the Bureau of Narcotics. Marijuana was alleged to cause insanity, to incite rape, and to cause users to develop delirious rages, making them irresponsible and prone to commit violent crimes. Factual distortions did not stop there, however. For instance, the bill was represented as one that was largely symbolic in that it would require no additional enforcement expenditures.

The evolution of drug policy since the initial legislation has also been, at least in part, shaped by bureaucratic competition, both between law enforcement and drug treatment bureaucrats over “ownership of the problem”—that is, over shares of federal, state, and local budgets—and between law enforcement bureaucracies themselves.


137. See ROBERT M. STUTMAN & RICHARD ESPOSITO, DEAD ON DELIVERY: INSIDE THE DRUG WARS, STRAIGHT FROM THE STREET (1992), for a description of the actual activities of a DEA agent, which reveals the tremendous amount of time and effort this agent spent competing for resources. The book also shows the significant role that politics play in determining the allocation of drug enforcement resources; its entire argument could be easily set in the context of the Breton-Wintrobe model of bureaucratic entrepreneurship.

138. BRETON & WINTROBE, supra note 136, at 39. This is possible, in part, because of the high costs of monitoring bureaus.


141. See supra note 135 and accompanying text.
As the perceived responsibility for some social ill (e.g., crime) is shifted from outside forces to the government and to the bureaucracy, bureaucrats seek to shift the blame elsewhere. Blaming crime on people crazed by drugs takes advantage of such an opportunity. As a consequence, a good deal of false or misleading information emanating from police bureaucrats about the relationship between drugs and crime has clearly characterized the evolution of drug policy. In fact, it was chiefly as a result of information promulgated by police that drug crime came to be widely held as the root cause of much of what is wrong with society. In particular, the contention that property crime is a major source of income for drug users (i.e., so drug use is thus the leading cause of property crime) has been made to justify political demands for the criminal justice system to “do something” about the drug/crime problem, demands that largely emanate from the police lobbies. In turn, that contention has been used to justify an emphasis on the control of illicit drug traffic as a means of general crime prevention. With these arguments as justification, state and federal legislators have been passing increasingly strict sentencing requirements for drug offenders, police have reallocated resources to make more drug arrests, and judges have sentenced increasingly large numbers of drug offenders to prison.

142. See Breton & Wintrobe, supra note 136, at 149-51.
143. Kaplan, supra note 139; see Kaplan, supra note 21; Lindesmith, supra note 135; Rasmussen & Benson, Economic Anatomy, supra note 11, at 119-50; Richards, supra note 139; Robert J. Michaels, The Market for Heroin Before and After Legalization, in DEALING WITH DRUGS: CONSEQUENCES OF GOVERNMENT CONTROL 289 (Ronald Hamowy ed., 1987).
144. See Randy E. Barnett, Public Decisions and Private Rights, 3 CRIM. JUST. ETHICS 50, 56 (1984) (book review) (explaining that specialists are not reliable guides for effective drug policy); supra notes 128 & 137 and accompanying text.
145. For instance, see the OFFICE OF NAT’L DRUG CONTROL POLICY, EXECUTIVE OFFICE OF THE PRESIDENT, NATIONAL DRUG CONTROL STRATEGY 2 (Jan. 1990).
147. Breton & Wintrobe, supra note 136, at 150-51, offer two reasons to explain why bureaucrats advocate policies of directly controlling “a source of blame” for a problem such as crime (e.g., alcohol prohibition, criminalization and prohibition of various drugs after 1914 and 1937, increased emphasis on drug control in the mid-1960s, and then again in the mid-1980s), even though such policies have a history of failure. First, there is always opposition to such policies, so when they fail, opponents can be blamed for not allocating sufficient resources to combat the problem. Second, because policy outcomes depend jointly on the inputs of several different groups and bureaus and the set of possible control methods is very large, when the subset selected fails, the bureaucrats can argue that: (1) while they advocated a control policy they favored a different subset of control tools (e.g., more severe punishment of drug offenders or greater spending on supply interdiction efforts) so they are not responsible for the failure, and/or (2) the other groups whose contributions were necessary to make the effort successful (e.g., witnesses, judges, legislators who approve prison budgets, other law enforcement agencies) did not do their share. Indeed, a policy can fail completely, while at the same time entrepreneurial bureaucrats expand their
Civil asset forfeiture procedures also provide a powerful motivation for law enforcement at all levels to increase drug arrests. As asset forfeiture laws provide law enforcement with a weapon to deter drug offenders, but they encourage more drug enforcement because, since the Federal Comprehensive Crime Act of 1984, most of the proceeds of these *in rem* proceedings go to the agency. Before discussing the implications of the allocation of seizures, note that the 1984 Crime Bill’s change in asset forfeiture law was a bureaucratically-demanded innovation that was justified as a means of achieving expanded inter-bureau cooperation.

Forfeiture policies were also just-
circumvent state laws and constitutions prohibiting certain forfeitures or limiting law enforcement use of seizures. For example, North Carolina law requires that all proceeds from [the sale of] confiscated assets go to the County School Fund. Law enforcement agencies in North Carolina, and in other [jurisdictions] where state law limited their ability to benefit from confiscations, began using the 1984 federal legislation to circumvent [the restrictions] by routinely arranging for federal “adoption” of forfeitures [so the seized assets could be repatriated to] the state and local law enforcement agencies. [Note that this adoption process raises significant doubts about the inter-agency cooperation justification for this legislation, other than the political cooperation that allowed police agencies at all levels to expand their budgets without going through the general budgeting process.] As education bureaucrats and others affected by this diversion of benefits recognized what was going on, they began to advocate a change in the federal law. They were successful, [at least initially]: the Anti-Drug Abuse Act of 1988, passed on November 18, 1988, changed the asset forfeitures provisions that had been established in 1984. Section 6077 of the 1988 [Act] stated that the attorney general must assure that any [seized asset] transferred to a state or local law enforcement agency “Is not so transferred to circumvent any requirement of State Law that prohibits forfeiture or limits use or disposition of property forfeited to state or local agencies.” This provision was designated to go into effect on October 1, 1989, and the Department of Justice interpreted it to mandate an end to all adoptive forfeitures.

State and local law enforcement officials immediately began advocating the repeal of Section 6077. Thus, the U.S. House Subcommittee on Crime heard testimony on April 24, 1989, advocating repeal of Section 6077 from such groups as the International Association of Chiefs of Police, the Florida Department of Law Enforcement, the North Carolina Department of Crime Control and Public Safety, and the U.S. Attorney General’s Office. Perhaps the most impassioned plea for repeal was made by Joseph W. Dean of the North Carolina Department of Crime Control and Public Safety, who admitted both that law enforcement bureaucracies were using the federal law to circumvent the state’s constitution and that without the benefits of confiscations going to those bureaus, substantially less effort would be made to control drugs.

“Currently the United States Attorney General, by policy, requires that all shared property be used by the transfer for law enforcement purposes. The conflict between state and federal law [given Section 6077 of the 1988 Act] would prevent the federal government from adopting seizures by state and local agencies.

. . . This provision would have a devastating impact on joint efforts by federal, state and local law enforcement agencies not only in North Carolina but also in other affected states . . .

Education is any state’s biggest business. The education lobby is the most powerful in the state and has taken a position against law enforcement being able to share in seized assets. The irony is that if local and state law enforcement agencies cannot share, the assets will in all likelihood not be seized and forfeited. Thus no one wins but the drug trafficker . . .

. . . If this financial sharing stops, we will kill the goose that laid the golden egg.”

This statement clearly suggests that law enforcement agencies focus resources on enforcement of drug laws because of the financial gains for the agencies arising from forfeitures.


Id. Citations omitted.
tified because they imply that the proceeds from drug crime are used to recoup public monies spent combating drug crime, as emphasized in a manual designed to help jurisdictions develop a forfeiture capability. \(151\) While pointing out that less tangible law-enforcement benefits (such as deterrence) should be counted as benefits, the manual emphasizes that the determining factor for pursuit of a forfeiture is the "jurisdiction's best interest." \(152\) This interest, of course, is viewed from the perspective of law enforcement agencies, a view that might put more weight on the benefits perceived by agency heads and somewhat less weight on the uncertain community-wide benefits of deterrence. Research indicates that police department discretionary budgets rise when they seize assets and that departments respond to this incentive by increasing drug arrests relative to arrests for other offenses. \(153\) Thus, while civil asset forfeiture programs may be an effective tool in a war on drugs, it is clear that they provide powerful incentives for police agencies to increase drug enforcement relative to other activities. Indeed, there is no research to demonstrate that seizure activity has a major deterrent effect, and there is some evidence that, at least in some jurisdictions, a substantial amount of assets are seized from innocent people. \(154\)

Consider the seizures made in Volusia County, Florida, for instance. The Sheriff's Department had a drug squad that collected over $8 million (an average of $5,000 per day) from motorists on Interstate 95 during a forty-one-month period between 1989 and 1992. \(155\) These seizures were “justified” as part of the war on drugs.

---

152. Id. at 40 (emphasis added).
153. See Mast et al., supra note 11, at 287-89. Benson, Rasmussen and Sollars show that forfeiture activity raises budgets, Benson et al., Police Bureaucracies, supra note 11, at 37, refuting the possibility that these funds are fungible and offset by reduced regular budget allocations. See also Eric Blumenson & Eva Nilsen, Policing for Profit: The Drug War’s Hidden Economic Agenda, 65 U. Chi. L. Rev. 35, 56 (1998); John L. Worrall, Addicted to the Drug War: The Role of Civil Asset Forfeiture as a Budgetary Necessity in Contemporary Law Enforcement, 29 J. Crim. Just. 171 (2001).
154. Blumenson & Nilsen, supra note 153, at 46-47, refer to the “ancient legal fiction” that presumes the property is guilty, thereby allowing police to seize assets even when the owner is not charged for a crime. Dramatic examples of police seizure activity can be found in an Arts and Entertainment Network investigative report on civil asset forfeiture entitled Seized by the Law (Kurtis Productions, Ltd. for A & E television broadcast 1996).
155. In a Pulitzer prize winning series of Orlando Sentinel articles printed during June 14-16, 1992, Jeff Brazil and Steve Berry describe in vivid detail the asset seizure program in Volusia County. See Jeff Brazil & Steve Berry, Series: Tainted Cash or Easy Money?, ORLANDO SENTINEL, June 14-16, 1992. Many other examples of abuses of police discretion in the pursuit of forfeitures can also be cited. For instance, Dennis Cauchon and Gary Fields demonstrate this in a special report titled Abusing Forfeiture Laws in USA TODAY, May 18, 1992. More recently, narcotics task forces in Texas were revealed to have seized more than $194 million between 1987 and 2000. See Jim Henderson, Big Numbers Don’t Add Up to Success in Texas War on Drugs, HOUS. CHRON., Dec. 24, 2000, at State 1.
but most Volusia County seizures actually involved southbound rather than northbound travelers, suggesting that the drug squad was more interested in seizing money than in stopping the flow of drugs. More significantly in this context, no criminal charges were filed in over 75% of the county’s seizure cases. Traffic citations were not even issued, let alone drug related charges. Indeed, it appears that a substantial amount of money was apparently seized from innocent victims. Three-fourths (199) of Volusia County’s seizures made during this period were contested. Money was not returned even when the seizure was challenged, no proof of wrongdoing or criminal record could be found, and the victim of the seizure presented proof that the money was legitimately earned. The sheriff employed a forfeiture attorney at $44,000 per year to handle settlement negotiations. By 1992 only four people had gotten their money back, one lost at trial and was appealing, and the rest settled for 50–90% of their money after promising not to sue the sheriff’s department. How many were drug traffickers? No one knows, since no charges were filed and no trials occurred; but, it is clear that several were innocent victims. Thus, asset seizures are not simply a way of using ill-gotten gains to finance criminal justice.

The Volusia County Sheriff’s Department is not the only law enforcement agency that has benefited from asset seizure laws. “Over 90% of the police departments serving a population of 50,000 or more, and over 90% of the sheriffs’ departments serving 250,000 or more residents, received money or goods from a drug asset forfeiture program” in 1990. Moreover, asset forfeiture by law enforcement

157. Id.
158. Id.
160. Id.
161. A twenty-one-year-old naval reservist had $3,989 seized in 1990, for instance, and even though he produced Navy pay stubs to show the source of the money, he ultimately settled for the return of $2,988, with 25% of that going to his lawyer. In similar cases the sheriff’s department kept $4,750 out of $19,000 (the lawyer got another $1,000); $3,750 out of $31,000 (the attorney got about 33% of the $27,250 returned); $4,000 of $19,000 ($1,000 to the attorney); $6,000 out of $36,990 (the attorney’s fee was 25% of the rest); and $10,000 out of $38,923 (the attorney got one-third of the recovery). Brazil & Berry, supra note 156.
162. BRIAN A. REAVES, U.S. DEP’T OF JUSTICE, DRUG ENFORCEMENT BY POLICE AND SHERIFFS’ DEPARTMENTS, 1990, at 1 (May 1992). As of September 30, 2001, the ONDCP reported that its asset forfeiture fund has a total of 20,703 assets valued at $831.5 million. OFFICE OF NAT’L DRUG CONTROL STRATEGY, EXECUTIVE OFFICE OF THE PRESIDENT, NATIONAL DRUG CONTROL STRATEGY: FY 2003 BUDGET SUMMARY 88 (Feb. 2002). Worrall, supra note 153, at 177-79, provides survey results indicating that local law enforcement agencies are dependent on asset seizure to supplement their budgets.
agencies has become increasingly controversial throughout the nation. Highly publicized criticism in the print and electronic media has raised constitutional issues such as the erosion of Fourth Amendment rights, protection of innocent parties, and a lack of proportionality of punishment to the crime.\textsuperscript{163} Whether large portions of the seizures come from criminals cannot be determined because many do not involve arrests, and the costs associated with recovering wrongfully seized assets can run into thousands of dollars.

Agency interest in budget enhancement is not limited to seizing the assets of citizens; drug policy is a significant vehicle by which public and quasi-public agencies compete for resources. Police agencies actively promulgate false and misleading information about the relationship between drugs and crime, blaming unsolved crime on people under the influence of drugs in order to increase their budget allocations. Increasing drug arrests tend to raise total arrests but at the cost of less enforcement of other offenses, creating for police agencies a “virtuous cycle” of more output as measured by arrests and more need for police as measured by crime, both of which lead to higher budget allocations.\textsuperscript{164} Further, innovations in drug markets are routinely presented by drug policy entrepreneurs as a new public policy problem because of the chemical, physiological, or psychological novelty of the new drug. Zimring and Hawkins note that this occurs because “allegations of a drug’s uniqueness can be used as a rhetorical device to shield proponents of a prohibitory policy from counterarguments based on the history of earlier efforts at the state regulation of other substances or of the same substance in different forms or settings.”\textsuperscript{165} By alleging unique threats posed by each new drug, criminal justice interests can claim as irrelevant any references to past policy experiences with any other drug. In a stunning exposé of bureaucratic politics designed to increase his own agency’s budget, Drug Enforcement Agent Robert Stutman described how he created a media and political campaign to generate hysteria about what was alleged to be an extraordinarily destructive new drug, crack cocaine.\textsuperscript{166}

\begin{footnotesize}
\begin{enumerate}
\item[163.] E.g., \textit{Seized by the Law}, supra note 154; \textit{War on Drugs, a War on Ourselves} (ABC television broadcast, July 30, 2002) (emphasizing the perverse incentives generated by asset forfeiture laws); \textit{supra} note 155 and accompanying text.
\item[164.] Police have a vested interest in keeping crime rates relatively high: if crime rates drop too much, support for more police and larger budgets wanes, and “[l]ike all bureaucracies, criminal justice agencies can hardly be expected to implement policies that would diminish their importance.” Michael E. Milakovich & Kurt Weis, \textit{Politics and Measures of Success in the War on Crime}, 21 CRIME \& DELINQ. 1, 10 (1975).
\item[165.] FRANKLIN E. ZIMRING \& GORDON HAWKINS, \textit{THE SEARCH FOR RATIONAL DRUG CONTROL} 51 (1992).
\item[166.] \textit{STUTMAN \& ESPOSITO}, supra note 137; see RASMUSSEN \& BENSON, \textit{ECONOMIC ANATOMY}, \textit{supra} note 11, at 141-46, for a discussion of Stutman and Esposito’s work in the context of the rise in drug arrests between 1984 and 1989.
\end{enumerate}
\end{footnotesize}
ers,\textsuperscript{167} uneducable children,\textsuperscript{168} and an unprecedented crime wave.\textsuperscript{169} By the time the scientific community can evaluate such claims and demonstrate that they are either false or exaggerated, drug market entrepreneurs can introduce a new product variety that police can depict in the same fashion.

Advocacy for a war on drugs is not limited to police agencies, however. Legislators and prosecutors reap immediate political benefits by appearing tough on drugs, while the social costs of these policies are not readily apparent to their constituents and occur over a long time period that is relatively less salient to political entrepreneurs seeking re-election.\textsuperscript{170} Tough sentences for drug offenders also increase the demand for prison beds,\textsuperscript{171} providing an incentive for contractors who build prisons and both public and private providers of prison services to advocate tough drug policies. Treatment providers also have an incentive to advocate harsh penalties for drug use, because they generate demand for their services through court ordered treatment and private spending for treatment driven by the fear of punishment.\textsuperscript{172}

\textsuperscript{167} The National Household Survey of Drug Abuse in 1998 reported that the 25-34 age group, in 1996, had the highest lifetime prevalence of crack use, 4.4%. That cohort reported a last month prevalence of 0.5% that year, about 11.4% of the lifetime rate. The same cohort reported a monthly rate of marijuana and/or hashish use that was 16.6% of the lifetime rate, suggesting that the tendency to be a regular user (as measured by use during the last 30 days) among crack users is not substantively different from that of marijuana users. CJ STATISTICS 2000, supra note 55, at 260, 261 tbls. 3.92-3.94.

\textsuperscript{168} There is some evidence that crack cocaine use among pregnant women poses some education problems for their children due to low birth weight because such infants are more likely to require special education. See H. Naci Mocan & Kudret Topylan, \textit{Illicit Drug Use and Health: Analysis and Projections of New York City Birth Outcomes Using a Kalman Filter Model}, 62 S. Econ. J. 164, 164 (1995); see also BENNETT & DILORENZO, supra note 139, at 243-45 (arguing that pregnant crack-using mothers are less likely to get badly needed prenatal care, however, because the hysteria around crack babies may cause them to believe that their children are hopeless).

\textsuperscript{169} But see Goldstein, supra note 29; Johnson et al., supra note 98.


In recent years many member nations of the European Community have been experimenting with drug reform. Although the sources of this emerging trend are not well researched at this time, it is worth noting that these countries, unlike the U.S., do not have elected sheriffs and prosecutors. This may reduce the incentives of some agencies to advocate high levels of drug enforcement.

\textsuperscript{171} In 1996, 36.8% of all felony defendants in the largest seventy-five counties were convicted on drug charges, so punishments for drug offenders have a significant effect on the overall demand for prison beds. CJ STATISTICS 2000, supra note 55, at 460 tbl. 5.47.

C. Tragedy in the Criminal Justice Commons

Criminal justice resources are scarce. In fact, competing demands for their use substantially exceed their supply. This means that public law enforcement must be rationed, and since they are not rationed by price in a market setting, some other allocating method must be used. Police, prosecutors, public defenders, probation services, and prisons are allocated in a common pool environment, implying that there are few incentives to use these resources efficiently. The commons problem emerges because no individuals or institutions “own” these public resources, and therefore decision-makers do not suffer the full consequences when these resources are not put to their highest and best use. Since the best use of these public resources involves social benefits or costs that are not captured or borne by the decision-makers, there is a tendency to undervalue these benefits and costs relative to those which do accrue to the agency and its administrators.

Public officials are presumably characterized by the same utility maximizing behavior that motivates people in private markets. The institutional framework of public officials may differ from that of private sector employees, but their fundamental objectives should include many that are common to both groups such as job security, prestige, discretion, advancement, leisure, and promotion of whatever they believe is in the public interest. Budget allocations to police agencies are likely to rise when they increase drug arrests, which is a measure of “performance” in the police budgeting process. Furthermore, an important measure of the need for police resources in this process is the amount of crime reported in the jurisdiction. As explained above, when police allocate more of their limited resources

---

173. Some might argue that criminal justice resources are “free” because everyone supposedly has equal access to them, but because all demands cannot be met in a timely fashion, there must be some rationing criteria, even if it is not explicit. See BENSON, supra note 146, at 97; BRUCE L. BENSON & LAURIN A. WOLLAN, JR., JAMES MADISON INST., PRISON OVERCROWDING AND JUDICIAL INCENTIVES (1989); Carl S. Shoup, Standards for Distributing a Free Governmental Service: Crime Prevention, 19 PUB. FIN. 383 (1964).

174. For a description of the “Tragedy of the Commons” see Garrett Hardin, The Tragedy of the Commons, 162 SCIENCE 1243-48 (1968). For applications of the commons analysis to the criminal justice system, see BENSON, supra note 146; BENSON & WOLLAN, supra note 173, at 2-8; RICHARD NEELY, WHY COURTS DON’T WORK, 164-86 (1982); RASMUSSEN & BENSON, ECONOMIC ANATOMY, supra note 11, at 17-37; Randy E. Barnett, Pursuing Justice in a Free Society: Part Two—Crime Prevention and the Legal Order, 5 CRIM. JUST. ETHICS 30, 31-35 (1986).

175. See supra note 9 for references to the “public choice” literature that emphasizes the role of incentives in bureaucratic agencies.

176. See BRETON & WINTROBE, supra note 136, at 27. From this list, increasing budgets would seem particularly important since many other objectives might be well served with more agency resources.

177. See Sollars et al., Drug Enforcement and Deterrence, supra note 118, at 36.

178. See Benson et al., Property Crime, supra note 118, at 679.
to drug offenses, the total number of arrests can increase and the number of property crimes also rises due to the reduced deterrent effect occurring because of the reallocation of police effort.\textsuperscript{179} By allocating too many resources to drug enforcement, police agencies can increase their measured productivity (total arrests) and generate a higher measure of need for police services (reported crime), with the net effect being added political support for increased budget allocations.\textsuperscript{180}

Enforcement dominates drug policy not because it is the most effective policy, but because it best serves the interests of the people who work in the institutions that dominate the drug policy decision-making process.

\section*{VII. Toward an Optimal Level of Drug Enforcement}

Drug policy alternatives are a continuum of options ranging from complete legalization to what is commonly called a “war on drugs,” and criminal enforcement plays an increasingly important role as policy approaches the terminal state of drug war. Suspension of civil rights\textsuperscript{181} and confiscation of property\textsuperscript{182} have historically been domestic consequences of U.S. wars, and they are integral parts of the effort to combat drugs.\textsuperscript{183} Drug war also involves undermining the legal principle of proportionality because punishments are often far in excess of any plausible estimate of the social harm.\textsuperscript{184} Reducing such abuses of the current war on drugs, without necessarily dealing with

\begin{itemize}
\item \textsuperscript{179} See \textit{supra} note 118 and accompanying text.
\item \textsuperscript{180} Consistent with the “public choice” perspective, police agencies are the primary source of information that is used to determine agency budgets. See Barnett, \textit{supra} note 144; Michaels, \textit{supra} note 143.
\item \textsuperscript{181} The Sedition Act (1798) was probably the first example of compromising the Bill of Rights when the national security was allegedly threatened. Sedition Act, ch. 74, 1 Stat. 596 (1798).
\item \textsuperscript{182} Leonard W. Levy, A License to Steal: The Forfeiture of Property 50-61 (1996).
\item \textsuperscript{183} When referring to efforts to reduce problem drug use, it might be useful to choose our words carefully. Conditions of war permit undermining of civil liberties, whereas the term campaign is a synonym without the same connotations. Another concession to clarity of purpose would eliminate the title “czar” from the drug policy lexicon—a title that connotes absolute power and is certainly not associated with enlightened public policy or even competence in serving the self-interest of the ruling institutions.
\item \textsuperscript{184} Harsh penalties for possession of illicit drugs have been justified, for example, because drugs allegedly impose heavy moral and welfare costs on society. James Q. Wilson, \textit{Drugs and Crime}, in \textit{Drugs and Crime, supra} note 12, at 527 n.2, argues that these costs associated with heavy use can be “so large that society should bear the heavy burden of law enforcement, and its associated corruption and criminality, for the sake of keeping the number of people regularly using heroin and crack as small as possible.” Wilson’s view is in direct contradiction to that of harm reduction advocates, many of whom argue that the criminalization of drugs generates more social harms than drug use itself. Drug war is generally defended by an appeal to abstract costs and moral values rather than empirically grounded arguments.
\end{itemize}
the politically contentious issue of an individual’s right to consume drugs if they do not affect others, is another task for drug policy reform.\textsuperscript{185}

Drug enforcement is clearly excessive in the efforts to reduce the harms of drug use, but this reliance on a drug war is no surprise given the institutional environment in which drug policy is formed. Most influential players in the formulation of drug policy, and all the dominant ones such as legislators, police agencies, and prosecutors, have incentives to conduct a war on drugs even though more effective policies are readily available. In the following Section we examine fundamental institutional changes that could be the foundation for a more rational drug policy. Few of our suggestions relate to specific changes in drug laws or policies; most are aimed at leveling the playing field so drug policy can be formulated in a way that better reflects objective analysis and community values, rather than the current institutional structure that favors political and bureaucratic interests.

There is no magic bullet that will solve the “drug problem,” however it is defined. Eschewing the extremes of the moral positions taken by some policy advocates (i.e., those stating that people have a right to consume drugs or those claiming that any drug use is immoral because it causes great individual and social harm), reasonable people can differ as to whether the consequences of the myriad of policy options that may influence drug use are, on net, desirable. Drug war advocates may argue that better information is needed about the consequences of alternative policies before drug policy entrepreneurs and legislators will consider reforming drug policy. However, justifying current policy on the grounds that we do not know enough about the policy alternatives, of course, is circular reasoning if we are not willing to experiment and discuss the alternatives and analyze their consequences because of our commitment to enforcement-based policy. Even if the research base for assessing all options is not entirely adequate, a careful delineation of the strengths and weaknesses of the existing knowledge about the consequences of alternative drug policies leads MacCoun and Reuter to conclude that “ignoring specific proposals, the desirability of major reform has a reasonable empirical and ethical basis. To scorn discussion and analysis of such major change, in light of the extraordinary problems

\textsuperscript{185.} See THOMAS SZASZ, THE THERAPEUTIC STATE: PSYCHIATRY IN THE MIRROR OF CURRENT EVENTS 271 (1984), for a prominent argument that the state has no legitimate interest in regulating drug use and the desire to “self medicate.” For arguments that the issue is more complicated because family members are economically, socially, and spiritually impoverished by this behavior and that there are social ramifications of these individual choices, see KLEIMAN, supra note 28, at 48-49, and Wilson, supra note 184, at 523.
associated with current policies, is frivolous and uncaring.”

Given our economic analysis suggesting that increased enforcement is not likely to substantially reduce drug use in the presence of persistent demand, but that greater enforcement can increase the harms of drug use through product substitution by users and sellers and compromise public safety, a strong case can be made that alternative drug policies should be explored.

Considering the drug policy experimentation in various European countries and in some U.S. jurisdictions, a large number of options might lead to a more efficacious drug policy. Initiatives that have been tried to reduce problems associated with hard drugs such as heroin include no penalties for use, needle exchange programs, experiments with methadone maintenance and heroin maintenance, compulsory treatment, and treatment as an alternative to prison. There is increasing appreciation in Europe of making a clear distinction between hard and soft drugs, in part because the harms of soft drugs such as marijuana and hashish are relatively modest, but also because it seems desirable to separate these markets from those for hard drugs. Some European initiatives and experiments do not provide definitive information about their efficacy, and some might be inappropriate for some local U.S. jurisdictions given the great variation in the intensity and type of drug use among localities in the U.S. Nevertheless, experimentation with alternative policies is clearly warranted. There have been nascent attempts at medical marijuana in the United States, and several U.S. states have experimented with the decriminalizing of possession of small amounts of marijuana. These alternative policies that have been used both here and abroad provide information and models of potential reform, but they are often scorned and rarely integrated into U.S. drug policies whose avowed purpose is to diminish significant problems of substance abuse.

The crucial question is how the U.S. can wean itself from its excessive reliance on drug enforcement policy and allow jurisdictions to at least carefully consider alternative policies. Some analysts may suggest that police agencies and legislators will not consider alterna-

186. MacCoun & Reuter, supra note 14, at 409.
187. Id. at 209.
rative policy options until public opinion is firmly behind such efforts. The fact is that escalating enforcement has led rather than followed public opinion, however, so it does not seem plausible that policy reform must await changes in public opinion. Indeed, where public opinion clearly supports policy change, such as the medical marijuana movement and its successful referenda in several states, federal drug enforcement authorities have resisted, often with at least tacit support from state and local law enforcement. Given the incentives outlined in Section VI that encourage legislators, criminal justice officials, and police agencies to bias drug policy toward enforcement, a serious discussion of policy alternatives that might influence public opinion is not likely to occur. Our analysis suggests that significant drug policy reform will not emerge from a battle over policy alternatives fought in the court of public opinion because public opinion is too easily swayed by misleading information released by political and bureaucratic interests. Nor will more careful policy evaluations turn the tide. The problem to be confronted is that too many U.S. elected officials and employees of public agencies have incentives to conduct a war on drugs, to ignore or resist policy alternatives, and most assuredly not to explore alternatives that might more effectively address the problems generated by substance abuse.

A. A Federalist Drug Policy

A basic tenet of economic theory is that individuals in all institutional environments respond to incentives and constraints. We have seen in Section VI that the criminal justice system is a commons in which virtually every decision-maker can reap benefits from policies while not directly bearing the costs. Keys to reforming drug policy, in our view, are to be found in changing the institutional environment in such a way that agencies, bureaucrats, and legislatures can be more closely held responsible for the costs of their policies. Correspondingly, institutional reform is required to prevent drug policy moguls and criminal justice authorities from defending policies that are not just ineffective, but that also generate substantial social costs simply because they are mandated by some higher authority. Key to these reforms, in our view, is the devolution of drug policy from the federal government to state and local jurisdictions. In what follows we make no claim as to what policies should be chosen (even though we both have strongly held and sometimes conflicting opinions about

190. Supra note 128 and accompanying text.
191. Supra text accompanying note 128.
192. With the exception of New Mexico, where Republican Governor Gary Johnson has advocated drug policy reform, grass roots organizations have initiated most of the proposals for medical marijuana, treatment for drug offenders, and other "harm-reduction" measures.
the “best” policy), but argue that only by decentralizing drug policy will we stimulate policy analysis, discussion, and innovation that result in policies that will reflect anything close to their actual opportunity costs. When decision-makers are responsible for most of the costs imposed by their policies, they are more likely to choose policy options that generate the highest social benefits.

B. Reform International Drug Policy Agreements

Throughout the twentieth century, the United States has been the guiding force behind the development of an international drug policy that has heavily influenced domestic policy. Three United Nations’ agreements currently compose the system of global prohibition: the Single Convention on Narcotic Drugs of 1961, the Convention on Psychotropic Substances of 1971, and the Convention Against Illicit Traffic in Narcotic Drugs and Psychotropic Substances of 1988. The first of these conventions established the worldwide drug temperance ideology that characterizes U.S. drug policy by banning all non-medical uses of drugs, with particular emphasis on cannabis, cocaine, and opium. The 1971 Convention added psychotropic substances such as LSD and peyote to the list of drugs that are to be prohibited except for research purposes but also allowed for treatment and other alternatives to criminal sanctions.

Drug war was not explicitly mandated as the policy approach for prohibition, but then the Vienna Convention, unlike the previous agreements, urged signatories to “provide for maximizing the use of criminal law and to ensure that the goal of deterrence is adequately
taken care of.” This convention is highly consistent with America’s drug war, but there is an increasing divergence from this policy among European nations as reflected in substantial experimentation with harm reduction policies and strategies to decouple soft and hard drugs. Nadelman contends that the U.S. role is of “exceptional scale and scope,” and argues that its international efforts are the product of the failure of domestic drug policy. The consequence is that American drug policy has a “Catch-22” character. Advocates of the U.S. drug war can argue that international agreements, which America engineered, require that continuation of excessive enforcement and that reform efforts abrogate these commitments. In the meantime, other industrialized nations are taking small steps toward drug policy independence.

America’s unwavering commitment to abide by the international agreements it engineered as long as four decades ago freezes drug policy in time and is, in a sense, a commitment to ignorance since it discards new evidence in favor of past prejudice. Advances in knowledge about the physiological effects of various drugs, the consequences of alternative policy regimes, and the unintended consequences of excessive enforcement cannot influence a policy that is committed to past perspectives, circumstances, and information. Similarly, new understanding of the pharmacological properties of drugs and changing socio-economic characteristics of drug users should also be integrated into a dynamic drug policy. Thus, a first step toward rationalizing U.S. drug policy is to reform its international drug treaty commitments, allowing other nations the right to determine a drug policy they deem appropriate and, in the process, free America to develop a dynamic drug policy that is appropriate for the treatment of a multi-faceted and ever changing problem.

199. Albrecht, supra note 196, at 55.
200. Supra text accompanying note 188.
201. Ethan A. Nadelmann, Global Prohibition Regimes: The Evolution of Norms in International Society, 44 INT’L ORG. 479, 508 (1990) (“American promoters of the drug control regime have argued that their international efforts are necessary to reduce the extent and costs of drug abuse in the United States.”). See Bruce H. Bullington, Christopher P. Krebs & David W. Rasmussen, Drug Policy in the Czech Republic, in ILICIT DRUGS IN EUROPE 73 (A. Springer & A. Uhl eds., 2000), for an example of American pressure in the drug policy of other nations. The Netherlands, Germany, Spain, Italy, and England are among the European nations that are declaring some independence from the Single Convention, despite pressure from the U.S. MACCOUN & REUTER, supra note 14, at 205-07.
202. Recall that the protagonist in Joseph Heller’s 1955 novel with this title would only be able to leave the army in World War II if he were declared insane, but any attempt to get out of the insane war situation would be unsuccessful because it was obviously the same thing to do. JOSEPH HELLER, CATCH-22 (1961).
203. The unilateral U.S. withdrawal in 2001 from the 1972 Anti-Ballistic Missile Treaty illustrates that nations often desire to change international agreements over time. Such flexibility is desirable because circumstances change over time. See, e.g., Barbara Ko-
C. Decentralize National Drug Policy

Despite the commitment to a federal drug policy, the extent and type of substance abuse varies enormously among policing jurisdictions across the nation, suggesting that there is no uniform “drug problem.”\textsuperscript{204} In this light, federal intrusion into state and local efforts to deal with local production or use of drugs is unjustifiable, and the case for state sovereignty in regulating substance abuse is compelling.\textsuperscript{205} That federal drug policy has a chilling effect on state and local innovation is apparent from the federal government’s resistance to state based legalization and medical marijuana initiatives.\textsuperscript{206} Nor should the federal government help states enforce their drug laws through block grants. Political processes inevitably cause such funds to be widely distributed among jurisdictions, guaranteeing that the funding formula will assure that places without a serious drug problem will “find” a problem in order to receive funds.\textsuperscript{207} Since marijuana is the illicit drug that is most widely used, such assistance provides local law enforcement officials with an incentive to conduct their drug war against this relatively benign drug because to do otherwise is to forfeit federal grants for law enforcement.

Even in the absence of a fundamental reinterpretation of international treaty arrangements, removing marijuana and hashish from the federal list of Schedule I substances is a reform with enormous

\textsuperscript{204} John G. Haaga & Peter Reuter, The Limits of the Czar’s Ukase: Drug Policy at the Local Level, 8 YALE L. & POL’Y REV. 36 (1990), point out that only marijuana and cocaine are used throughout the country and that most other drugs have been concentrated in isolated areas. More recent data from the Drug Arrest Monitoring System is consistent with this observation. For example, among men in the thirty-four cities in this testing program, in 1999 opiate use varies from 1.4% in Ft. Lauderdale to 20.1% in Chicago, two cities with virtually identical rates of positive tests for cocaine. Among these cities cocaine use rates among adult male arrestees varies from 13.7 to 51.3%, while the range for marijuana use is from 28% to 51.2%. CJ STATISTICS 2000, supra note 55, at 394 tbl. 4.30.

\textsuperscript{205} See, e.g., United States v. Lopez, 514 U.S. 549, 564 (1995) (noting that “States historically have been sovereign” in the area of crime control); United States v. Bass, 404 U.S. 336, 349 (1971) (“Congress has traditionally been reluctant to define as a federal crime conduct readily denounced as criminal by the States. This congressional policy is rooted in the same concepts of American federalism that have provided the basis for judge-made doctrines.”); see also William H. Rehnquist, The 1998 Year-End Report of the Federal Judiciary, 31 THIRD BRANCH 1, 2 (1999) (“Federal courts were not created to adjudicate local crimes, no matter how sensational or heinous the crimes may be. State courts do, can, and should handle such problems.”).

\textsuperscript{206} See United States v. Oakland Cannabis Buyers’ Coop., 532 U.S. 483, 498-99 (2001) (holding that medical uses of marijuana were prohibited because the legislature’s intent was clear in the Controlled Substance Abuse Act, which classifies marijuana as a Schedule I substance). Marijuana use was expressly prohibited in any way except for government approved research projects.

\textsuperscript{207} See ZIMRING & HAWKINS, supra note 165, at 166.
potential benefits.\textsuperscript{208} Accounting for 44\% of all drug arrests in 1997,\textsuperscript{209} significant savings of police, prosecutorial, and prison resources might be achieved by changes in the criminal treatment of cannabis. Since almost all of these arrests are made by local police officers and marijuana possession is illegal to some degree in all states, the real impact of this policy change would appear only if states alter their enforcement practices or change their laws when they are empowered to do so. An atmosphere of candid policy discussion uncharacteristic of the drug war might arise, however, if politicians in localities and states are responsible for both the budgets for drug enforcement and the decision regarding how to deal with individual drugs. Furthermore, as one state experiments with an alternative policy, others may adopt innovations that are deemed successful when compared to current practices.\textsuperscript{210} Marijuana policy reform is more likely given the enormous benefits that could accrue to citizens who can be casualties of the drug war.\textsuperscript{211} Foremost among potential casualties are the millions of high school youth who have used marijuana.\textsuperscript{212} Normal youth experiment with drugs,\textsuperscript{213} but they are probably physically safer using marijuana than the other primary drug of choice, alcohol,\textsuperscript{214} and drug war rhetoric about marijuana being a gateway drug\textsuperscript{215} undermines the legitimate concerns about drug abuse.

Removing marijuana from Schedule I would allow the states to operate as “laboratories for democracy” so they can experiment with marijuana reforms that are consistent with their political climate. It is interesting that we can have a spirited debate on tobacco policy but it is currently difficult to do so with respect to drugs. It appears

\textsuperscript{208} As noted below, the most fundamental reform would have the current list of Schedule I substances as being illegal for importation into the U.S., and perhaps illegal to transport between states. State law would otherwise prevail.

\textsuperscript{209} CJ STATISTICS 2000, supra note 55, at 393 tbl. 4.29. Over 86\% of these arrests were for possession. It is interesting to note that marijuana’s share of drug arrests has trended downward since 1982 when this drug accounted for 72\% of these arrests. \textit{Id.}

\textsuperscript{210} See DAVID OSBORENE, LABORATORIES OF DEMOCRACY (1988) (examining the proposition that decentralized policy making in the states leads to more innovative and cost effective public policy). These laboratories, in effect, provide policy demonstrations for other jurisdictions.

\textsuperscript{211} Advocates of drug war, by ignoring unintended consequences that in the rhetoric of war are called collateral damage, seem to assume that only persons directly involved in the drug trade would take casualties.

\textsuperscript{212} Over 23\% of members of the high school class of 1999 used marijuana during the thirty days prior to the survey, and 49.7\% had experimented with the drug at least once. CJ STATISTICS 2000, supra note 55, at 246 tbl. 3.72.

\textsuperscript{213} \textit{Supra} note 44 and accompanying text.

\textsuperscript{214} Over 74\% of the members of the class of 2000 drank alcohol in the last 12 months, and 50\% report use during the last thirty days. CJ STATISTICS 2000, supra note 55, at 245 tbls. 3.70-3.71. Further, 46.7\% of 1999 college students reported that they were frequent binge drinkers in high school. \textit{Id.} at 256 tbl. 3.83. In 1998, 39\% of all traffic fatalities were in alcohol-related crashes. \textit{Id.} at 275 tbl. 3.117.

\textsuperscript{215} See MACCOUN & REUTER, supra note 14, at 351.
that policy makers may want harm reduction in the context of tobacco, and they certainly are willing to discuss many policy alternatives except criminalization of tobacco. The tendency to allow smokers to consume tobacco as long as others are not directly or indirectly harmed is in stark contrast to the treatment of marijuana smokers who make what are probably less damaging choices for themselves and others but are required to take total responsibility and suffer enforced treatment and incarceration. Once policy experimentation is allowed to flourish, careful evaluation should accompany reforms, and if the benefits of such reform prove to be as large as the preceding economic analysis suggests, reform of other illicit drug policies might be investigated. In some cases, of course, policy experimentation may reveal that the consequences of reform are unacceptable, at least in some states or communities. For example, witness the wide variety of state and local laws dealing with alcohol.

Local jurisdictions also have an important role in the search for effective drug policy because most law enforcement activity is done by local agencies and, importantly, drug problems vary markedly within states. Rural communities, affluent suburban areas, and distressed central city neighborhoods within a state are likely to face very different patterns of drug use because their populations are likely to perceive long-term opportunities very differently due to family circumstances and neighborhood environment. Therefore, many aspects of drug policy should be left up to local jurisdictions, and the creativity of their policy responses should not be stunted by perverse incentives imposed by state and federal law.

216. See Haaga & Reuter, supra note 204, at 38-46.
217. Relatively affluent suburban youth might use drugs for recreation, while youth in dysfunctional central cities and depressed rural areas might engage in self-medication to relieve the psychological stress caused by their environment and modest future prospects. Drug policy advocates may reject the claim that both reasons for drug use are understandable and legitimate but nevertheless recognize that the very different circumstances of communities may require alternative remedies to alleviate problem drug use.
218. Local jurisdictions could, of course, target their enforcement efforts on minorities and routinely violate civil rights. State and federal law obviously should restrain such local excess, but the problem with drug policy is that federal drug enforcement grants and asset forfeiture laws provide incentives that encourage the violation of citizens’ rights. Horror stories of local jurisdictions’ drug enforcement run amok are easy to find, but the particularly egregious case of Tulia, Texas has received national attention. In a town of 4500, forty-three “drug dealers” were arrested; all were black, except for three whites that had close ties with the black community. Due process was regularly violated and all the convictions were obtained on the unsubstantiated claims of an undercover officer of suspicious character. See Jim Yardley, The Heat Is on a Texas Town After the Arrests of 40 Blacks, N.Y. TIMES, Oct. 7, 2000, at A1. The Tulia drug task force was just one of forty-nine narcotics task forces in Texas that have received over $300 million in federal funds to conduct the war on drugs that between 1987 and 2000 made 189,586 arrests and seized over $184 million in assets. Henderson, supra note 155. By 2002, abuses led Governor Rick Perry to order “the Department of Public Safety begin monitoring the state’s forty-nine narcotics task forces following allegations that some of the drug teams were little more
Local governments obviously have the capacity for innovative drug policies. Drug courts were developed in Miami, Florida, and have spread throughout the nation, for example.\textsuperscript{219} Furthermore, local innovations are more likely to be incremental changes that do not receive much national attention. Consider, for instance, the common problem with excessive drug enforcement that drug crackdowns in one part of town simply result in dealers moving to another part of town.\textsuperscript{220} In Tampa, Florida, rising enforcement in black neighborhoods caused dealers to move into white neighborhoods that had not experienced an active drug trade.\textsuperscript{221} These and other failures of the drug war caused community leaders to pressure the police to change their tactics, to emphasize community safety rather than more easily measured outputs such as arrests. The resulting innovative program focused on community involvement, and police officials expected that “the most that would happen to the drug trade . . . would be that it moved indoors. That seemed, nonetheless, a worthy goal.”\textsuperscript{222} When local officials are held accountable for the costs of their policies, they are more likely to implement policies that are cost effective and are somewhat less prone to engage in the drug war rhetoric that is especially appealing when the costs of policy are borne by another level of government.

D. Removing Perverse Policy Incentives and Constraints

1. Asset Forfeiture

Even if control of policy is substantively devolved to states and local jurisdictions, states need to avoid imposing incentives that bias policy toward enforcement and constraints that potentially increase the social costs of drug enforcement. Reforming the perverse incentives generated by asset forfeiture legislation at both the state and federal level, which allow police to keep the proceeds and therefore increases drug enforcement activity beyond the level they would choose in the absence of this incentive, should be high on the policy...
reform agenda. If seizing the assets of drug offenders has a strong deterrent effect, although there is no empirical evidence to suggest that it does, and neglecting the obvious injustices that these in rem proceedings can engender, then a minimal reform would simply require that all forfeited assets be turned into general revenue rather than into accounts dedicated to the law enforcement agency. Such a straightforward reform would eliminate this important incentive to combat drugs at the expense of other crimes, the advantage apparently being more arrests for property offenses. A more thorough reform would authorize asset forfeiture only in those cases where the owner is convicted of a drug offense and the seizures do not make the punishment disproportionate to the crime.

2. Minimum Mandatory Sentences

Minimum mandatory sentences are a constraint that legislated drug policy places on the judicial system, often wreaking havoc with fairness in sentencing because judges are stripped of the capacity to calibrate a punishment that is proportionate to the crime. Aside from their tendency toward injustice, minimum mandatory sentences also tend to disproportionately increase drug enforcement because of institutional incentives that influence police and prosecutors. The output of prosecutors is convictions, and tough minimum mandatory sentences make plea bargaining easier and conviction more likely, thus providing prosecutors added impetus to conduct a war on drugs.

3. Unintended Consequences of Well-Intentioned Reforms: Mandated Treatment and Agency Incentives

Recognizing that incarcerating non-violent drug offenders is an expensive and ineffective way to combat substance abuse, there is increasing interest among some states and local governments in mandating treatment for first or second drug offenses. Although this reform is intended to reduce drug abuse and the social costs of drug enforcement, the details of implementation could actually reduce the efficacy of drug policy. In the most draconian scenario, offenders are mandated into a treatment regime and if they fail to complete it or recidivate, they are incarcerated for a longer time than they otherwise would have been. Mandating treatment also means that the demand for treatment will greatly increase relative to the supply of competent programs, putting problem drug users at further risk of

223. Supra text accompanying note 148.
224. Supra note 150.
225. For a scathing indictment of mandatory sentences, see United States v. Hiveley, 61 F.3d 1358, 1363-66 (8th Cir. 1995) (Bright, J., concurring).
226. Rydell & Everingham, supra note 125.
being found incorrigible when they recidivate. Finally, such a law is very likely to squander valuable treatment resources by imposing treatment on casual experimenters who have no discernable drug problem. While the move away from incarceration toward treatment is in principle desirable, success of such programs crucially depends on how they are integrated into, or separated from, the system of punishments. Further, they are fundamentally flawed if they are predicated on the proposition that all substance use is misuse and should be punished, or treated, by law.

In the context of a decentralized experimental environment, it may turn out that centralization of some drug policy issues will still be justified for some problems that are national in scope. For instance, the central government might have a comparative advantage as a potential disinterested third party in comparing the consequences of various local experiments and providing information about what appears to work. Similarly, research on the medical and psychological consequences of various types of drug use may be needed to determine what the most effective policies should be, and the federal government could be the source or at least the disseminator of such research. If criminal justice remains the major focus of drug policy and combating the importation of illegal drugs from abroad is a desirable part of that policy, then that also falls within the purview of the federal government. Under these circumstances, a case might also be made for federal involvement in combating interstate trade of illegal drugs, but it must be recognized that if the central government is in charge of controlling the flow of drugs into a state, they are in a position to influence local policy on use in ways that may be inappropriate.

VIII. CONCLUSION

Having battle plans for a drug war drawn up in Washington D.C. by a czar has led to policies that are about as effective as those usually imposed by pre-Soviet Russian emperors. Even without resorting to first principles that would question the right of the government to regulate the use of substances when other persons are not adversely affected, it seems clear that the U.S. drug war has generated more collateral damage than can conceivably be warranted by any realistic assessment of the associated benefits. Given the great variation in drug use among jurisdictions in this country, many decisions about drug policy should devolve from Washington to states and local communities, and funding of these programs should come from local resources. When state and especially local officials are responsible for

227. See Shedler & Block, supra note 44.
the financial and intangible costs of drug enforcement, they are more likely to be held accountable for their actions and therefore are more likely to carefully consider the full costs of a largely ineffective drug policy.
THE MEANING OF “APPROPRIATE VALIDATION” IN
DAUBERT V. MERRELL DOW PHARMACEUTICALS,
INC., INTERPRETED IN LIGHT OF THE BROADER
RATIONALIST TRADITION, NOT THE NARROW
SCIENTIFIC TRADITION

EDWARD J. IMWINKELRIED*

I. WHAT THE TRILOGY DOES NOT STAND FOR: SPURIOUS CLAIMS ABOUT THE
MEANING OF DAUBERT, JOINER, AND KUMHO.................................................... 740
A. The Language of the Opinions in the Trilogy.............................................. 740
B. Generally Held Epistemological Views........................................................ 742

II. HOW IT MUST BE VALIDATED: THE TRILOGY SHOULD NOT BE CONSTRUED AS
CONFINING THE PROPONENT TO EMPIRICAL TESTING AND INDUCTION AS THE
METHOD OF VALIDATION..................................................................................... 744
A. The Language of the Opinions in the Trilogy.............................................. 744
B. Generally Held Epistemological Views........................................................ 746

III. WHAT MUST BE VALIDATED: THE TRIAL JUDGE SHOULD NOT ONLY FOCUS
ON THE SPECIFIC CLAIM BEING MADE BY THE EXPERT; THE JUDGE SHOULD
ALSO DISTINGUISH BETWEEN PURELY DESCRIPTIVE AND INFERENTIAL
CLAIMS............................................................................................................... 749
A. The Required Foundation as a Variable of the Proponent’s Claim............ 749
B. The Varying Nature of Expert Claims........................................................... 751

IV. HOW MUST IT BE VALIDATED?: IN THE CASE OF INFERENTIAL CLAIMS, THE
TRIAL JUDGE SHOULD ASK WHETHER THE PROPONENT’S FOUNDATIONAL
SHOWING OF THE RESULTS OF USING THE EXPERT’S THEORY OR TECHNIQUE
WOULD CONVINCE A SKEPTICAL RATIONALIST THAT ITS USE ENABLES THE
EXPERT TO ACCURATELY MAKE THE SPECIFIC DETERMINATION TO WHICH THE
EXPERT PROPOSES TESTIFYING................................................................. 756
A. A Proposed Standard for Inferential Claims.............................................. 757
B. Illustrative Applications of the Proposed Standard................................. 760

V. CONCLUSION........................................................................................................ 764

[W]e have cast off the naive doctrine that all science is necessarily
true and that all true knowledge is necessarily scientific . . . .1

The Supreme Court has rendered a trilogy of cases explicating the
standards governing the admissibility of expert testimony under
Federal Rule of Evidence 702. In 1993, in the watershed case of
Daubert v. Merrell Dow Pharmaceuticals, Inc.,2 the Court initially
held that the traditional common law, general acceptance admissibil-
ity standard for scientific evidence is no longer good law. The Court
remarked that the ‘rigid ‘general acceptance’ requirement [was] at
odds with the ‘liberal thrust’ of the Federal Rules and their ‘general
approach of relaxing the traditional barriers to opinion testimony.’”3

* Visiting Professor of Law, University College Dublin; Professor of Law, University of
California at Davis; B.A. 1967, J.D. 1969, University of San Francisco; former Chair,
Evidence Section, American Association of Law Schools.

1. JOHN ZIMAN, RELIABLE KNOWLEDGE: AN EXPLORATION OF THE GROUNDS FOR
BELIEF IN SCIENCE 2 (1978) (emphasis omitted).
3. Id. at 588 (quoting Beech Aircraft Corp. v. Rainey, 488 U.S. 153, 169 (1988)).
The Court noted that Article VII of the Federal Rules, governing opinion testimony, omitted any language that could reasonably bear the construction that it codified a “general acceptance” requirement. In the Court’s words, “[g]iven the Rules’ permissive backdrop,” that omission had the effect of abolishing the traditional requirement.

Then the Daubert Court proceeded to derive a new standard from the statutory text of Rule 702. Justice Blackmun pointed out that the statute refers to “scientific . . . knowledge,” and opted for a methodological definition of “scientific knowledge.” He ruled that, to be sufficiently reliable to be admissible, a purportedly scientific proposition must be supported by “appropriate validation.” To provide guidance to trial judges, Justice Blackmun listed several factors that trial judges may weigh in evaluating the soundness of the underlying scientific methodology, including such considerations as whether the expert’s hypothesis has been tested; whether the theory has been subjected to peer review; “the known or potential rate of error”; and “the existence and maintenance of standards controlling the technique’s operation.” However, the Justice emphasized both that the list was not “definitive” and that “[t]he inquiry envisioned by Rule 702 is . . . a flexible one.”

In 1997, the Court revisited the topic of scientific testimony and handed down its decision in General Electric Co. v. Joiner. The formal holding in Joiner is that on appeal, the proper scope of review of a trial judge’s ruling is whether the judge abused his or her discretion. The Court took the occasion to amplify on the validation standard enunciated in Daubert. In the lead opinion in Joiner, Chief Justice Rehnquist asserted:

> [N]othing in either Daubert or the Federal Rules of Evidence requires a district court to admit opinion evidence that is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.

---

4. *Id.* at 589.
8. *Id.* at 593.
9. *Id.*
10. *Id.* at 594.
11. *Id.*
12. *Id.* at 593.
13. *Id.* at 594.
15. *Id.* at 143.
16. *Id.* at 146.
In 1999, the Court completed the trilogy by deciding *Kumho Tire Co. v. Carmichael*. While the proponents of the testimony in *Daubert* and *Joiner* characterized their evidence as “scientific” in nature, in *Kumho* the proponents argued that their testimony qualified as non-scientific expert evidence. As they correctly pointed out, Rule 702 refers in the alternative to “scientific, technical or other specialized knowledge.” The question posed was the degree, if any, to which *Daubert* extended to non-scientific expertise. On the one hand, Justice Breyer, writing for the majority, acknowledged that it would be unduly rigid to mandate that the trial judge assess the admissibility of non-scientific expertise solely in terms of the factors listed in *Daubert*. After all, that list had been devised with scientific methodology in mind. The species of non-scientific expertise are so variegated that the Court could “neither rule out, nor rule in, for all cases and for all time the applicability of the factors mentioned in *Daubert*. On the other hand, the Court forcefully rejected any suggestion that non-scientific expertise is wholly exempt from “Daubert-style scrutiny” of its reliability. Justice Breyer explained:

> In *Daubert*, the Court specified that it is the Rule’s word “knowledge,” not the words (like “scientific”) that modify that word, that “establishes a standard of evidentiary reliability.” Hence, as a matter of language, . . . Rule [702] applies its reliability standard to all “scientific,” “technical,” or “other specialized” matters within its scope.

The Court reiterated its position in *Joiner* that the trial judge should not accept *ipse dixit* from an expert. While *Joiner* recognized that the trial judge enjoys discretion in applying the *Daubert* factors to gauge the admissibility of scientific testimony, *Kumho* accorded the judge a second, deeper type of discretion—a latitude to select factors which strike the judge as “reasonable measures of the reliability” of the non-scientific expertise in question.

One would think that the Court’s rendition of three opinions in this doctrinal area in such a short period of time would have brought exceptional clarity to the area and settled many, if not most, of the controversies. Moreover, the scholarly commentary in this area has

---

18. FED. R. EVID. 702.
20. *Id.* at 150.
21. *Id.* at 158.
22. *Id.* at 147 (quoting *Daubert* v. Merrell Dow Pharms., Inc., 509 U.S. 579, 589-90 (1993)) (citation omitted).
23. *Id.* at 157.
24. *Id.* at 152.
25. *Id.* at 150-52.
been extensive and often insightful. However, despite the volume of commentary and the Court’s frequent pronouncements in this area, there are still large areas of uncertainties. The uncertainties relate to two basic issues: What must be validated, and how can it be validated?

On both issues, the opinions include passages that are susceptible to the interpretation that the Court has established onerous requirements for introducing expert testimony. For example, on the first issue, *Kumho* contains language implying that, at least in some cases, the trial judge should pass on the “global[ ]27 question of the reliability of the expert discipline itself. In his opinion in *Kumho*, Justice Breyer mentioned a category of cases in which “the discipline itself lacks reliability, as, for example, do theories grounded in any so-called generally accepted principles of astrology and necromancy.”

To compound the proponent’s burden, on the second issue—how it must be validated—the *Daubert* opinion implies that “experimental” validation is a “canonical” requirement for demonstrating the reliability of expert testimony. The majority approvingly quoted the late Sir Karl Popper’s assertion that “the criterion of the scientific status of a theory is its falsifiability, or refutability, or [empirical] testability.” It is no accident that at the very beginning of his list of relevant factors, Justice Blackmun observed that:

> Ordinarily, a key question to be answered in determining whether a theory or technique is [reliable] scientific knowledge . . . will be whether it can be (and has been) tested [empirically].

Considered together, the placement of that factor at the head of the list and the Justice’s stress on that factor indicate that that consideration may be more than a mere factor. The empirical testing of the proposition is arguably a full-fledged requirement for admissibility under *Daubert*.22

---


27. Risinger, supra note 26, at 773.


31. Id. at 593.

32. See sources cited supra note 29.
While the opinions in the trilogy are susceptible to the interpretation that they carry these two implications, this Article argues that both implications are unsound. Part I of the Article advances the negative contention that the trilogy of cases should not be construed as endorsing either implication.

Part II of the Article turns to the more difficult, positive task of identifying the proper interpretation of the trilogy on the two fundamental questions of “what” and “how.” Subpart II.A advocates the position taken recently by Professor D. Michael Risinger,\(^33\) that the thing that must be validated is the hypothesis that the expert can perform the specific “task at hand” involved—not, for example, the broad question of whether forensic document examination is a valid discipline, but instead the narrower issue of a document examiner’s claimed ability to identify the author of hand-printed Japanese characters.\(^34\) Rather than being required to demonstrate the validity of the field that the expert practices in, the proponent’s only obligation is to show that the theory or technique in question enables the expert to accurately make the determination that the expert proposes testifying to.

Subpart II.B then contends that it would be wrongminded for the courts to place exclusive or even primary emphasis on empirical testing as a validation method. The courts should be open to a variety of validation techniques, including, but not limited to, empirical induction and mathematical deduction. Adopting the attitude of a skeptical rationalist, the judge ought to inquire whether the results of the use of the technique in question demonstrate that the technique “works”; that is, whether the technique enables the expert to accurately perform the specific task at hand. As the task itself varies, as when the expert puts the technique to different uses and applications, the required validation must also change.

Thus, this Article conceives of “appropriate validation” as a flexible, relative concept. There is no invariable requirement that the proponent of expert testimony demonstrate that the expert’s theory has been validated by empirical testing and induction. Modernly, rationalist, skeptical decisionmakers commonly rely on other modes of validation. Those modes should also be acceptable in the courtroom. Further, there is a close relationship between the “how” and the “what” issues. The judge cannot intelligently decide which modes of validation are acceptable until the judge focuses on the precise claim made by the expert. The judge cannot determine which types of vali-

\(^{33}\) Risinger, \textit{supra} note 26.

\(^{34}\) \textit{Id.} at 798-800 (discussing United States v. Fujii, No. OO CR 17, slip op. at 3-4 (N.D. Ill. Sept. 25, 2000)). The official citation for \textit{Fujii} is 152 F. Supp. 2d 939 (N.D. Ill. 2000).
dation are apropos until the judge has identified the precise claim and task that must be validated.

I. WHAT THE TRILOGY DOES NOT STAND FOR: SPURIOUS CLAIMS ABOUT THE MEANING OF DAUBERT, JOINER, AND KUMHO

As previously stated, one passage in Kumho implies that the proponent has the burden of establishing the general reliability of the proffered expert's discipline. In his lead opinion, Justice Breyer commented that sometimes “the discipline itself lacks reliability.” The Justice cited two examples, namely, astrology and necromancy. However, the trilogy contains other language pointing to the more sensible conclusion that the proponent need demonstrate only that the expert's theory or technique can enable the expert to accurately make the specific determination which he or she proposes to testify about.

A. The Language of the Opinions in the Trilogy

As Professor Risinger has emphasized, in the formal summary at the end of his opinion in Daubert, Justice Blackmun stated that the proponent's foundation must convince the trial judge that the expert's theory or technique is sufficiently “reliable” to perform “the task at hand.” Earlier in the opinion, in the process of explaining the requirement that the theory or technique “fit” the case, the Justice added that the theory or technique needs “a valid scientific connection to the pertinent inquiry.”

Joiner lends itself to the same interpretation as Daubert. In Joiner, Chief Justice Rehnquist analyzed the question of whether the animal studies cited by the plaintiff were an adequate basis for the expert's opinion as to the cause of Joiner's small-cell lung cancer. The Chief Justice initially listed the criticisms of the animal studies. The Chief Justice then wrote:

35. See supra notes 21 & 22.
37. Id.
39. Risinger, supra note 26, at 772.
41. Id. at 591-92.
42. Id. at 592.
44. Id.
Respondent [plaintiff] failed to reply to this criticism. Rather than explaining how and why the experts could have extrapolated their opinions from these seemingly far-removed animal studies, respondent chose “to proceed as if the only issue [was] whether animal studies can ever be a proper foundation for an expert’s opinion.” 864 F. Supp., at 1324. Of course, whether animal studies can ever be a proper foundation for an expert’s opinion was not the issue. The issue was whether these experts’ opinions were sufficiently supported by the animal studies on which they purported to rely.45

Significantly, the emphasis indicated by the italics appeared in the Chief Justice’s original opinion.

The Kumho opinion is cast in the same mold.46 To begin with, Kumho echoes some of the key language in Daubert. For instance, the Kumho Court refers to “the task at hand”47 and the need for a demonstrated connection between the expert’s theory and “the pertinent inquiry.”48 More importantly, though, in reviewing the foundation laid by the plaintiffs for Carlson’s expert opinion, Justice Breyer engaged in a highly particularized analysis:

[C]ontrary to [plaintiffs’] suggestion, the specific issue before the [trial] court was not the reasonableness in general of a tire expert’s use of a visual and tactile inspection to determine whether overdeflection had caused the tire’s tread to separate from its steel-belted carcass. Rather, it was the reasonableness of using such an approach, along with Carlson’s particular method of analyzing the data thereby obtained, to draw a conclusion regarding the particular matter to which the expert testimony was directly relevant.49

Again, the emphasis is in Justice Breyer’s own text. The Justice acknowledged that “as a general matter, tire abuse may often be identified . . . through visual or tactile inspection of the tire.”50 However, Carlson claimed to have developed a more “particular” method—namely, a theory that there are four characteristic signs of tire abuse and that the absence of at least two of the signs indicates that the accident was caused by a manufacturing defect in the tire.51 A bit later in the opinion, the Justice emphasized that Carlson had not rested his opinion simply on the general theory that, in the absence of evidence of abuse, a defect will normally have caused a tire’s separation. Rather, the expert employed a more specific theory to establish the

47. Id. at 141.
48. Id. at 149.
49. Id. at 153-54.
50. Id. at 156.
51. Id. at 154.
existence (or absence) of such abuse.52

Still later, Justice Breyer underscored that “the question before the trial court was specific, not general.”53 In the next paragraph, the Justice stated that “[t]he particular issue in this case concerned” the reliability of the specific theory that Carlson had employed.54

B. Generally Held Epistemological Views

Restricting the validation requirement to the specific theory or technique the expert relies on is not only consistent with most of the language in the opinions forming the trilogy, but is also in accord with epistemological views generally held in our society. Any other interpretation of the trilogy would lead to the admission of junk expertise as well as to the exclusion of demonstrably reliable evidence.

Assume, for example, that the expert belongs to an established field of science in which there is a huge body of literature documenting quality experimentation validating many of the propositions circulating in the field. Yet, at any given time, the discourse in the field will undoubtedly include a spectrum of types of propositions. In some cases, the propositions have such substantial supporting data that we can be relatively confident that we “know” the proposition to be true, at least as a working assumption in everyday life and industry. However, the discourse is also likely to include unsubstantiated conjectures, and worse still, speculations that will later be exposed as invalid.55 At one extreme, it would be premature to permit testimony about any theory circulating in the field simply because many, if not most, of the propositions being discussed in the field’s discourse have passed the muster of empirical validation. Inferring the truth of one proposition in the field from the truth of another proposition in the same field can be a non sequitur. Consider, for example, the discipline of forensic pathology. It is true that the courts routinely accept pathologists’ testimony on a wide range of subjects.56 Judicial receptivity to pathologists’ opinions on many subjects, such as the estimation of stature from skeletal remains, is justifiable, since there is a substantial body of research investigating the reliability of those estimations.57 However, the literature in the pathology field also includes discussions of many novel conjectures, especially with respect to the determination of time of death.58 Despite the respected status

52. Id.
53. Id. at 156.
54. Id. at 157.
55. ZIMAN, supra note 1, at 130-33.
56. 2 PAUL C. GIANNELLI & EDWARD J. IMWINKELRIED, SCIENTIFIC EVIDENCE § 19-10(B) (3d ed. 1999).
57. Id. § 19-4(B), at 103.
58. Id. § 19-8(A).
of the field of forensic pathology, it would be fallacious to leap to the conclusion that pathologists should be allowed to testify about any theory that has garnered a measure of attention in the field. By the same token, even if one posits the general reliability of the field of questioned document examination, that assumption does not dictate the admissibility of proffered testimony identifying the author of handwritten Japanese characters.59

At the other extreme, it is equally foolish to bar all testimony even about such maligned fields as astrology—one of the areas of claimed expertise that Justice Breyer cited in Kumho as a classic example of a discipline that “itself lacks reliability.”60 Suppose that a testator left a substantial bequest to an institute for the express purpose of “fostering the study of astrology.” The law of wills allows testamentary gifts for the purpose of promoting any lawful activity.61 Although most Americans may have no faith in astrological predictions,62 the study of astrology is perfectly lawful. Assume further that at some point after the testator’s death, the heirs bring a challenge, alleging that the institute is no longer using the bequeathed funds for the purpose specified in the decedent’s will. In addition to hearing testimony about the manner in which the institute was spending the funds, the court could undoubtedly accept testimony from experienced astrologers as to whether the funded activities related to “astrology,” as that term is generally understood by its avowed practitioners. To adjudicate the dispute, there is no need for the court to make a global judgment about the scientific reliability of astrologers’ predictions. Rather, the “task at hand” is deciding whether the proffered witness is familiar enough with the state of the discipline to determine whether the funded activities in any way relate to the discipline.

The upshot is that despite Justice Breyer’s disparaging comments about astrology and necromancy in Kumho, it is neither necessary nor sufficient for a judge passing on the admissibility of an expert’s testimony to make a global judgment about the general reliability of the expert’s discipline. Even if the field has amassed a huge body of research verifying many of the propositions relied on by experts in the field, the proponent should be required to lay a foundation demonstrating the reliability of the specific technique the expert proposes to use to perform the “task at hand.” Conversely, even when the field

59. Risinger, supra note 26, at 798-800 (discussing United States v. Fujii, 152 F. Supp. 2d 939 (N.D. Ill. 2000)).


61. 79 A.M.JUR. 2D Wills § 65 (1975); 95 C.J.S. Wills § 72 (2001).

62. ROPER CENTER FOR PUBLIC OPINION RESEARCH, Question ID: USGALLUP:96SEP3 RO4H 1999 (stating that only 25% of the respondents expressed belief in astrology) (on file with author).
has little or no research to validate its leading tenets, the judge cannot substitute bias for an analysis of the question of whether the proponent has laid a foundation demonstrating the reliability of the expert’s use of a specific theory or technique to make a particular determination. In short, the judge’s focus ought to be narrow and precise when the judge defines what must be validated.

II. HOW IT MUST BE VALIDATED: THE TRILOGY SHOULD NOT BE CONSTRUED AS CONFINING THE PROPONENT TO EMPIRICAL TESTING AND INDUCTION AS THE METHOD OF VALIDATION

As in the case of the question of what must be validated, the trilogy contains passages implying that the proponent has a heavy burden on the second question. The passages suggest that the required—or at least preferred—method of validation is empirical testing. Justice Blackmun began his list of relevant factors by highlighting the “key question . . . [of] . . . whether it can be (and has been) tested.” In the course of explaining that factor, the Justice cited several authorities indicating that he meant testing in the sense of controlled, empirical experiments. One citation was to Hempel, asserting that the “statements constituting a scientific explanation must be capable of empirical test.” The very next citation is to Popper’s declaration that the “criterion of the scientific status of a theory is its falsifiability, or refutability, or testability.” In his separate opinion, the Chief Justice made it clear that he assumed that Justice Blackmun was referring to “empirical testing.”

A. The Language of the Opinions in the Trilogy

However, the parallel to the trilogy’s treatment of what must be validated continues. Once again, the trilogy contains other language making it reasonably clear that the Court is not mandating the use of controlled empirical testing as the sole or even primary method of validating expert testimony.

Rather than naively lauding the scientific technique of empirical testing, the Daubert Court acknowledged the limits of the scientific enterprise. The Court rejected the myth of the infallibility of science

63. Crump, supra note 29; Jasanoff, supra note 29, at 9, 11-12.
65. Id. (citing Carl G. Hempel, Philosophy of Natural Science 49 (Elizabeth & Monroe Beardsley eds., 1966)).
66. Id. (citing Karl Popper, Conjectures and Refutations: The Growth of Scientific Knowledge 37 (5th ed. 1989)).
67. Id. at 600 (Rehnquist, C.J., concurring in part and dissenting in part).
and asserted that “arguably, there are no certainties in science.”

Empirical induction cannot yield absolute certainty; even when a large number of experiments yield observations consistent with the truth of an hypothesis, the hypothesis can be accepted only provisionally because there are always other conceivable experiments that could be devised. The possibility of another experiment raises the possibility of subsequent falsification. It would be anomalous if the Court demanded reliance on empirical induction in the same opinion in which it frankly confronted the limits of that methodology. That demand would also be inconsistent with Justice Blackmun’s caution that “[t]he inquiry envisioned by Rule 702 is, we emphasize, a flexible one.” The Court insisted on “appropriate validation,” not controlled experimental validation.

The Kumho opinion undercuts any contention that empirical testing and induction are the mandatory or preferred means of validating expert theories and techniques. Near the beginning of his opinion, Justice Breyer echoes Daubert by reading it as holding that “the test of reliability is ‘flexible.’” In a later part of the opinion, the Justice appeared to concur with the Solicitor General when stating:

As the Solicitor General points out, there are many different kinds of experts, and many different kinds of expertise. See Brief for United States as Amicus Curiae 18-19, and n. 5 (citing cases involving experts in drug terms, handwriting analysis, criminal modus operandi, land valuation, agricultural practices, railroad procedures, attorney’s fee valuation, and others).

Although some of the propositions in these disciplines, such as handwriting analysis, can certainly be tested by systematic scientific experiments, other propositions related to such topics as drug trafficking argot, criminal modus operandi, and attorney fee valuation

69. Daubert, 509 U.S. at 590.
71. Daubert, 509 U.S. at 594.
72. Id. at 590.
74. Id. at 150.
75. SCIENCE AND THE LAW, supra note 26, § 4-1.2; see also United States v. White Horse, 177 F. Supp. 2d 973 (D.S.D. 2001) (discussing the inadequate testing of parts of the Abel Assessment for Sexual Interest).
do not lend themselves to that mode of validation. As the Advisory Committee’s Note accompanying the December 1, 2002 amendment to Federal Rule of Evidence 702 observes, “[s]ome types of expert testimony will not rely on anything like a scientific method . . . .” Active drug dealers are highly unlikely to knowingly participate in controlled studies of their jargon or modus operandi, and researchers conducting any such studies could run the risk of criminal responsibility for misprision of felony.

**B. Generally Held Epistemological Views**

This broad view of the modes of permissible validation is not only consistent with most of the language in the opinions forming the trilogy—as in the case of narrowing the scope of the validation requirement to the specific theory or technique the expert relies on—but this view is also in accord with widely held epistemological notions. To be sure, in some cases empirical testing and induction are an appropriate and adequate means of validation. Suppose that a DNA typing laboratory is attempting to determine how much difference can be expected in measurements of the length of different DNA sample fragments from the same source—the so-called “match window.” That question cannot be answered *a priori* or by deduction. Rather, to validate the match window for its equipment and procedures, the laboratory must engage in empirical testing and rely on Baconian, inductive reasoning.

However, that type of reasoning is only one of the recognized branches of logic. To determine the sum of the degrees in all the corners of a square, a person could resort to empirical testing and, after measuring the degrees in a large number of squares, induce the answer. Alternatively, though, the person could turn to geometric deduction to find the answer. In this situation, mathematical deduction is not only a viable option, but in some respects, it is superior to empirical testing. Standing alone, the latter cannot yield certainty. Since there is always the possibility of invalidation in a subsequent test, induction can yield only probability. In contrast, if a proposi-
tion is deduced by strict mathematic logic, the proposition may be regarded as demonstrated or proven.

In other cases, especially in the medical field, extensive, collective clinical experience can suffice to validate a proposition even when the experience cannot be precisely quantified. If tens of physicians give the same course of treatment to hundreds of patients suffering from similar symptoms and then observe that the symptoms disappear, there is a reasonable inference that the treatment is an effective cure for the illness evidenced by that constellation of symptoms. Concededly, unlike some of the phenomena studied in fields such as physics, the behavior of patients cannot be reduced to “a precise logico-mathematical language.” In clinical experience with patients, variables cannot be controlled to the same extent as in a classic chemistry experiment. Yet, as Popper himself noted, the inductive empirical method is essentially “commonsense writ large.” Common sense strongly suggests that extensive clinical experience involving a wide range of patients can produce a reliable generalization as to the efficacy of a treatment.

Furthermore, widespread, collective experience can be an epistemologically sound basis for validating an expert proposition, even when the experience is compiled by persons lacking formal scientific or medical training. The examples abound. In *Kumho*, Justice Breyer appeared to approve of the Solicitor General’s position that testimony by undercover police officers about the meaning of drug terms is sufficient non-scientific expertise. The Advisory Committee’s Note to the recent amendment to Rule 702 also indicates that the courts should be receptive to that type of non-scientific expert testimony. Testimony by automobile mechanics provides a further ex-

---

86. ALBUREY CASTELL, AN INTRODUCTION TO MODERN PHILOSOPHY 197 (2d ed. 1963) (discussing the work of David Hume); see also Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 590 (1993) (“[A]rguably, there are no certainties in science.”).
87. ZIMAN, supra note 1, at 150, 163.
88. CASTELL, supra note 86, at 176 (discussing the epistemology of John Locke).
89. Kevin Patterson, What Doctors Don’t Know (Almost Everything), N.Y. TIMES MAG., May 5, 2002, at 74 (discussing the reliance on clinical experience in developing the medical state of the art).
90. Before the American Psychiatric Association released the latest version of its diagnostic criteria in the current version of *The Diagnostic and Statistical Manual of Mental Disorders*, the A.P.A. subjected many of the criteria to field trials involving more than 7,000 subjects at eighty-eight universities and research institutions. William D. Weitzel, Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV), ADVOCATE, AUG. 1994, at 25-26.
91. ZIMAN, supra note 1, at 160.
92. Id. at 166-68.
93. Id. at 135 (quoting KARL POPPER, THE LOGIC OF SCIENTIFIC DISCOVERY 22 (1959) (emphasis omitted)).
95. FED. R. EVID. 702 advisory committee’s note:
ample. The Note to the amendment to Rule 702 asserts that the judge may weigh the consideration that “the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give.”96 For decades, we have relied on automobile mechanics to diagnose engine malfunctions.97 Mechanics have rendered diagnostic and repair services that the public has found useful—in innumerable instances, mechanics have “fixed” the problems that prompted the customer to bring the car to the mechanic. There is a common sense inference from this extensive, successful experience, even though most of both the mechanics diagnosing and the lay customers relying on the diagnoses lack engineering degrees.

In summary, the cases forming the trilogy do not require that the trial judge demand that the proponent use empirical testing and induction as the sole, or even primary, mode of validation. Any such requirement would fly in the face of the general social consensus that there are modes of verifying propositions other than controlled scientific experimentation. However, having said that, the challenge remains: developing an affirmative understanding of the meaning of “appropriate validation”98 for the proponent’s claim that the expert can perform “the task at hand,”99 the specific determination that the expert claims to be able to make. That challenge leads us to Part II. We recur to the same two questions we began with: What must be validated, and how can it be validated?

When a law enforcement agent testifies regarding the use of code words in a drug transaction, the principle used by the agent is that participants in such transactions regularly use code words to conceal the nature of their activities. The method used by the agent is the application of extensive experience to analyze the meaning of the conversations. So long as the principles and methods are reliable and applied reliably to the facts of the case, this type of testimony should be admitted.

See also Mark Hansen, Dr. Cop on the Stand, 88 A.B.A. J., May 2000, at 30, 34 (Professor David Faigman expressing disappointment at the courts’ willingness to routinely permit this type of testimony by “police officers, who have little research or data to support their opinions.” Professor Faigman “call[s] for a more discriminating analysis of the scientific methods and techniques underlying such testimony. At the very least, he says, the courts should inquire into the nature and extent of a police officer’s asserted expertise.”). In this regard, Professor Faigman is certainly correct. The trial judge should not presume that every police officer has the requisite experience. For that matter, it should not suffice for the officer to simply assert that he or she has experience as an undercover agent in illegal drug transactions. Rather, the officer should be required to establish that in several such drug transactions, he or she has encountered the use of the same or similar code words. The question is the witness’s possession of significant, similar experience, not the witness’s status as an undercover police officer.

96. Fed. R. Evid. 702 advisory committee’s note.
99. Id. at 597.
In order to intelligently answer the question of “appropriate validation” in sequence, the judge must pose the two questions of what and how.

III. WHAT MUST BE VALIDATED: THE TRIAL JUDGE SHOULD NOT ONLY FOCUS ON THE SPECIFIC CLAIM BEING MADE BY THE EXPERT; THE JUDGE SHOULD ALSO DISTINGUISH BETWEEN PURELY DESCRIPTIVE AND INFERENCEAL CLAIMS

Professor Risinger has helped clarify this area of law with his proposal that the trilogy be interpreted as requiring the judge to focus on the specific theory or technique the expert is relying on, rather than a global judgment about the reliability of the expert’s field or discipline.100 However, one further refinement is necessary: the courts must also identify the particular use to which the expert proposes to put the theory or technique. What is the specific determination that the expert claims that use of the theory or technique will enable the expert to make?

A. The Required Foundation as a Variable of the Proponent’s Claim

In many areas of evidence law other than expert opinion, the courts have long recognized that the required foundation depends not only on what item of evidence the proponent is offering, but also on the purpose for which the item is proffered.

The authentication area is illustrative of this. An accused is charged with murder but claims self-defense. At trial, the defense attorney contemplates introducing the accused’s testimony that just before his encounter with the alleged victim, the accused received a threatening letter purportedly written by the victim.101 Suppose that the defense attorney offered the testimony for the limited purpose of showing that the accused had an honest fear of the victim. On that theory of logical relevance, it would suffice for the accused to identify the letter as one he received. The accused would not have to establish that the alleged victim actually wrote the letter. The purpose of the evidence is to show the accused’s state of mind.102 Even if the letter is a forgery and the accused’s belief about its authorship is mistaken, the accused’s receipt of the letter is relevant to substantiate the subjective element of his self-defense claim.

However, assume that the defense attorney wants to put the same evidence to another use. Now the defense counsel wants to treat the letter as evidence that the alleged victim hated the accused and was

100. Risinger, supra note 26.
102. Id.
more likely to have started the fight. Although the defense is still offering the same exhibit, the defense would have to lay a quite different foundation. The defense will have to establish the victim’s authorship of the letter.103 “The prior foundation is adequate relative to the claim about the defendant’s state of mind, but insufficient relative to the claim about the alleged victim’s state of mind and conduct.”104

In addition, consider the hearsay area. Assume, for example, that the plaintiff has filed a personal injury action against the defendant. The plaintiff pedestrian alleges that she was injured on March 1, 2002, when the brakes on the defendant’s car failed, and the car struck the plaintiff. The plaintiff’s complaint seeks both compensatory and punitive damages. As the basis for the punitive damages claim, the plaintiff alleges that the defendant was driving recklessly in conscious disregard of the risk that his brakes would fail. The plaintiff avers that on February 1, 2002, the defendant learned that his brakes were defective.

At trial, the plaintiff offers an exhibit into evidence. The exhibit purports to be a report given to the defendant by an automobile mechanic. The report, dated February 1, 2002, explicitly states that the brakes on the defendant’s car are “dangerously thin and in need of immediate replacement.” The plaintiff could offer the exhibit on her punitive damages claim. If the plaintiff proffers this exhibit for the limited purpose of establishing the defendant’s notice of the existence of the dangerous condition, the exhibit would be nonhearsay.105 The plaintiff would be offering the exhibit only to show the effect of the contents of the exhibit on the state of mind of the reader; namely, putting the defendant on notice of the brake problem.106 If so, the only necessary foundation would be proof that the exhibit is a document that the mechanic handed the defendant. Alternatively, though, the plaintiff might offer the exhibit on her compensatory damage claim. If she did so, she would have to establish the truth of the report’s assertion that the defendant’s brakes were defective. On that theory, the exhibit constitutes hearsay,107 and the plaintiff would have to lay a very different foundation to bring the exhibit within the business entry hearsay exception.108 In short, the foundation depends both on what item of evidence the proponent offers and on which theory of logical relevance the proponent relies.

103. Id.
104. Id. at A18.
105. FED. R. EVID. 801(c).
107. FED. R. EVID. 801(c).
108. IMWINKELRIED, supra note 106, § 10.05.
The same holds true in the character evidence area. Change the facts in the prior hypothetical. Assume that rather than suing the driver, the plaintiff sues the driver’s employer. Once again, her complaint seeks both compensatory and punitive damages. She alleges that she was injured as a result of the employee’s careless driving in the scope of his employment. Moreover, to support her punitive damage claim, she alleges negligent entrustment. She avers that when the defendant hired the employee, the defendant knew or should have known that the employee is an habitually careless driver.

At trial, during her case-in-chief, the plaintiff calls the driver’s former employer. The plaintiff seeks to elicit the former employer’s testimony about a number of serious traffic accidents caused by the driver while he was in this witness’s employ. If the plaintiff offers this testimony solely to show that the driver is a careless driver, the only required foundation will be the witness’s personal knowledge that the driver caused the prior accidents.\(^\text{109}\) Given the negligent entrustment claim, the employee’s character trait for careless driving is in issue,\(^\text{110}\) and the employee’s prior, specific acts of careless driving are admissible to prove that character trait.\(^\text{111}\)

However, assume that the plaintiff wants to offer the identical testimony on her punitive damage claim in order to prove that when the defendant hired the employee, the defendant knew of the employee’s propensity. If the plaintiff offers the testimony for that purpose, an additional foundation is necessary; that is, proof that the employee’s former employer communicated the information to the defendant before the defendant hired the employee. Again, the nature of the required foundation turns on which theory of logical relevance the proponent advances, as well as on the content of the proffered testimony.

**B. The Varying Nature of Expert Claims**

As in the case of the authentication, hearsay, and character rules, the requisite foundation for expert testimony ought to vary with the purpose for which the evidence is being offered.\(^\text{112}\) In the past, there

\(^{109}\) **FED. R. EVID. 602.**

\(^{110}\) **Id. at 405(b).**

\(^{111}\) James W. McElhaney, *Don’t Be Locked Out: The Right Strategy Can Open Doors to Evidence That Might Otherwise Be Inadmissible at Trial*, 85 A.B.A. J., May 1999, at 64:

Sue a trucking company for its driver’s negligent injury of a pedestrian, and the driver’s bad record for speeding and reckless driving is not admissible in evidence. But change your pleadings to add a count of negligent entrustment of a truck to someone the company should have known was a terrible driver, and the driver’s record is admissible.

\(^{112}\) Edward J. Imwinkelried, *The Escape Hatches From Frye and Daubert: Sometimes You Don’t Need to Lay Either Foundation In Order to Introduce Expert Testimony!*, 23 AM.
has been an unfortunate tendency to lump together the various uses of expert testimony:

courts and commentators have tended to refer to a “Daubert foundation” as though one size predicate fit all. However, the use of the proposed expert testimony at trial determines the proponent’s claim about the expertise. In turn, the claim determines the required validation.  

The expert’s potential claims about his or her theory or technique fall into at least two broad categories: descriptive and inferential claims. In some cases, the expert merely describes or summarizes experience within his or her field. Suppose, for instance, that in a contract lawsuit, there is a dispute over the meaning of a term in the written agreement. To support her interpretation of the term, the plaintiff calls an experienced member of the industry as an expert witness. The witness proposes to testify that within the industry, there is a trade custom or usage as to the meaning of that term. The expert’s specific theory is that the usage exists within the industry. So long as the witness testifies that he or she has been a member of the industry for a certain period of time and has encountered that usage of the term on several occasions by industry members, the foundation ought to be deemed adequate. Standing alone, that experience suffices.

The same rationale explains the approving mentions of police testimony about drug argot in both Justice Breyer’s Kumho opinion and the 2000 Advisory Committee’s Note to Federal Rule 702. That
testimony is a variation of the evidence described in the preceding paragraph. Drug trafficking is a business. Just as a term can acquire a specialized meaning for members of a lawful commercial trade, a term can take on a peculiar significance for criminal drug traffickers. Hence, just as a veteran member of the meat scrap industry could testify as to the meaning of “50% protein” in a lawful contract between two industry members, an experienced undercover officer may testify as to the meaning of “lid” in an unlawful agreement for the purchase of a contraband drug. As the Advisory Committee’s Note states, in this situation “experience alone . . . may . . . provide a sufficient foundation for expert testimony.”

However, in other cases the proponent of the expert testimony wants the witness to do far more than merely recite or summarize experience as to fact A. Instead, the proponent contemplates inviting the expert to draw an inference from the witness’s experience. The expert evaluates the experience and draws a further inference as to fact B. Consider the possible uses, for example, of testimony by a psychologist about rape trauma syndrome (RTS).

The proponent could conceivably use such testimony for purely descriptive purposes. Assume, for example, that a medical board’s certification examination included a number of questions about RTS. The board gave a particular applicant a failing grade on the examination and denied him certification. The applicant later filed suit to challenge the examination. In particular, he disputed the accuracy of the board’s proposed answer to a question about whether a certain phobia is commonly regarded as a symptom of RTS. At the trial of that lawsuit, the applicant should be permitted to use the psychologist’s testimony for descriptive purposes. The psychologist might testify: He has been a practicing psychologist for twenty years; during that time, he has closely followed the RTS literature; and that literature generally indicates that the phobia in question is not symptomatic of RTS. The witness is merely relating or summarizing his experience within the field. The witness’s specific theory is that there is
a certain belief, custom, or practice in a field, and the witness’s experience validates that hypothesis.124

However, the proponent could put RTS testimony to radically different uses. Suppose, for instance, that the proponent wants to use the testimony as credibility evidence.125 The plaintiff files a premises liability action against a hotel. The plaintiff alleges that while staying in one of the defendant’s rooms, the plaintiff was raped due to lax hotel security. However, the plaintiff did not report the alleged offense to the police until seventy-two hours after the alleged incident. On cross-examination, the defense attorney forces the plaintiff to concede the delay in reporting the claimed rape; the suggestion is that the delay is pretrial conduct inconsistent with her trial testimony of a rape.126 Later in the plaintiff’s case, the plaintiff calls the psychologist to rehabilitate the plaintiff’s credibility. The plaintiff’s attorney wants to elicit the witness’s testimony that because of the social stigma attached to sexual assault, many rape victims delay reporting the offense.

Is the prior foundation still adequate for the further credibility inference? The answer is no. The proponent no longer is offering the testimony for purely descriptive purposes. The question is not whether the RTS profile exists within psychological circles. Rather, the proponent wants the witness to draw a credibility inference as to a connection between certain conduct (the delay) and a state of mind (the person’s honest belief that she has been raped). The prior foundation does not validate the existence of the connection. The expert’s specific theory is that if an alleged victim delays reporting a rape but otherwise matches the RTS profile, the match strengthens the infer-

124. Notice that the witness has testified to more than his or her membership in the discipline or field. The witness has added that he has had extensive experience reviewing the RTS literature. Without more, a witness’s testimony that he or she is a member of the pertinent field would not suffice. Even if he or she is a practitioner of a certain discipline, he or she may not have had any encounters with the relevant belief, custom, or practice. The key question is whether the witness has had a significant number of experiences similar to the experience involved in the litigation. See Edward J. Imwinkelried, The Next Step After Daubert: Developing a Similarly Epistemological Approach to Ensuring the Reliability of Nonscientific Expert Testimony, 15 CARDOZO L. REV. 2271, 2290-94 (1994) (discussing the quantitative and qualitative dimensions of the witness’s experience).

There is a similar doctrine with respect to reputation character evidence. If a witness proposes testifying about a person’s reputation for a character trait, it is not enough that the witness is a member of the same community as the person. In addition, the witness must vouch that he or she is familiar with the reputation. IMWINKELRIED, supra note 106, § 6.02[2]. Standing alone, membership in the person’s community does not guarantee that the witness has heard any discussions or mention of the person’s character.

125. 1 GIANNELLI & IMWINKELRIED, supra note 56, § 9-4(B).

126. 1 EDWARD J. IMWINKELRIED ET AL., COURTROOM CRIMINAL EVIDENCE § 712, at 267 (3d ed., Lexis Law Pub’g 1998) (“Like prior inconsistent statements, prior inconsistent acts are admissible to impeach.”).
ence\textsuperscript{127} that the victim honestly believes she had been raped. It no longer suffices to show that many or most psychologists subscribe to the theory that the RTS profile includes certain conduct.

Suppose, however, that the witness added the following testimony:

\begin{quote}
[I]n his experience and that of other rape counselors, many self-described rape victims delay reporting the offense [for the stated reason of] a sense of embarrassment. Furthermore, when these women were treated as rape victims, in most cases their mental health improved.\textsuperscript{128}
\end{quote}

It is true that the inference from this collective experience is not indisputable; some of the self-described rape victims might have been lying.\textsuperscript{129} However, this successful clinical experience with the therapy\textsuperscript{130} is highly probative: “[I]f there are numerous reports and an evident improvement in the reporter’s mental health in many cases, there is a plausible inference that most reporters were subjectively truthful.”\textsuperscript{131} Given this foundation, the testimony is rehabilitative; it demonstrates that even many women who subjectively believe that they have been raped engage in the seemingly impeaching conduct of delayed reporting.

Change the facts again. Assume that in the same premises liability case, the plaintiff’s attorney calls the psychologist to provide substantive corroboration that there was a rape. The attorney wants to argue that the fact that the plaintiff matches the profile is evidence on the merits that a rape occurred.

Once again, the original foundation would be inadequate. As in the case of the expert drawing the credibility inference, this witness proposes to do more than describe his experience that practitioners in his specialty accept RTS. Again, the expert’s underlying theory is not merely that a certain belief, custom, or practice—the RTS profile—exists within the specialty. This expert contemplates drawing a substantive inference as to a connection between certain symptoms and the prior occurrence of a particular type of historical event, that is, a rape. In this variation, the expert’s specific theory is that if the

\textsuperscript{127} The expert is not claiming that without more, the match either supports an inference of truthfulness or dictates such an inference. Testimony of the former type would rest on the specific theory that if the alleged victim delayed reporting the rape but otherwise matched the profile, the match is adequate to prove that the victim honestly believes that she was raped. Testimony of the latter sort would be based on the theory that if the alleged victim delayed reporting but otherwise matched the profile, the match certainly proved that the witness honestly believed that she had been raped. Since the underlying theories differ, different—in these cases, additional—validation would be required.

\textsuperscript{128} Imwinkelried, \textit{supra} note 101, at A18.

\textsuperscript{129} Id.

\textsuperscript{130} State v. Saldana, 324 N.W.2d 227, 230 (Minn. 1982).

\textsuperscript{131} Imwinkelried, \textit{supra} note 101, at A18.
alleged victim delayed reporting but otherwise matched the RTS profile, the match strengthens the inference that there was a rape. The original foundation falls even shorter here.

Even the foundation, embellished with testimony about numerous successful clinical interventions based on the RTS diagnosis, would be inadequate. Together with evidence of such clinical success, a large database consisting of self-reports of claimed rape victims can support the credibility inference. However, to establish that a syndrome extracted from the database is an accurate profile of rape victims, there must be some showing of the truth of the reports included in the database—emergency room reports that the reporters displayed physical signs of violent sexual assault, police reports that post-report investigations led to confessions of sexual assault, or perhaps convictions on rape charges prompted by the reports. The clinical successes may justify the credibility inference; but in the clinical mental health context, treatment often depends on the patient’s beliefs “even though [the patient’s subjective] beliefs about her illness [or the precipitating events] are entirely erroneous.”132 Here, though, the plaintiff’s attorney wants to use the profile substantively as “a fact-finding tool.”133 The attorney should be permitted to do so only if the expert can demonstrate that the reports used to generate the RTS profile are in fact based on rapes. Relative to this substantive claim, even the embellished foundation is lacking.

IV. HOW MUST IT BE VALIDATED?: IN THE CASE OF INFERENTIAL CLAIMS, THE TRIAL JUDGE SHOULD ASK WHETHER THE PROponent’S FOUNDATIONAL SHOWING OF THE RESULTS OF USING THE EXPERT’S THEORY OR TECHNIQUE WOULD CONVINCE A SKEPTICAL RATIONALIST THAT ITS USE ENABLES THE EXPERT TO ACCURATELY MAKE THE SPECIFIC DETERMINATION TO WHICH THE EXPERT PROPOSES TESTIFYING

As we have seen, the validation of purely descriptive expert claims is a relatively straightforward matter. As the 2002 Advisory Committee’s Note points out, “experience alone” can be adequate validation for this type of claim.134 If the witness has had a large135 number of similar136 experiences as to fact A, the witness ought to be permitted to describe or summarize that body of experience for the jury.

133. Saldana, 324 N.W.2d at 230.
134. Fed. R. Evid. 702 advisory committee’s note.
135. Imwinkelried, supra note 124, at 2290-92 (discussing Hume’s insistence upon “a repetition of similar” experiences).
136. Id. at 2292-94.
However, as Section III.A indicated, the foundational hurdle is higher when the expert wants to do more than vouch for the existence of a belief, custom, or practice within the discipline or field. For example, when a psychologist proposes relying on the RTS profile as the basis for a further inference as to credibility or historical events, a different foundation is necessary. How should a trial judge assess the adequacy of such a foundation under the trilogy?

A. A Proposed Standard for Inferential Claims

It is submitted that to assess the adequacy of the foundation, the trial judge should demand a showing of the results of the use of the expert’s theory or technique and then inquire whether a skeptical rationalist would consider that showing sufficient to demonstrate that the theory or technique “works,” namely, that it enables the expert to accurately make the inference the expert proposes to draw on the witness stand.

To begin with, the judge should adopt the stance of a skeptic. By that, I do not mean a total skeptic who believes as a matter of epistemology that we cannot be confident that we know anything. Rather, I mean a person with a critical mind, who does not accept assertions at face value. That has long been the mindset of the law of evidence. “In the everyday affairs of business and social life,” if you receive a letter purporting to come from a particular sender, you typically assume that the purported author is the sender. In contrast, both the common law of evidence and modern evidence statutes such as the Federal Rules adopt a different approach, requiring an affirmative showing of the authenticity of the letter. The very existence of the authentication requirement is perhaps the best evidence that “the common law [of evidence is] imbued with a spirit of skepticism.”

The trilogy reflects that skeptical attitude. For its part, Daubert rejects the traditional, general acceptance test. The Daubert court took the position that even if the entire specialty community vouches for a theory or technique, the collective ipse dixit is insufficient; according to Daubert, the trial judge can and should demand more by way of validation. Joiner applies the same questioning attitude to the individual expert. As previously stated, the Joiner Court declared that “nothing in either Daubert or the Federal Rules of Evidence re-

---

137. ZIMAN, supra note 1, at 109.
138. 2 McCORMICK, EVIDENCE § 218, at 37 (5th ed. 1999).
139. Id.
140. FED. R. EVID. 104(b), 901(a).
quires a district court to admit opinion evidence that is connected to . . . data only by the *ipse dixit* of the [individual] expert”¹⁴³—a statement reiterated in *Kumho*.¹⁴⁴ Neither the word of the individual expert nor even the collective word of his or her colleagues is adequate for the skeptical mind.

In the grand scheme of American evidence law, the judge ruling on admissibility is not only a skeptic, he or she is also a rationalist in the tradition of empiricism founded by John Locke, the seventeenth century English philosopher.¹⁴⁵ Enlightenment Rationalism had a profound effect on British and then American evidence law.¹⁴⁶ In particular, Locke’s empirical epistemology influenced many of the most important writers on evidence doctrine, including Gilbert,¹⁴⁷ Bentham,¹⁴⁸ Wills,¹⁴⁹ and Best.¹⁵⁰ To one degree or another, nearly all the leading Anglo-American writers on evidence have embraced these basic epistemological views.¹⁵¹ Their writings and the judicial opinions influenced by their writings gave birth to the celebrated Rationalist Tradition in evidence law.¹⁵²

The essence of Lockean empiricism¹⁵³ was his belief, stated in the *Essay Concerning Human Understanding*, that all our knowledge is founded in experience.¹⁵⁴ The tradition determines which modes of reasoning are considered valid.¹⁵⁵ For instance, in Book I of his essay, Locke repudiated the notion that certain ideas are innate to the human mind.¹⁵⁶ Instead, as Locke explained in Book II,¹⁵⁷ knowledge had to be derived directly or indirectly from experience.¹⁵⁸ As a result of experience, we directly gain “sheer data”¹⁵⁹ in the form of sensations. Upon careful reflection,¹⁶⁰ we can draw indirect, reliable inferences from the sheer data.¹⁶¹ However, in Locke’s view empirical inference encompassed more than strictly scientific experiments.¹⁶²

¹⁴⁶. Id. at 74.
¹⁴⁷. Id. at 35.
¹⁴⁸. Id. at 38.
¹⁴⁹. Id. at 47.
¹⁵⁰. Id. at 48.
¹⁵¹. Id. at 40.
¹⁵². Id. at 32-91.
¹⁵³. Id. at 83 n.26.
¹⁵⁴. CASTELL, supra note 86, at 170-71.
¹⁵⁵. Id. at 178; see also TWINING, supra note 145, at 80.
¹⁵⁶. CASTELL, supra note 86, at 170-71.
¹⁵⁷. Id. at 183-84.
¹⁵⁸. Id. at 175-77.
¹⁵⁹. Id. at 176.
¹⁶⁰. Id. at 175.
¹⁶¹. Id. at 176.
¹⁶². Id. at 171.
Both experts and laypersons are viewed as possessing the cognitive competence to draw trustworthy inferences from experiential data.163

In gauging the adequacy of the proponent’s foundation, what question should this skeptical, rationalist decision-maker ask? The skeptical, rationalist judge must demand proof that the expert’s specific theory or technique works; that is, the use of the theory or technique enables the expert to accurately make the inferential determination that the expert contemplates testifying to. As one of the leading commentators on expert testimony, Professor Paul Giannelli, has pointed out, when the concern is the “validity” of a theory or technique, the essential question is whether the theory or technique enables the expert to determine what he or she claims that it helps determine—“its accuracy.”164 The wording of Federal Rule of Evidence 901 is instructive. Subdivision 901(a) codifies the general requirement that proponents authenticate their evidence. In the words of the statute, “The requirement of authentication . . . as a condition precedent to admissibility is satisfied by evidence sufficient to support a finding that the matter in question is what its proponent claims.”165 Subdivision 901(b)(9) expressly extends the requirement to “[e]vidence describing a process or system used to produce a result and showing that the process or system produces an accurate result.”166 The accompanying Advisory Committee’s Note gives two scientific instruments, X-rays and computers, as illustrations of the reach of (b)(9).167 As the statute and Note indicate, the mark that the theory or technique must hit is accuracy.

What evidence should the trial judge weigh in resolving the crucial question of whether the theory or technique enables the expert to accurately make the inferential determination? Consistently with the mindset of empiricism, the judge ought to focus on the evidence of the results of the use of the theory or technique. In the macrocosm of society, why do we place faith in science? We do so in large part because there is an “immense body of results”168 proving “its worth in the realm of material technique.”169 Those concrete results—science’s many practical,170 successful applications and technological

163. TWINING, supra note 145, at 80.
165. Fed. R. Evid. 901(a).
166. Id. at 901(b)(9).
168. ZIMAN, supra note 1, at 6-7.
169. Id. at 2.
170. Id. at 127.
achievements— are the pragmatic basis for the belief in the validity of systematic experimental testing and induction.

In the microcosm of the courtroom as well, the focus should be on results. The Supreme Court reflected that focus in the trilogy. In his list of factors in Daubert, Justice Blackmun stated that “the court ordinarily should consider the known or potential rate of error.” Of course, the only way to ascertain that rate is to use the theory or technique and then to observe the results of that use. In Kumho, Justice Breyer asserted that in evaluating the reliability of an expert’s methodology, the judge should consider “how often an . . . expert’s . . . methodology has produced erroneous results.” Moreover, the 2000 Advisory Committee’s Note to amended Rule 702 states that a pertinent consideration is the “results” reached when the theory or technique is utilized.

B. Illustrative Applications of the Proposed Standard

Under this proposed standard, how should a judge rule on the following states of the record?

The proponent’s foundation consists of only the expert’s personal voucher that the theory or technique is valid. Joiner and Kumho supply the answer here. Joiner teaches that the trial judge need not accept the “ipse dixit” of an individual expert, and Kumho reaffirms that teaching. That teaching is in accord with the skeptical spirit of American evidence law. The trial judge should not simply take the expert’s word for the validity of the theory or technique.

Suppose that the proponent supplements the foundation with additional testimony that the vast majority of the specialists in the field believe that the theory or technique is valid. Now Daubert controls. There Justice Blackmun rejected general acceptance within the relevant specialist discipline as a dispositive consideration. In Kumho, Justice Breyer indicated that even a collective voucher by the practitioners of either astrology or necromancy would not suffice.

171. Id. at 10.
172. Id. at 127.
173. Id. at 10.
176. FED. R. EVID. 702 advisory committee’s note (“Whether the field of expertise claimed by the expert is known to reach reliable results for the type of opinion the expert would give.”).
178. Kumho, 526 U.S. at 146.
181. Kumho, 526 U.S. at 151.
part, the Advisory Committee’s Note to the 2000 amendment to Rule 702 states that a general voucher by the practitioners of clinical ecology would be an inadequate foundation. That statement is also consistent with the skeptical attitude with which the trial judge should approach the proponent’s foundation.

Assume that the proponent goes farther and offers at least some instances of results indicating that on occasion, reliance on the theory or technique has permitted an expert to accurately make the inferential determination that the witness proposes to draw on the stand. One or a few isolated anecdotes should not suffice. While a handful of successful results may warrant a serious, larger-scale investigation into the validity of the theory or technique, scanty anecdotal evidence is inadequate as a foundation. Empiricists, such as the eighteenth century Scottish philosopher David Hume, insisted upon a showing of “many instances”—a definite pattern of consistent outcomes—before they were willing to infer a relationship or connection.

Alternatively, suppose that the proponent can show that the theory or technique has been repeatedly used. Is that showing satisfactory as a foundation? The answer must be no when the expert is making an inferential claim. The longstanding use of the technique or theory is persuasive evidence that the members of that field accept the theory or technique, but general acceptance suffices only when the expert’s claim is descriptive in nature. Before embracing an inferential claim about the theory or technique, an empiricist should demand proof of the results of its use. Again, both Daubert and Kumho manifested concern about the results of reliance on the theory or technique. That concern is relevant in several forensic settings, including polygraphy and arson analysis. In some of the research cited in support of the admissibility of polygraph evidence, there were no follow-up studies to independently verify that the polygraphist’s conclusion was correct. Although the polygraph has

182. Fed. R. Evid. 702 advisory committee’s note (citing Sterling v. Velsicol Chem. Corp., 855 F.2d 1188 (6th Cir. 1988) (rejecting testimony based on ‘clinical ecology’ as unfounded and unreliable)).
183. Cavallo v. Star Enter., 100 F.3d 1150, 1158-59 (4th Cir. 1996).
185. Castell, supra note 86, at 191-93; see also Fed. R. Evid. 702 advisory committee’s note (stating that in assessing the adequacy of the proponent’s foundation, the judge should consider whether “the expert has unjustifiably extrapolated”).
186. Ziman, supra note 1, at 6, 10, 46, 75.
been used extensively for decades, an empiricist would not treat these studies as adequate validation without a showing of the accurate use of the technique, namely, the “ground truth”\(^{190}\) of the examiners’ findings. Arson analysts’ opinions are vulnerable to the same type of doubt. Arson investigators rely on certain clues such as concrete “spalling”\(^{191}\) and char depth\(^{192}\) to determine the point of origin of a fire. These clues are plausible and widely accepted by fire department arson investigators.\(^{193}\) The difficulty, though, is that there have been few “full-scale burns” of buildings to verify that such factors accurately identify the starting place of a fire.\(^{194}\) To make matters worse, “[n]othing in the natural world ‘tests’ an arson investigator’s expertise. If an arson investigator is wrong, nothing runs aground or burns down.”\(^{195}\) These clues are in widespread use, but there is little objective evidence that their use yields accurate results. Without any effort to detect error and evaluate the results of the use of the technique, the analysts might simply be repeating the same mistakes over and over again.

In the above states of the record, given the Supreme Court’s trilogy of decisions, without more the trial judge should rule the foundation inadequate. At the polar extreme, however, there are many foundational showings that the judge should rule sufficient.

In light of the Daubert Court’s emphasis on scientific testing, induction from empirical tests is certainly an acceptable method of validating an expert theory or technique. The ideal situation is one in which: The expert constructs a large database;\(^{196}\) the database is representative of the relevant universe;\(^{197}\) the test was conducted under conditions approximating those in the pending case;\(^{198}\) and the error rate is negligible.\(^{199}\) When the proponent’s foundation paints that impressive a picture of the state of the scientific research supporting the theory or technique, the judge ought to allow the proponent to submit the theory or technique to the trier of fact. The courts often

---

\(^{190}\) 1 FAIGMAN ET AL., supra note 189, § 14-3.3.2[3].

\(^{191}\) 2 GIANNELLI & IMWINKELRIED, supra note 56, § 26-4(A), at 557.

\(^{192}\) Id. § 26-4(A), at 556.

\(^{193}\) Id. § 26-4(A).

\(^{194}\) Vincent Brannigan & Jose Torero, The Expert’s New Clothes: Arson ‘Science’ After Kumho Tire, 43 FIRE CHIEF 60 (1999) (the authors are professors in the Department of Fire Protection Engineering at the University of Maryland).

\(^{195}\) Id. at 61.

\(^{196}\) EDWARD J. IMWINKELRIED, THE METHODS OF ATTACKING SCIENTIFIC EVIDENCE § 10-6(a) (3d ed., Lexis Law Publ’g 1997).

\(^{197}\) Id. § 10-6(b), at 297-98.

\(^{198}\) Id. § 10-6(b), at 298.

encounter such foundations for genetic marker analysis, such as DNA typing.200

Although the classic technique of experimental testing and induction is a permissible method of validation, that technique does not exhaust the possibilities. Members of the empirical school of epistemology recognized other valid modes of reasoning such as mathematical deduction.201 If algebraic or geometric reasoning demonstrates the logical necessity of a proposition, that demonstration would be acceptable to an empirical epistemologist; and the same demonstration should be adequate as a foundation in court.

However, assume that the proponent calls neither a molecular biologist nor a mathematician as her expert witness. Rather, the proponent invites an automobile mechanic to testify as an expert on engine malfunctions. Although the mechanic’s academic background might not be as impressive as that of the biologist or the statistician, this evidence should also be ruled admissible. There is far more than the bare fact that for decades, mechanics have used their techniques of automotive diagnosis and repair. In addition, during that period of time car owners have come to rely on mechanics to diagnose and repair engine malfunctions.202 The point is that car owners have turned to mechanics for such a lengthy period of time precisely because they have generally been satisfied with the results of the mechanics’ expert diagnosis and work. In a large number of cases, the mechanics succeeded in eliminating or reducing the problem that caused the owner to consult the mechanic in the first instance.203 The mechanics’ lay customers themselves are not experts; but, as previously stated, rationalist epistemologists agree that laypersons also possess the cognitive competence to evaluate results and draw inferences from empirical observations.204

These two sets of hypotheticals represent polar extremes on the spectrum of cases that will be presented to judges. Under the standard proposed at the outset of this Subpart, the proper outcome in these hypotheticals is fairly clear cut. However, other fact situations will fall in the middle of the spectrum. These fact situations can pose difficult, close calls for the trial judge. Joiner was such a case. There the plaintiffs presented several, large polychlorinated biphenyl (PCB) studies.205 Yet, like the trial judge, the Court was concerned about

200. 2 GIANNELLI & IMWINKELRIED, supra note 56, § 18-3(C), at 21-22 (reviewing some of the research validating the short tandem repeat (STR) method of DNA typing).
201. ZIMAN, supra note 1, at 102, 136, 150, 163; see also CASTELL, supra note 86, at 171, 195; TWINING, supra note 145, at 80.
202. Imwinkelried, supra note 97, at 35.
203. Id.
204. TWINING, supra note 145, at 80.
the dissimilarities between the facts of the pending case and the test conditions in the studies: The plaintiff was an adult human being while the animal studies involved infant mice; the mice had received massive doses, but, relative to body size, the plaintiff’s exposure was “far less”;206 the plaintiff’s exposure was dermal while the doses were injected into the mice; and the plaintiff developed a small-cell carcinoma, but the mice had alveologenic adenomas. Given those dissimilarities, the Court concluded that the trial judge did not abuse his discretion in barring the testimony. However, before sustaining the lower court ruling, the Court went to lengths to underscore that on review, the standard is abuse of discretion.207 Pointedly, concurring in part and dissenting in part, Justice Stevens added: “it bears emphasis that the Court has not held that it would have been an abuse of discretion to admit the expert testimony.”208 The standard proposed in this Subpart should reduce the uncertainty in this area of law, but there will still be cases such as Joiner, presenting the courts with tough judgment calls.

V. CONCLUSION

Daubert is a landmark decision. Much of the attention for the decision has been devoted to the Court’s formal holdings that the general acceptance test is no longer good law and that Rule 702 requires a validation foundation for scientific testimony. However, the real significance of the opinion may lie in the Court’s frank recognition of the limits of the scientific enterprise.209 With the benefit of a number of amicus briefs filed by scientific organizations,210 the Court moved beyond the flawed, popular notion that at least the hard sciences can yield absolute certainty.211 The Court confronted the reality that there are erratic phenomena that defy even the most meticulous application of the scientific method.212 Citing the amicus briefs, Justice Blackmun therefore wrote that “arguably, there are no certainties in science.”213 Daubert should help to disabuse the lower courts from a simplistic faith in the scientific method. It would be ironic if the trilogy, begun in Daubert, was interpreted as standing for the proposition that the scientific method is the only method of validating expert

206. Id. at 144.
207. Id. at 139, 141-43.
208. Id. at 150, 155 (Stevens, J., concurring in part and dissenting in part).
209. Imwinkelried, supra note 68.
210. Id. at 64-65. The amici included the American Association for the Advancement of Science, the National Academy of Sciences, the American Medical Association, and the Carnegie Commission on Science, Technology, and Government.
211. Id. at 59-60.
212. Id. at 60-63.
testimony or even that there is a bias in favor of empirical testing and induction.

"[W]e have cast off the naive doctrine that all science is necessarily true . . . ."\textsuperscript{214} It is silly to presume that at any given time, all the propositions circulating within the scientific community—or even the specialized scientific communities of biology, chemistry, or physics—are sufficiently validated. Quite to the contrary, the discourse in any scientific circle is likely to include a mix of long validated propositions, newly validated ones, conjectures, and propositions that have already been discredited.\textsuperscript{215} In that light, the answer to the question, "What must be validated?," should be clear. The focus must be on the specific theory or technique that the expert proposes to rely on to perform "the task at hand,"\textsuperscript{216} the determination the expert contemplates testifying about.

We have also "cast off the [equally] naive doctrine that . . . all true knowledge is necessarily scientific."\textsuperscript{217} In the long term, perhaps the most important insight in \textit{Daubert} is the Court's recognition of the fallibility and inherent limitations of scientific methodology. Anglo-American evidence law rests on empiricist epistemology. That school of epistemology recognizes the validity of modes of reasoning other than scientific induction. Mathematical deduction is also permissible, and even laypersons are viewed as possessing the cognitive competence to draw reliable inferences from sensory data. An answer to the question, "How must it be validated?," thus emerges. If the expert makes a limited, descriptive claim, there is adequate validation when the foundation demonstrates that the expert has had a large number of similar experiences. When the expert makes an inferential claim, the question becomes whether the proponent has made a showing of the results of the use of the theory or technique that would convince a skeptical rationalist that the theory or technique accurately does what the expert claims.

It would be a grave mistake to interpret the trilogy as if it mandated that the judge assess expert foundations from the narrow perspective of the scientific tradition. Rather, the trilogy should be read through the broader lens of the Rationalist Tradition of Anglo-American evidence law.\textsuperscript{218} Prescribing a scientific perspective might simplify the trial judge’s task; in many cases, it will be perfectly clear that the proponent’s foundation does not qualify as controlled, experimental verification. However, adopting that perspective will also result in the exclusion of a vast amount of evidence that satisfies

\textsuperscript{214} ZIMAN, supra note 1, at 2.
\textsuperscript{215} Id. at 130-33.
\textsuperscript{216} Daubert, 509 U.S. at 597.
\textsuperscript{217} ZIMAN, supra note 1, at 2.
\textsuperscript{218} TWINING, supra note 145, at 32-91.
empirical epistemology, as Locke conceived it. Ease of use makes a narrow scientific perspective tempting; but as Justice Breyer wisely cautioned in *Kumho*, “Too much depends upon the particular circumstances of the particular case at issue.”\(^{219}\) Unlike dogmatic\(^{220}\) scientism, the Rationalist Tradition has the intellectual breadth and flexibility to accommodate those myriad circumstances.


\(^{220}\) Schwartz, *supra* note 29.
THE PROPERTY WARS OF LAW FIRMS: OF CLIENT
LISTS, TRADE SECRETS AND THE FIDUCIARY
DUTIES OF LAW PARTNERS

ROBERT W. HILLMAN*

I. TRADE SECRETS ...................................................................................................  770
   A. In General .................................................................................................  770
   B. Customer Lists as Trade Secrets ...............................................................  772
   C. Law Firm Client Lists as Trade Secrets ....................................................  773
      1. Law Practice and Information “Used in One’s Business” .................  773
      2. The “Independent Economic Value” of Client Information ...........  773

II. THE LEGAL ETHICS PERSPECTIVE ...................................................................  775

III. THE LAW FIRM CASES ..................................................................................  777
   A. Fred Siegel Co. v. Arter & Hadden .........................................................  777
   B. Early, Ludwick & Sweeny v. Steele .........................................................  780

IV. A CLOSER LOOK AT THE TRADE SECRET FORMULA AS APPLIED TO LAW FIRMS.  781
   A. Ascertainability .......................................................................................  781
   B. Efforts to Maintain Secrecy .....................................................................  782
   C. Misappropriation .....................................................................................  784
   D. Remedies ................................................................................................  785

V. TRADE SECRET PROTECTIONS AND FIDUCIARY DUTIES COMPARED .......  786

VI. A FINAL NOTE: THE ROLE OF CONTRACTUAL REMEDIES .........................  788

The dramatic rise in litigation among former law partners is a
well-documented feature of the legal profession’s landscape. An
impressive body of law underlying lawyer mobility has developed in the
relatively short period during which law partner litigation has be-
come commonplace.¹ With each reported decision, the litigation has
yielded substantive legal standards regulating the relationships of
partners associated in a law practice.²

An important theme emerging from the lawyer mobility litigation
is the primacy of the interests of clients in resolving disputes among
former law partners. No other consumer class can match clients of
law firms in terms of protections accorded by law. Already intense
competition for clients has been bolstered by rigorous enforcement of
bans on anticompetitive practices by law firms. The law firm cannot
secure its investment in a client through a long-term contract bind-

* © 2003 by Robert W. Hillman. Professor of Law, University of California, Davis.
ing the client to the firm. The law firm cannot impose economic penalties on those who leave the firm and take its clients. And more to the point, the law firm cannot enforce against departing members contracts that in any way limit those members in competing for the firm’s clients. Litigation has not been kind to law firms seeking to protect their client base.

One of the casualties of this short but intense period of litigation among former law partners is the traditional notion that a law firm enjoys a proprietary interest in its clients. A little more than twenty years ago the Pennsylvania Supreme Court chastised lawyers who took cases from a firm by noting that “[n]o case on the list, however, was [theirs]. Rather, each case was [a firm] case on which . . . [they] were working.” Indeed, a lower court judge in the same case put it even more succinctly: “If they want their own firm, let them get their own clients.” The Pennsylvania cases speak the language of the past. If the subsequent explosion of lawyer mobility litigation has demonstrated anything it is that neither law firms nor attorneys within the firms may claim clients as their own. In short, neither firm nor lawyer has a proprietary interest in clients.

Although the view of clients as property of the firm is no longer part of the lawyer mobility landscape, there exists a variation of the property approach that may have some vitality. Law firms, like other professional service organizations, develop knowledge relating to their practices. For want of a better term, this may be referred to as the intellectual property of law firms, a term that may include, for example, client lists, billing data, form files, compensation practices,

3. See, e.g., In re Cooperman, 633 N.E.2d 1069, 1072-73 (N.Y. 1994) (voiding a nonrefundable retainer agreement because it “inappropriately compromises the right to sever the fiduciary services relationship with the lawyer”).

4. See, e.g., Jacob v. Norris, McLaughlin & Marcus, 607 A.2d 142, 148 (N.J. 1992) (voiding partnership agreement varying payouts depending upon whether departing partners took clients and employees from the firm).

5. See generally Hillman, supra note 1, § 2.3.3.

6. Epstein, 393 A.2d at 1177.


8. It seems to be standard practice in commentary on “intellectual property” not to define the meaning of the term. See, e.g., Peter A. Alces & Harold F. See, The Commercial Law of Intellectual Property (1994); Michael A. Epstein, Epstein on Intellectual Property (4th ed. 1999). As the term is here used, law firm intellectual property simply means nonpublic information (i.e., information not generally known or readily ascertainable) that has value to the firm. The issue of whether the particulars of the firm’s rights in such information are protected by law, however, is another issue. Cf. Donald A. Gregory et al., Introduction to Intellectual Property Law 1 (1994) (“Intellectual property is an intangible that is not easily described. The law creates the property by defining what will be protected from others. Intellectual property is created and protected based upon policy considerations as to what types of intellectual activities should be encouraged.”).
and personnel information.9

The status of a law firm’s information and the right to control its use are issues of great practical importance. In an environment in which “portable” business drives lawyer mobility,10 a firm entertaining the possibility of hiring a lateral partner likely will want to know something about the partner’s clients, billings and income as well as the resources that will be required in order to support the lawyer’s practice. Information needs grow and become more complex when the partner proposes to move an entire team (partners, associates, paralegals, etc.) from one firm to another. Whether one more person or many are involved in the move, there remains the larger issue of whether information acquired while at one firm may be used after affiliation with a new firm.

There exist a variety of legal doctrines relevant to a lawyer’s use of information acquired while a member of a law firm. The fiduciary duties of partners and, in the case of associates, employees may bear on the activities of withdrawing lawyers before and after they leave their firms. The difficulty with fiduciary standards, however, lies in the indeterminacy of their application.11 In Gibbs v. Breed, Abbott & Morgan, for example, a New York appellate court took a harsh view of law partners who shared with another firm.12 They were contemplating joining information on associate salaries and billing.13 Signaling the difficulty of issues pertaining to the use of firm information, however, Gibbs included a forceful dissent that found nothing at all inappropriate about sharing such data with another firm.14 Fiduciary duties are more likely to be indeterminate when used to define rights in information than they have been in when used to order other aspects of the relations among law partners.

Although fiduciary duties are the logical starting point for evaluating the conduct of partners who utilize for private benefit informa-

9. Normally, rights in intellectual property are transferable. Insofar as information relates to client matters, however, law firms operate under restrictions that have effectively eliminated the option of selling the property. More recently, a number of states have amended their ethics rules to allow, under defined circumstances, the sale of the goodwill associated with a law practice. The easing of restrictions follows the 1990 amendment of the Model Rules of Professional Conduct to allow lawyers to buy and sell law practices when certain conditions are satisfied. See Model Rules of Prof’l Conduct R. 1.17 (2001); see Hillman, supra note 1, § 2.5.3.
10. See Hillman, supra note 1, § 1.1.
13. The information included salaries, annual billable hours, and billing rates. Id. at 580. The partners planned to recruit the associates to their new firm. Id.
14. Id. at 590 (Saxe, J., concurring in part and dissenting in part) (disagreeing with the majority as to the confidentiality of the information disclosed and calling the information “the greatest unkept secret in the legal profession”).
tion they acquire while at a firm, firms seeking redress may be expected to employ the distinct but compatible body of law governing trade secrets. They will receive considerable support for this effort from the decision of the Ohio Supreme Court in Fred Siegel Co. v. Ar
ter & Hadden, the first state high court opinion to conclude that a law firm’s client list may be a trade secret.15 Perhaps the most sur-
prising aspect of Siegel is that the trade secret analysis has come so late to the world of lawyer mobility litigation. Now that it has ar-
rived, we may expand allegations of trade secret misappropriation so that it will become a routine part of litigation among former law
partners.

Whether viewed through the lens of fiduciary norms or trade secret protections, one of the more important and unresolved issues raised by lawyer mobility trends is the extent to which a law firm has a recognizable interest in the information pertaining to clients that it possesses. If such an interest exists and is protected by law, the information may properly be regarded as the intellectual property of law firms. In settings outside of the legal profession, disputes over information rights typically are limited to two claimants—the original possessor of information and the party who presently desires to use the information. In the law firm setting, however, a third party—the client—is introduced, and in most respects the interests of this party are superior to those of either the firm or its departing partner.

This Article explores the nature of information about clients generally and client lists specifically as the intellectual property of law
firms. It examines whether client information is protectible as a trade secret and the extent to which fiduciary norms preclude use of such information by withdrawing partners.

I. TRADE SECRETS

A. In General

Trade secret law recognizes and protects some types of valuable business information.16 As one court put it, “[t]he trade secret is a type of intellectual property, in effect, a property right in discovered knowledge.”17 In core concepts, this is ancient law, traceable at least to Roman times.18 In its Anglo-American development, trade secret law is closely connected with norms of commercial morality. Indeed,

one commentator stated it well:

Imagine for a moment a commercial society that has never heard the term “fiduciary obligation,” and where no one owes duty of fair play to anyone else. . . . A moment’s reflection about such an abominable business climate leads to the recognition of practical reasons for the development of the law of trade secrets. 19

Reflecting the substantial common law jurisprudence on trade secrets, the *Restatement (First) of Torts*, promulgated in 1939, stated that an individual “who discloses or uses another’s trade secret, without a privilege to do so, is liable to the other if . . . his disclosure or use constitutes a breach of confidence reposed in him by the other in disclosing the secret to him.” 20 The common law underpinnings of trade secret law are now reflected in statutes that have been adopted in most states. For the most part, the statutes implement the *Uniform Trade Secrets Act* (UTSA), which defines a trade secret as follows:

> [I]nformation, including a formula, pattern, compilation, program, device, method, technique, or process, that:
> (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and
> (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy. 21

Information is not a trade secret simply because the person or firm in possession of the information regards it as confidential. 22 Factors relevant to according information the status of a trade secret include:

- the extent to which the information is known outside the business and by employees and others involved in the business, the measures taken by the employer to guard the secrecy of the information, the information’s value to the employer and to competitors, the resources the employer expends in developing the information,

---

19. Melvin F. Jager, *Trade Secrets Law* § 1.03 (2000). Mr. Jager also notes, however, that a useful definition of trade secrets “is not readily discernable from the cases.” *Id.* § 2.01.

20. *Restatement (First) of Torts* § 757 (1939) (emphasis added). The section also states that liability exists if the trade secret is discovered by improper means. *Id.* A more current restatement of trade secret protections is offered in the *Restatement (Third) of Unfair Competition* §§ 39-45 (1995).


and the ease or difficulty with which the information could be properly acquired or duplicated by others.23

Also relevant to this equation is a balancing of the “conflicting rights of an employer to enjoy the use of secret processes and devices which were developed through his own initiative and investment and the right of employees to earn a livelihood by utilizing their skill, knowledge and experience.”24 To some extent, the product of this balancing may be seen in authority that would disable the employer from barring use of the information if the employee is able to derive the information through independent invention or “reverse engineering” (i.e., starting with the trade secret and working backward to find the method by which it was developed).25

B. Customer Lists as Trade Secrets

Information concerning the consumers of a firm’s products or services may be valuable business information. In some cases, the information may have value to the firm, or a competitor, even if it offers little more than the names and contact information of individuals with whom the firm has had a business or professional relationship.

Although only a small number of states that have adopted the UTSA expressly apply the act to a list of actual or potential customers26 or the equivalent, many jurisdictions have, through case law, applied trade secret precepts to protect customer lists.27 Moreover, as the Restatement (Third) of Unfair Competition states, “[t]he general rules that govern trade secrets are applicable to the protection of information relating to the identity and requirements of customers.”28 Of course, not all customer lists are trade secrets. Customer lists derived from readily-identifiable sources may not be protected, although specific information concerning customers (such as contact persons or special needs) may qualify as a trade secret even when the list does not.29 Moreover, firms may have a particularly difficult time in preventing the use of client information by individuals having

25. Id. at 476.
29. GREGORY ET AL., supra note 8, at 206.
close relationships with the clients.  

In part because trade secret status may be defeated by nondefinitional factors such as the failure to take steps to maintain the secrecy of information, the line between those customer lists that are trade secrets and those that are not is blurred.

C. Law Firm Client Lists as Trade Secrets

1. Law Practice and Information “Used in One’s Business”

By defining a trade secret as “any formula, pattern, device, or compilation of information which is used in one’s business,” the Restatement of Torts seemed to exclude protection for information that has value but is not actually used in a business. For years, debate has ensued over whether the practice of law is a business or a profession, and depending on where one stands in this debate the Restatement, read literally, may negate the status of information in the possession of law firms as trade secrets. For two reasons, however, the distinction between a business and profession should have little effect on the trade secret analysis. First, the view that law practice to a significant extent is a business (albeit with the underpinnings of a profession) has gained the upper hand in recent years. Second, and more to the point, the UTSA drops the “use in business” portion of the definition of a trade secret, leading to the perhaps unintended effect that protection under the statutes may extend to information used in non-business ways.

2. The “Independent Economic Value” of Client Information

To qualify as a trade secret under the UTSA, information must derive “independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper

30. See Moss, Adams & Co. v. Shilling, 179 Cal. App. 3d 124, 129 (Cal. Ct. App. 1986) (holding that names of clients employees had worked for prior to their withdrawal from accounting firm were not trade secrets).

31. See infra text accompanying notes 76-81.

32. Restatement (First) of Torts § 757 cmt. b (1939).

33. But cf. Restatement of Unfair Competition § 39 (1995) (defining a trade secret as “information that can be used in the operation of a business or other enterprise”).

34. Consider, for example, the widely followed American Lawyer rankings of firms on the basis of profitability per partner. America’s Highest-Grossing Law Firms in 2001, 24 AM. LAW. 203 (July 2002).

35. See, e.g., Alces & See, supra note 8, at 83-84.

[O]ne may have a secret method for applying paints to model airplanes and use that purely for hobby purposes. There are many other hobbyists who might wish to have such information if it were available and, therefore, this method would derive significant economic value from not being generally known. Under the Uniform Act, such information would appear to qualify as a trade secret; under the Restatement it would not.

Id.
means by, other persons who can obtain economic value from its disclosure or use.”

Normally, the identity of law firm clients is readily apparent and often even is a matter of public record. For this reason, information concerning the identity of a specific client normally is easily ascertainable with a minimum of effort and should not qualify for trade secret protection. As is the case with the customer list of a commercial enterprise, the real value of client information lies in the list. The longer and more complex the list becomes, the greater is the difficulty of independently creating the information. The very difficulty in the recreation is one element of the value of the client list.

Difficulty of replication is not the sole determinant of the value of the information contained in a client list. The relationship that exists or existed between a firm and its clients distinguishes the firm from countless other firms which may wish to compete for the attentions of the clients and provide noisy communications to this effect. The relationship is of value not only to the firm but also to any of its members whom clients may identify as forming part of that relationship. The goodwill associated with the relationship may be used to good advantage by a lawyer within the firm who wishes to solicit the client. Indeed, in some cases, the client may have greater allegiance to the soliciting lawyer than to the firm. Even when the relationship between soliciting lawyer and client is more attenuated, however, a departing lawyer or former member of the firm may use successfully the relationship between the firm and the client as the basis for the lawyer’s solicitation attempts. For this purpose, a client list may be of real assistance to the soliciting lawyer.

The potential value of the client list is greatest when the number of clients is large and, therefore, the information reflected on the list is extensive. Consider, for example, the strategic use of a client list in an early and infamous lawyer mobility case. In the case cited earlier in this Article as a paradigm of the now-defunct clients as assets of the firm view, lawyers in the firm notified approximately four hundred of the firm’s clients that they were forming a new firm and that the clients were free to retain counsel of their choice. The contacts were by phone as well as letter. Although the associates had worked on the client matters, the use of a client list undoubtedly facilitated the communications. The case illustrates the value of client lists in practices involving large numbers or relatively unsophisticated clients.

37. See infra text accompanying notes 45-46.
For practices of this type, client lists may seem to possess the requisite “independent economic value” to justify protection as trade secrets. This assumes, however, that a law firm is no different from other commercial enterprises in its entitlement to trade secret protection for client information. Although there is growing acceptance that law partnerships are profit-seeking firms, they do operate under unique ethics standards designed to protect the consumers of their services. As is discussed below, these standards may limit significantly trade secret protections that may otherwise be available.

II. THE LEGAL ETHICS PERSPECTIVE

Norms of legal ethics have shaped the ground rules that regulate the relationships of lawyers associated in law firms. In particular, the principle of client choice, which allows clients to freely discharge lawyers or firms, undermines the expectations of firms regarding continued client patronage and loosens the bonds that tie lawyers associated in law practices.\(^{39}\) Interestingly, law is unique among the professions in the degree to which the will of clients is a paramount value in shaping ethics norms,\(^{40}\) although it should be emphasized that the policy reason supporting such a privileged status for consumers of legal services has not yet been articulated.\(^{41}\)

Client choice operates to void traditional contractual means of tying clients to firms. The classic contractual device for restraining future competition is the restrictive covenant that prohibits an individual from post-withdrawal competition with a firm. The restrictive covenant, however, is not an option for law partners desiring to restrain future competition. Courts invariably decline to enforce provisions of partnership agreements that restrict the ability of lawyers to compete with the firms from which they have withdrawn.\(^{42}\) Even cli-

---

\(^{39}\) See, e.g., Charles W. Wolfram, Modern Legal Ethics § 9.5.2, at 545 (1986) (“It is now uniformly recognized that the client-lawyer contract is terminable at will by the client. For good reasons, poor reasons, or the worst of reasons, a client may fire the lawyer.”).

\(^{40}\) See, e.g., Riordan v. Barbosa, No. 395945, 1999 Conn. Super. LEXIS 446, at *21 (Conn. Super. Ct. Mar. 1, 1999) (enforcing a restrictive covenant against an accountant and noting that “[t]here is no per se distinction between so-called professional people and other members of the work force with respect to the reasonableness of a noncompetition covenant”). See Hillman supra note 1, § 2.3.3 (discussing the legal, accounting, and medical professions).

\(^{41}\) Inexplicably, professional responsibility standards seemingly provide greater protections for clients than provided throughout the counterpart ethics standards of the other professions. A recent indication of the disparity may be seen in restrictions placed on the ability of lawyers to take advantage of new firm structures offering limitations of liability. For a discussion of this point, see Robert W. Hillman, Entity Rationalization and Professional Service Firms, 58 Bus. Law. (forthcoming August 2003).

\(^{42}\) Both the Model Code of Professional Responsibility and the Model Rules of Professional Conduct void restrictive covenants by prohibiting a lawyer from agreeing to restrict the lawyer’s practice after termination of the relationship created by the agreement. Model Rules of Prof’l Conduct R. 5.6 (1983); Model Code of Prof’l Responsibility
ents who have had retained firms to prosecute litigation on a contingent fee basis may change firms mid-stream and compensate the discharged firms on limited, *quantum meruit* basis, thus denying the firms the benefits of the bargains they earlier struck with their clients. The ease with which clients may change law firms not only sets the stage for intense competition for clients but also undermines any treatment of clients as *assets* of the firms they presently retain.

Given the environment of intense competition for clients, information concerning the specific clients of a particular firm may be useful to those who plan to compete with the firm through a future solicitation of its clients. The greatest risk to the firm arises when those who are planning for future competition are presently members of the firm and have relatively easy access to the client information. Viewed solely from the commercial perspective, the information may be protectible as a firm trade secret. Introducing the principle of client choice to the analysis, however, renders problematic trade secret protection for the client list. When broad prohibitions against use of client information operate to restrict choices available to clients, the firm’s interest in maintaining its competitive position is likely to be subordinated to the clients’ interest in retaining the law firms of their choice. Stated another way, the right to control use of information about law firm clients vests at least as strongly in the clients as it does in the law firms.

There exists a further, deeper-rooted difficulty with extending trade secret protection to information concerning a law firm’s clients. A recurring question in professional responsibility is whether client relationships run to firms or to their lawyers. The canons of ethics emphasize lawyer-client relationships, while courts commonly assume the relevant relationships are between clients and their firms. From the clients’ perspective, the question may be more nuanced than either the canons or the courts recognize. Clients sometimes view their relationships as with firms, and at other times with lawyers within the firms. Sophisticated clients seem especially inclined

---

43. *See supra* text accompanying notes 26-31.

44. *See supra* text accompanying notes 26-31.

45. *See* ABA CANONS OF PROF’L ETHICS Canon 44 (1931).
to “hire the lawyer rather than the firm.”\textsuperscript{46} The firm’s expectation that information concerning a client is confidential and will not be used to the firm’s disadvantage is undermined when the client views the primary relationship as running to the lawyer rather than the firm. The weakening of that expectation calls into question the correctness of protecting client lists as law firm trade secrets.

III. THE LAW FIRM CASES

Only sparse case law exists concerning the status of client information as a trade secret. The most important of the reported opinions are the 1999 decision of the Ohio Supreme Court in \textit{Fred Siegel Co. v. Arter & Hadden}, which expressly stated that a law firm’s client list may be a trade secret,\textsuperscript{47} and \textit{Early, Ludwick & Sweeny, LLC v. Steele}, a Connecticut lower court decision that carefully distinguished law firms from commercial enterprises in rejecting claims that a client list was entitled to protection as a law firm’s trade secret.\textsuperscript{48}

A. \textit{Fred Siegel Co. v. Arter & Hadden}

In \textit{Siegel}, an associate leaving the firm took with her a sixty-three page list that included the names, addresses, and phone numbers of hundreds of clients.\textsuperscript{49} She later used the list to solicit the clients.\textsuperscript{50} Applying the Ohio statute, the court concluded that the list could be a trade secret and that issues of material fact made the lower court’s award of summary judgment for the defendants on this point in error:

\begin{itemize}
\item \textsuperscript{46} In a Corporate Legal Times Roundtable, for example, one corporate counsel commented, “[T]he old adage that you don’t hire the firm, you hire the lawyer is so true. And I want the lawyer who’s working on my matters day-in and day-out to have that service mentality. It’s just so important to making sure that our needs are met.” Successfully Moving Up: Hiring in Times of Change: Tips for the Restless, Corp. Legal Times, May 1996, at 44. The point that clients hire lawyers rather than firms is developed more fully in Robert W. Hillman, Loyalty in the Firm: A Statement of General Principles on the Duties of Partners Withdrawing from Law Firms, 55 Wash. & Lee L. Rev. 997, 1011 n.54 (1998).
\item \textsuperscript{47} 707 N.E.2d 853, 864 (Ohio 1999).
\item \textsuperscript{49} 707 N.E.2d 853.
\item \textsuperscript{50} See id. at 861.
\end{itemize}
[L]istings of names, addresses, or telephone numbers that have not been published or disseminated, or otherwise become a matter of general public knowledge, constitute trade secrets if the owner of the list has taken reasonable precautions to protect the secrecy of the listing to prevent it from being made available to persons other than those selected by the owner to have access to it in furtherance of the owner's purpose.\(^{51}\)

The court found that issues of material fact existed concerning whether the firm took reasonable steps to protect its interest in the list.\(^{52}\) On this point, the record showed that access to the list on a computer required use of a password and that “[h]ard copies of the list were stored within office filing cabinets, which were sometimes locked.”\(^{53}\) The firm’s senior partner testified that he probably had told employees the information was confidential.\(^{54}\) The defendants, on the other hand, argued the information was a matter of public record and was capable of being independently created.\(^{55}\)

To assist the trial court in determining whether the client list is a trade secret, the court offered the following guidance:

Where information is alleged to be a trade secret, a factfinder may consider, e.g., the amount of effort or money expended in obtaining and developing the information, as well as the amount of time and expense it would take others to acquire and duplicate the information. . . . The extensive accumulation of property owner names, contacts, addresses, and phone numbers contained in the Siegel client list may well be shown at trial to represent the investment of Siegel time and effort over a long period.

The purpose of Ohio’s trade secret law is to maintain commercial ethics, encourage invention, and protect an employer’s investments and proprietary information. That purpose would be frustrated were we to except from trade secret status any knowledge or process based simply on the fact that the information at issue was capable of being independently replicated.\(^{56}\)

Assuming a firm follows the hints offered in Siegel (i.e., it locks the file cabinet, changes the password frequently, and stamps the document confidential), it has an arguable claim that a lengthy client list is a trade secret of the firm. Whether this result is desirable, and how far the logic of trade secret thinking extends, are distinct and important questions.

\(^{51}\) Id. at 862.
\(^{52}\) Id.
\(^{53}\) Id. (emphasis added).
\(^{54}\) Id.
\(^{55}\) Id.
\(^{56}\) Id. at 862-63 (citations omitted).
As to the desirability of the result, what is missing from the majority’s opinion is any consideration of the interests of clients.\textsuperscript{57} In this regard, the purpose for which the information is used should be considered in assessing the nature of the firm’s rights in the information. For example, a withdrawing lawyer’s use of otherwise protected information for the purpose of informing clients the lawyer has represented of the lawyer’s change in firm affiliation may be appropriate in light of the fundamental right of clients to choose their lawyers. Although the associate in \textit{Siegel} took a list containing information on all the firm’s clients, apparently she did so for the purpose of using the information to contact clients for whom she had worked while at the firm.\textsuperscript{58} If that is so, treating the portions of the list used by the associate as a trade secret of the firm would directly undermine the ability of clients to choose their lawyers.

An additional fact in \textit{Siegel} bears emphasis because it suggests an important limitation on application of trade secret law to client lists. While at the firm, the associate had maintained a rolodex with information on clients with whom she worked.\textsuperscript{59} When she left the firm, she took the rolodex with her.\textsuperscript{60} The court of appeals held, simply, that the rolodex “contains the names of clients she worked for and she could properly retain this information.”\textsuperscript{61} The firm did not appeal on this issue.\textsuperscript{62} If rolodex information is not a trade secret, then it is likely that information limited to data on clients that the withdrawing lawyer had represented is not protectible as a trade secret, a result that seems eminently sensible.\textsuperscript{63}

There was a spirited dissent in the case. Noting that “[c]lients may not be reserved to any lawyer or firm as a trade secret”\textsuperscript{64} in his dissent, Justice Cook distinguished between the identity of clients included on a list, which may not be trade secrets, and other information incorporated into a client list. As to the latter, he concluded

\begin{itemize}
\item \textsuperscript{57} Cf. Early, Ludwick & Sweeney, LLC v. Steele, No. CV9804090638, 1998 WL 516156, at *4 (Conn. Super. Ct. Aug. 7, 1998) (“[T]he court finds that the dispute presented here is not purely entrepreneurial or commercial. The dispute between the parties here is as to who is entitled to represent the clients concerned. It necessarily implicates the clients’ rights, including their right to choose their lawyers.”).
\item \textsuperscript{58} Siegel, 707 N.E.2d at 857.
\item \textsuperscript{59} Id. at 856.
\item \textsuperscript{60} Id. at 857.
\item \textsuperscript{61} Fred Siegel Co. v. Arter & Hadden, No. 71440, 1997 WL 428629, at *4 (Ohio Ct. App. July 31, 1997).
\item \textsuperscript{62} Siegel, 707 N.E.2d at 863.
\item \textsuperscript{63} See also Sonkin & Melena Co. v. Zaransky, 614 N.E.2d 807, 861 (Ohio Ct. App. 1992) (concluding the firm’s workers’ compensation client list maintained in its computer data base could be a trade secret but that the firm had failed to prove it had taken affirmative steps to protect the information).
\item \textsuperscript{64} Siegel, 707 N.E.2d at 864 (Cook, J., dissenting).
\end{itemize}
trade secret law would protect Siegel’s investment in developing the compilation aspects of its client lists—the cross-referencing of a given client’s name with parcel numbers, the names of other property owned or managed by that same client, billing names and phone numbers, the identity of clients if different from the named tax plaintiff, the identity of owners of properties, and the contact people of leased property.65

B. Early, Ludwick & Sweeny v. Steele

In Steele, the withdrawing lawyer contacted sixteen clients he had represented in pediatric lead poisoning cases while at the firm.66 The firm asserted “that the names, addresses and telephone numbers, guardians, blood lead levels, and insurance coverage” constituted a client list that was the firm’s trade secret.67 The court acknowledged that a client list may be a trade secret but concluded that the information at issue in this case did not so qualify because the requisite standard of secrecy had not been satisfied.68 It added:

Defendant Steele acquired the client list by proper means, in the course of work for plaintiff. Having worked on the cases in question and established relationships with these clients, Steele was entitled to notify them of his change of employer and to signify his willingness to represent them if they so desired. . . .

The efforts made by ELS to maintain secrecy of the alleged trade secrets appear to be no more than the usual precautions taken by a law firm to ensure clients’ files remain confidential. . . .

. . . . The central issue in this case—the use by Steele of the client list—and the relief sought by plaintiff—an order enjoining Steele from using the list—directly implicate the said clients’ rights to choose their representation. . . . Were the court to grant the relief requested, the clients’ right to change counsel would be restricted. This would clearly be contrary to public policy. . . .

. . . . The court notes in passing it is highly unlikely that the clients in question, in choosing ELS to represent them, contemplated that ELS thus acquired a proprietary interest in their names, addresses, telephone numbers, medical conditions and blood lead levels such as to restrict said clients’ freedom to change lawyers as the clients see fit.69

In at least two ways, Steele departs from the analysis offered in Siegel. First, Steele considers the effect on clients of extending trade

65. Id. at 865 (Cook, J., dissenting).
67. Id. at *3.
68. Id. at *2.
69. Id. at *3-*5.
secret protection to the client list. In contrast, Siegel offered a more generic trade secret analysis that might be applied without modification to the client list of an accounting firm or, for that matter, any commercial venture. As a matter of policy, Siegel's approach may be sensible, but the failure of the court to even recognize the elevated role of client choice in the law firm context disconnects the opinion from previously unchallenged norms of legal ethics.

Steele also differs from Siegel as to the inclusiveness of the information at issue. In Steele, the client information used by the departing lawyer was limited to sixteen matters for which the attorney had been responsible while at the firm. In contrast, the Siegel information consisted of sixty-six pages of data that included information on all of the firm's clients (numbering in the hundreds). To the extent information is overinclusive and extends well beyond the clients and matters handled by the attorney while at the firm, the argument that client interests preclude protection of the firm's interest in the information weakens. Stated another way, the differing results in the two cases may be understood by recognizing that the associate in Siegel took far more information than was necessary in order for the clients to be given a reasonable choice of whether they wish to be represented by the firm or its departed lawyer.

IV. A CLOSER LOOK AT THE TRADE SECRET FORMULA AS APPLIED TO LAW FIRMS

The awkwardness of treating a law firm client list as a trade secret becomes apparent upon examination of the conditions for the protection of information as a trade secret. To be sure, a client list rather easily may constitute a “compilation” as that term is used in the UTSA's definition of a trade secret. The Uniform Act's definition also includes, however, requirements that information be protected. The Act requires that the information is not readily ascertainable by proper means and that efforts have been made to maintain its secrecy. Moreover, it is only the act of misappropriation that entitles the firm to injunctive relief or damages. Each of these conditions merits closer examination.

A. Ascertainability

Courts vary in application of the standard that information must not be readily ascertainable by proper means. One court has reasoned that time is a factor in this analysis, so that information that can be reverse engineered only through the expenditure of consider-
able time is not readily ascertainable.\textsuperscript{73} Other courts, however, do not emphasize time and resource expenditures in the ascertainability inquiry.\textsuperscript{74} Under either approach, a law firm client list including hundreds of clients, each of whom has a relatively small claim, embodies information that likely is not readily ascertainable from other sources.\textsuperscript{75} As the number of clients decreases, however, the ease with which reverse engineering may be accomplished increases and trade secret status of client information becomes more problematic.

B. Efforts to Maintain Secrecy

Both \textit{Siegel} and \textit{Steele} properly emphasize that a firm that has not taken steps to protect the secrecy of client information will be unable to protect the information as a trade secret.\textsuperscript{76} As another court put it in a case involving a commercial customer list, “It would be anomalous for the courts to prohibit the use of information that the rightful owner did not undertake to protect.”\textsuperscript{77}

\textit{Siegel} found summary judgment against the firm was inappropriate in light of the facts that the “client list was maintained on a computer that was protected by a password. Hard copies of the list were stored within office filing cabinets which were sometimes locked . . . [and] employees [were informed] that the client list information was confidential and not to be removed from the office.”\textsuperscript{78} Because these measures do not differ from normal precautions taken by firms to protect the confidentiality of client information independent of trade secret concerns, \textit{Siegel} may be suggesting it should be a relatively simple matter for most firms to satisfy the secrecy condition for trade

\textsuperscript{73} Electro-Craft Corp. v. Controlled Motion, Inc., 332 N.W.2d 890, 899 (Minn. 1983); \textit{see also} ALCES & SEE, supra note 8, § 3.3.1, at 81 (1994) (“The term ‘readily ascertainable’ should be contrasted with the term ‘available after lengthy and expensive efforts.’”).

\textsuperscript{74} \textit{See}, e.g., Steenhoven v. Coll. Life Ins. Co. of Am., 458 N.E.2d 661 (Ind. Ct. App. 1984).

\textsuperscript{75} An example would be the client list of a firm that specializes in worker compensation claims.

\textsuperscript{76} \textit{Cf.} Fleming Sales Co. v. Bailey, 611 F. Supp. 507, 512 (N.D. Ill. 1985):

Of course the absence of specific confidentiality procedures (whether written or unwritten) does not itself negate the existence of ‘reasonable’ efforts to maintain secrecy—all the Act requires . . . So long as Fleming scrupulously limited distribution of customer list information to employees and outsiders whose access was necessary to Fleming’s successful pursuit of its business, it must be deemed to have satisfied the ‘reasonable efforts’ requirement, particularly if those given access to the information were also advised to preserve its confidentiality . . .

\textsuperscript{77} Dicks v. Jensen, 768 A.2d 1279, 1284 (Vt. 2001); \textit{see also} Surgidev Corp. v. Eye Tech., Inc., 828 F.2d 452, 455 (8th Cir. 1987) (observing that “reasonable” rather than “all conceivable” efforts are required and concluding that employer took adequate steps to protect the identity of ophthalmologists who were high-volume implanters of intraocular lenses even though it did not remind departing employees of secrecy of information).

\textsuperscript{78} Fred Siegel Co. v. Arter & Hadden, 707 N.E.2d 853, 862 (Ohio 1999).
secret status. Such a reading of the case may be errant, however, because the court was simply pointing to the existence of issues of material fact for summary judgment purposes.

An additional measure to preserve the secrecy of client lists not addressed in Siegel is to include in the partnership agreement or employment contract specific provisions according trade secret status to client information.79 At least as applied to information pertaining to clients with whom an attorney has a past relationship, however, the enforceability of such contractual provisions is dubious. Contracts restricting the right of a lawyer to practice law are void under prevailing standards of legal ethics.80 Contracts that seek to prevent the use of information by withdrawing lawyers are likely to share a common fate with contracts that impose economic penalties of withdrawing lawyers who compete. In their inability to enforce restrictive covenants, law firms are at a disadvantage when compared with other firms and may have comparatively greater difficulty in safeguarding confidential information.81

---

79. Such agreements may also bolster claims that information has been misappropriated. Cf. Nilssen v. Motorola, Inc., 963 F. Supp. 664, 679-80 (N.D. Ill. 1997) (“While an express confidentiality agreement may certainly suffice to define the duty of confidentiality necessary for action under Act § 2b(2)(b)(II), the existence of such an agreement is not a prerequisite to such an action. Rather a duty of confidentiality may be implied from the circumstances surrounding the parties’ relationship.”) (citations omitted).

80. See Model Rules of Prof’l Conduct R. 5.6 (2001).

81. In Fleming, 611 F. Supp. at 514, for example, the court noted that employers are not helpless in dealing with employees who leave and take confidential information because “[n]othing prevents such an employer from guarding its interests by a restrictive covenant.” The employer in the case was a manufacturer’s representative business. In some jurisdictions, enforcement of a restrictive covenant may be far more difficult than Fleming suggests. See, e.g., Dougherty, McKinnon & Luby, P.C. v. Greenwald, Denzik & Davis, P.C., 447 S.E.2d 94 (Ga. Ct. App. 1994) (refusing to enforce restrictive covenant against former employees of accounting firm in part because it was overbroad and extended to clients for whom they had not worked).

As this Article was going to press, a California intermediate appellate court rendered an opinion affirming a trial court’s finding that a law firm’s client list was a trade secret that had been misappropriated by withdrawing partners. See Reeves v. Hanlon, No. GC023679 (Cal. Ct. App. Feb. 20, 2003), available at http://www.courtinfo.ca.gov/opinions/documents/B151460.PDF (opinion certified for partial reporter publication; full opinion available at website, partial published opinion available at Reeves v. Hanlon, 106 Cal. App. 4th 433 (2003) (last visited Mar. 27, 2003) (on file with author). As to the firm’s efforts to protect the client list, the appellate court emphasized a number of factors:

[The trial court made elaborate findings regarding the Reeves firm’s measure to protect its client list from outsiders. . . . [The] firm was located in a building with a security guard; the attorney work area was closed to the public, absent an invitation, and monitored by receptionists; the client list was stored in a computer system requiring password access; confidentiality policies regarding client information were stated in the employee handbook, which was signed by all employees; and the importance of client confidentiality was discussed at employee meetings.

Id. at *23. Significantly, the factors emphasized in Reeves are standard practices for a large number of law firms.

As to the independent economic value of the client list, the court observed:
C. Misappropriation

The UTSA defines misappropriation by reference to “improper means” used to acquire a trade secret, which in turn is defined to include “breach or inducement of a breach of a duty to maintain secrecy.” For good reason, the definition rather neatly sidesteps questions of ownership of trade secrets. A trade secret right is not an exclusive property right tantamount to ownership of tangible property. If it were such an exclusive property right, difficult questions concerning the true owners of the information (the firm, the lawyers working with clients, or the client themselves) would demand attention. Instead, trade secret law protects information in the possession of one party from another party’s acquisition through improper means, including breach of a confidence.

One can with little difficulty find the requisite duty to maintain secrecy in many employer-employee relationships. Once again, however, law firms present difficult issues. Is there a difference between the duties of associates and partners in this regard? Can it be said that partners are under less of a duty by virtue of their co-ownership of the firm? Similar questions have been raised with respect to fiduciary duties of partners and associates, and the absence of adequate responses in that context suggests the issues raised may extend to questions of partner versus associate use of law firm information. Because trade secret protection need not turn on ownership of the in-

---

[T]he trial court found that the Reeves firm’s client list was confidential data "developed at great effort and expense over a period of 21 years of practice by advertising, client intake, representation and the good will developed therefrom . . . ." This finding is supported by Reeves's testimony, who indicated that his firm engaged in a specialized practice, and that his existing clients were a fertile source of new business. In view of this testimony, the trial court could properly conclude that the client list had independent economic value because it would allow a competitor to target its efforts to acquire clients.

Id. at *22-23.

82. UNIF. TRADE SECRETS ACT § 1(2) (1985).
83. Id. § 1(1).
84. Cf. E.I. Du Pont De Nemours Powder Co. v. Masland, 244 U.S. 100, 102 (1917), in which Justice Holmes noted:

The word 'property' as applied to trademarks and trade secrets is an unanalyzed expression of certain secondary consequences of the primary fact that the law makes some rudimentary requirements of good faith. Whether the plaintiffs have any valuable secret or not, the defendant knows the facts, whatever they are, through a special confidence that he accepted. The property may be denied, but the confidence cannot be. Therefore, the starting point for the present matter is not property or due process of law, but that the defendant stood in confidential relations with the plaintiffs or one of them.

The issue of whether trade secrets are property remains a matter of debate as well as artful expression. Compare ROGER M. MILGRAM, MILGRAM ON TRADE SECRETS § 1.01 (1981) ("Recognition of trade secrets as property is a basic conceptual step from which important aspects of trade secret law are derived."), with ALZER & SEE, supra note 8, § 3.4.1, at 84 ("A trade secret right is not profitably viewed as an exclusive property right to use the trade secret.").
formation or status of the persons who would use it in competition with the firm, however, there is little reason to draw a distinction between partners and associates in addressing trade secret issues.85

A more fundamental problem exists with broadly treating a law firm’s client information as a trade secret. Imposing on a particular lawyer the duty to maintain secrecy of the information necessarily requires the lawyer to refrain from using or disclosing the information following withdrawal from the firm. This, in turn, impedes the ability of the lawyer to communicate with and provide services to clients for whom the lawyer worked while at the firm. Such a result would serve as a direct restraint on competition with the firm and could undermine significantly the ability of clients to choose their lawyers.

Given the strength of the principle of client choice, there is little doubt that information will not be protected when clients effectively are asked to bear part of the costs of that protection. This is yet another instance of the development of law specially tailored for the legal profession reflecting the unique position of law firm clients among the larger class of consumers. In any event, perhaps it is concerns with impact on client choice of counsel that explains why in Siegel the associate’s taking of her rolodex became a nonissue on appeal.86

D. Remedies

The UTSA provides for injunctive relief87 and damages88 as remedies for a misappropriation of trade secrets. In addition, attorney’s fees may be awarded if a claim of misappropriation is made in bad faith or in cases of willful and malicious misappropriation.89

The UTSA’s injunctive relief provisions include a variation allowing “[i]n exceptional circumstances, an injunction may condition future use upon payment of a reasonable royalty.”90 Use of either traditional injunctive relief or the royalty option in cases of client information, however, will raise the usual concerns on frustration of client’s choice of counsel.91 Royalty payments would add the additional

---

85. Cf. Revised Unif. P’ship Act § 401(g) (1997) (“A partner may use or possess partnership property only on behalf of the partnership.”).
88. Id. § 3.
89. Id. § 4.
90. Id. § 2(b). Exceptional circumstances include “a material and prejudicial change of position prior to acquiring knowledge or reason to know of misappropriation that renders a prohibitive injunction inequitable.” Id.
91. See supra notes 39 to 46 and accompanying text.
V. TRADE SECRET PROTECTIONS AND FIDUCIARY DUTIES COMPARED

Even if firm information falls short of trade secret status, its use at the time of a lawyer’s withdrawal or thereafter may be restricted by the fiduciary duties under which partners operate. In *Gibbs v. Breed, Abbott & Morgan,* for example, the court found the surreptitious recruiting of associates by partners planning to withdraw from the firm together with the use of information concerning the associates compensation, billing rates, and billable hours constituted an egregious breach of fiduciary duty they owed their firm.

Sheehan’s disclosure of confidential BAM data to even one firm was a direct breach of his duty of loyalty to his partners. Because the memo gave Chadbourne confidential BAM employment data as well as other information reflecting BAM’s valuation of each employee, Chadbourne was made privy to information calculated to give it an unfair advantage in recruiting certain employees.

The information involved in *Gibbs* concerned associates rather than clients, but the case does illustrate that fiduciary duty and trade secret law may operate in similar ways to protect information a law firm regards as confidential.

Although similar, the relative protections accorded by the two doctrines will vary from case to case. Consider, for example, the associate in *Siegel,* the trade secret decision discussed earlier. On remand, the critical issue for trade secret purposes is whether the firm took sufficient steps to protect the secrecy of the information. From the fiduciary perspective, however, the measures the firm took to protect the information are not an essential element in defining the duties of the associate while she was at the firm. Instead, if the associate knew the firm regarded the client list as confidential and nevertheless removed it from the firm, she could have breached her duties as a fiduciary through the mere taking of firm property. Although efforts to preserve secrecy and misappropriation are distinct requirements for trade secret protection, a fiduciary claim may be developed on the basis of misappropriation (i.e., improper taking) alone. Firms

---

92. *Model Code of Prof’l Responsibility DR 2-107(A)* (1982) (allowing fee splitting only when the client consents, the division is proportionate to the services performed and responsibilities assumed, and the total fees are not unreasonable). An exception allows fee splitting when payments are made to a former partner or associate pursuant to a separation agreement. See id. DR 2-107(B). See generally *Hillman,* supra note 1, § 4.6.2.


94. *Id.* at 583.

95. *Id.*

96. See Fred Siegel Co. v. Arter & Hadden, 707 N.E.2d 853 (Ohio 1999).
unable to satisfy the secrecy threshold for trade secret protection thus may still be able to assert claims based on fiduciary duties.

In other respects, however, trade secret protection may be broader than fiduciary duties. Along this line, a rather ancient agency doctrine addresses the duties of the fiduciary with respect to post-withdrawal use of employer information by allowing the employee to take and use only the information he has committed to memory. The Restatement (Second) of Agency describes the post-withdrawal duties of the agent as

\[
\text{not to use . . . in competition with the principal or to his injury, trade secrets, written lists of names, or other similar confidential matters given to him only for the principal's use . . . . The agent is entitled to use general information . . . retained in his memory, if not acquired in violation of his duty as agent[.]}^{97}
\]

Although the statement begs the question to some extent by its references both to trade secrets and to information acquired in violation of a duty, it does make the distinction for fiduciary duty purposes between information taken in documentary form and information that has been memorized. A few courts have concluded memorized information cannot be a trade secret, but UTSA seemingly rejects such a limitation by extending trade secret protection to information. The matter is not free from doubt, but at the very least it

---

97. Restatement (Second) of Agency § 396(a) (1958).
98. Cf. Fleming Sales Co. v. Bailey, 611 F. Supp. 507, 513-14 (N.D. Ill. 1985) (noting that a former employee may use skills and knowledge acquired in the course of employment, provided that written records of compilations are not taken:

Any other rule would force a departing employee to perform a prefrontal lobotomy on himself or herself. It would disserve the free market goal of maximizing available resources to foster competition . . . . [I]t would not strike a proper balance between the purposes of trade secrets law and the strong policy in favor of fair and vigorous business competition.).

99. For example, in Avnet, Inc. v. Wyle Lab., Inc., 437 S.E.2d 302 (Ga. 1993), the Georgia Supreme Court described the common law distinction between lists containing customer information and a former employee's knowledge of customer information, noted the latter was not a trade secret, and concluded the Georgia Trade Secrets Act did not change this result. The state statute was subsequently amended to cover information without regard to form, but it has been held that the amendment applies to tangible forms of information and not memorized information. See Amerigas Propane, L.P. v. T-Bo Propane, Inc., 972 F. Supp. 685, 697-98 (S.D. Ga. 1997).

100. Unif. Trade Secrets Act § 1(4) (1985); see also Vigoro Indus., Inc. v. Cleveland Chem. Co., 866 F. Supp. 1150, 1162 (E.D. Ark. 1994) ("[T]he distinction between information which is written down and that which is memorized has little materiality under Arkansas law. The critical issue is whether the information, whether written or memorized, is entitled to protection as a trade secret."). Aff'd in part and remanded in part, 82 F.3d 785 (8th. Cir. 1996); cf. Cont. Dynamics Corp. v. Kantor, 408 N.Y.S.2d 801, 802 (N.Y. App. Div. 1978).

Since the names of potential customers were readily ascertainable from public sources, the defendants' solicitation of the plaintiff's customers from casual memory is not a legally cognizable wrong. However, where customer lists do not rise to the level of trade secrets, an employee's 'physical taking' or 'studied
would seem that use of information committed to memory distinction is less likely to be less significant for trade secret purposes than it is for the fiduciary duty analysis.

Another way in which trade secret rights may accord broader protections than fiduciary duties is that the status of trade secret information is not affected by the fact that the individual who has the information has left the firm. Fiduciary duties of partners, in contrast, diminish substantially when a partner withdraws from a firm. The same may be said of employees, such as associates, who are not partners of the firm. To be sure, fiduciary duties may be implicated if the information was acquired by improper means while the lawyer was still a member of the firm, but in such a case relief is sought for actions taken prior to departure rather than post-withdrawal competition with the firm.

VI. A FINAL NOTE: THE ROLE OF CONTRACTUAL REMEDIES

By focusing on the potential trade secret status of client lists, this Article has emphasized protections imposed by law. Information may also be protected by contract. For this reason, the UTSA makes clear that the statutory framework for trade secrets does not displace contractual remedies. The Comment adds:

This Act . . . is not a comprehensive statement of civil remedies. It applies to a duty to protect competitively significant secret information that is imposed by law. It does not apply to a duty voluntarily assumed through an express or implied-in-fact contract. The enforceability of covenants not to disclose trade secrets and covenants not to compete that are intended to protect trade secrets, for example, is governed by other law.

_id._ (citation omitted).

101. See, e.g., Gibbs v. Breed, Abbott & Morgan, 710 N.Y.S.2d 578, 590 (N.Y. App. Div. 2000) (Saxe, J., concurring in part and dissenting in part) (“I conclude that the information plaintiffs disclosed to Chadbourne should not be treated as a ‘trade secret’ or ‘confidential matter’ since if it were, a departing attorney might have a continuing obligation not to disclose it . . . .”).

102. Under the Revised Uniform Partnership Act (RUPA), for example, upon dissociation a partner’s duty to refrain from competing with the partnership terminates, although as to matters that arose before the dissociation the partner continues to have duties to account and to refrain from dealing with the partnership as or on behalf of an adverse party. See _REVISED UNIF. P’SHP ACT_ § 603(b) (1997).

103. Associates, as employees, have fiduciary duties to their firms.

104. Although the time at which the individual secured the information is an issue under the trade secret analysis, the protections accorded by this area of law are not limited to relief from actions taken while the partner was at the firm.

105. _UNIF. TRADE SECRETS ACT_ § 7(b)(1).

106. _Id._ cmt.
This allowance of private ordering as a means of defining rights in information is consistent with a view of fiduciary duties as “unspoken expectations” that, by virtue of their default character, are readily supplanted by explicit agreements among the parties.107

There is an intuitive appeal to deferring to private ordering to establish rights in information.108 This is particularly true when the parties are lawyers and presumably capable of bargaining on the subject. The difficulty with deferring to private ordering in establishing rights to information concerning law firm clients is that the results of the bargaining may restrict use of client information in ways that undermine the ability of clients to freely select their law firms. As has been discussed in the context of contractual measures as evidence of efforts to maintain secrecy,109 the principle of client choice operates to restrict private ordering among law partners on such basic matters as restrictive covenants and contractual disincentives to competition.

Whether the principle of client choice is generally in need of reexamination is beyond the scope of this Article. As long as it remains a fundamental tenet of legal ethics, the role of private ordering in restricting rights to client information is likely to be extremely limited.

108. Cf. Gibbs v. Breed, Abbott & Morgan, 710 N.Y.S.2d 578, 582 (N.Y. App. Div. 2000) (finding that withdrawing partners breached no duties to the firm in taking desk copies of recent correspondence, apparently without regard to client authorizations: “These [files] were comprised of duplicates of material maintained in individual client files, the partnership agreement was silent as to these documents, and removal was apparently common practice for departing attorneys.”) (emphasis added).
109. See supra text accompanying notes 79-81.
DISABUSING THE DEFINITION OF DOMESTIC ABUSE: HOW WOMEN BATTER MEN AND THE ROLE OF THE FEMINIST STATE

LINDA KELLY*

INTRODUCTION .......................................................................................................................... 792
I. THE STUDY OF INTIMATE VIOLENCE .................................................................................. 794
   A. The Early Studies ............................................................................................................... 794
      1. Use ............................................................................................................................... 795
      2. Frequency .................................................................................................................... 797
      3. Severity ....................................................................................................................... 798
   B. Recent Findings .............................................................................................................. 799
II. THE REACTIONS .................................................................................................................. 800
   A. Denying Female Violence ............................................................................................... 801
      1. By Woozles and Scare Tactics ..................................................................................... 801
      2. By Methodological Critique ...................................................................................... 802
         (a) CTS Challenges ...................................................................................................... 802
         (b) Gender Concerns ................................................................................................... 804
   B. Defending Female Violence ............................................................................................ 805
      1. Quantitative Criticisms ............................................................................................... 805
      2. Qualitative Criticisms ................................................................................................. 808
   C. Minimizing Female Violence ......................................................................................... 809
   D. Responding to the Critics: Why Female Violence Must Be Examined ....................... 810
III. THE EXPLANATIONS .......................................................................................................... 814
   A. Female Violence: A Theoretical Threat ......................................................................... 814
      1. The Development of the Patriarchal Definition of Domestic Violence .................. 814
      2. Challenging Domestic Violence, Challenging Feminism ......................................... 817
         (a) Through Theory ..................................................................................................... 817
         (b) Through Methodology .......................................................................................... 818
   B. Female Violence: A Practical Threat ............................................................................. 820
      1. Financial Concerns ...................................................................................................... 820
      2. Marketing Dilemmas ................................................................................................. 821
      3. Vindictive Possibilities ............................................................................................... 822
IV. LEGAL IMPLICATIONS ...................................................................................................... 822
   A. Academic ....................................................................................................................... 822
   B. The Feminist State of Domestic Violence ...................................................................... 824
      1. Arrest ........................................................................................................................ 824
         (a) The Development of Arrest Policies and Theories: The Minneapolis Experiment and U.S. Attorney General’s Report ................................................................. 824
         (b) The Gendered Promotion of Arrest ....................................................................... 830
         (c) The Statistics ......................................................................................................... 830
         (d) Thurman v. City of Torrington ............................................................................. 832
      2. Prosecution ................................................................................................................ 836
         (a) The Development of Prosecution Policies and Theories ....................................... 836
         (b) The Gendered Promotion of Prosecution ............................................................ 839

* Professor of Law, Indiana University School of Law, Indianapolis. B.A., 1988, University of Virginia; J.D., 1992, University of Virginia. Special thanks to John Hill, who makes everything possible.
INTRODUCTION

Domestic abuse. The term immediately conjures up images nationally spread through such highly publicized events as the murder trial of O.J. Simpson and the saga of John and Lorena Bobbit, hit tunes like Tracy Chapman’s *Behind the Wall*, and movies like Julia Robert’s *Sleeping with the Enemy* or Farrah Fawcett’s *The Burning Bed*. Everyone can also tell a more local story about domestic violence, be it one carried in a hometown newspaper or known about the neighbors. After a long history of hiding domestic violence behind closed bedroom doors, everyone now knows all about the existence and prevalence of domestic violence. Or do we? The images we associate with domestic violence depict the male as batterer and the female as victim. Yet, despite the critical importance of first acknowledging and then eradicating the male abuse of women, an equally important but untold story remains. Women can be batterers. Men can be victims.

Over the last twenty-five years, leading sociologists have repeatedly found that men and women commit violence at similar rates. The 1977 assertion that “the phenomenon of husband battering” is as prevalent as wife abuse is confirmed by nationally representative studies, such as the Family Violence Surveys, as well as by numerous

1. TRACY CHAPMAN, *Behind the Wall*, on TRACY CHAPMAN (Electra/Asylum 1983).
4. On the historic privacy of domestic violence and the successful efforts to make domestic violence a public issue, see infra note 163 and accompanying text.
5. Throughout this Article, my use of the terms domestic, intimate, and spousal violence or abuse are used to refer to men and women who have experienced physical violence at the hands of their partner. These violent couples may be married or not married, living together or apart. Certainly other forms of domestic violence exist, such as child abuse, elder abuse, homosexual abuse, and sibling abuse. However, purely as a shorthand measure, the terms are relied upon here exclusively to refer to violence between heterosexual partners. For recognition of the other forms of domestic violence, see infra notes 297-300 and accompanying text.
other sources. However, despite the wealth and diversity of the sociological research and the consistency of the findings, female violence is not recognized within the extensive legal literature on domestic violence. Instead, the literature consistently suggests that only men commit domestic violence. Either explicitly, or more often implicitly, through the failure to address the subject in any objective manner, female violence is denied, defended and minimized.

How is it that our general legal understanding of domestic violence as defined by the male abuse of women is so squarely contradicted by the empirical reality? Honestly answering this question requires tracing the history of both the theory and practice of domestic violence law. Undertaking such an exploration, one quickly finds that the “discovery” of domestic violence is rooted in the essential feminist tenet that society is controlled by an all-encompassing patriarchal structure. This fundamental feminist understanding of domestic violence has far-reaching implications. By dismissing the possibility of female violence, the framework of legal programs and social norms is narrowly shaped to respond only to the male abuse of women. Female batterers cannot be recognized. Male victims cannot be treated. If we are to truly address the phenomenon of domestic violence, the legal response to domestic violence and the biases which underlie it must be challenged.

Through an open discussion of domestic abuse, Part I of this Article endeavors to expose the fact that domestic violence is committed by women. In so doing, I introduce to legal literature the first extensive account of the Family Violence Surveys and various other studies completed over the last twenty-five years which have repeatedly found that men and women commit violence at similar rates. After exploring the tendency to deny, defend or minimize the violence of women in Part II and then arguing that female violence must be addressed, I assert in Part III that today’s refusal to react is a product

7. For a discussion of the Family Violence Surveys and related work on intimate violence, see infra notes 14-35 and accompanying text.
9. In relation to the thousands of articles on domestic violence, a LEXIS/NEXIS search found that the Family Violence Surveys are cited in only twenty-six articles. None of these articles gave any meaningful critical examination of the Family Violence Surveys. In fact, only one article gave anything more than a cursory footnote reference to the surveys’ coverage of female intimate violence. The one article which did contain any substantive discussion of the National Family Violence Surveys did so only in order to compare it to the National Violence Against Women Survey (which was completed by the article’s author) and the National Crime Victimization Survey. Patricia Tjaden, Extent and Nature of Intimate Partner Violence as Measured by the National Violence Against Women Survey, 47 LOY. L. REV. 41, 50-51 (2001). For further discussion of the National Violence Against Women Survey and the National Crime Victimization Survey, see infra notes 59, 158 and accompanying text.
of the feminist control over the issue of domestic violence. Female violence presents both a threat to feminist theory as well as to the practice of domestic violence law. Notwithstanding such concerns, today’s myopic understanding of domestic violence has serious implications. Limiting this examination to the criminal justice system, Part IV considers how the feminist definition of domestic violence has skewed arrest and prosecution philosophies, resulting primarily in having only male batterers criminally pursued. The Part also reviews how rehabilitative programs are geared toward treating domestic violence as the byproduct of a patriarchal society, thereby only producing programs which address male violence. Similarly, the services for domestic violence victims, in particular, the availability of shelters, have also been shaped by the feminist definition of domestic violence. In conclusion, Part V calls for challenging the existing gendered definition of domestic violence and thereby demands changing our norms and institutions so that we may honestly work toward addressing and eliminating domestic violence.

Having set out the prerequisite roadmap, it is important to emphasize and re-emphasize what this Article is not. It is not an attempt to minimize or question in any respect the magnitude or seriousness of the intimate violence inflicted on women by men. Rather, it is an attempt to look candidly at the possibility of the violence inflicted on men by women, look at the reaction to such revelations, discuss the explanations for and implications of such responses, and call for a more honest, open discourse on domestic violence.

I. THE STUDY OF INTIMATE VIOLENCE

The reality of husband abuse was first nationally exposed with the release of several studies during the 1970s. The findings of these early studies were startling. Not only were women engaging in intimate violence, but their propensity for such acts as compared to male abuse of intimate female partners was similar in a number of ways. Such radical reports ignited a controversy which continues today. Husband battering continues to be heatedly denied, defended and minimized. However, when the early studies of husband abuse, its confirmation by a variety of later research, and the attacks on the projects are examined together in a critical fashion, a simple truth remains. Women batter.

A. The Early Studies

In her 1977 work entitled The Battered Husband Syndrome, sociologist Suzanne Steinmetz was among the first to bring public and
academic attention to the “phenomenon of husband battering.” 10
While the article was brief, it was explosive. Professor Steinmetz
studied the use of physical violence by husbands and wives in five
independent surveys conducted by various family violence research
teams. 11 Such surveys relied primarily upon the “Conflicts Tactic
Scales” (CTS). Developed in 1971 to measure family violence, CTS
breaks physical force and violence into eight categories ranging from
(1) throwing things; (2) pushing, shoving or grabbing; (3) hitting or
slapping; (4) kicking, biting or hitting with a fist; (5) hitting or trying
to hit with something; (6) beating up; (7) threatening with a knife or
gun; and finally, (8) using a knife or gun. 12

1. Use

In four out of the five studies reported, Professor Steinmetz found
that husbands and wives are roughly equal in their use of any form
of physical violence. While data from three of the reported studies
finding gender parity was based on her individual work, 13 the fourth
set of data was from the landmark National Family Violence Survey

10. Steinmetz, supra note 6, at 499.
11. In addition to the empirical data, Steinmetz’s work examines spousal violence by
comparing the historic treatment of husband and wife abuse and its portrayal in comic
strips. Id. at 499-501. It is from relying on such a variety of sources that Steinmetz con-
cludes the “phenomenon of husband battering” exists and can no longer be ignored. Id. at
499, 503-08.
12. In all, CTS is comprised of eighteen items, divided into three approaches to resolv-
ing conflict: (1) rational discussion, (2) verbal and non-verbal hostile expressions (which
are not physical) and (3) the use of physical force and violence. It is this last stage of con-
flict resolution, the physical violence continuum, which was, and continues to be, the most
controversial. For a more detailed explanation of the development and use of CTS, see
Murray A. Straus et al., Behind Closed Doors: Violence in the American Family 26-
28, 253-66 (1980). For a further discussion of CTS and its critics, see infra notes 52-59 and
accompanying text.

It should be noted that Steinmetz’s study comparison completely omitted the CTS physi-
cal violence category of “beating up.” Further, several of the studies compared by
Steinmetz did not ask questions specifically addressing kicking, threatening with a knife
and gun, or use of a knife or gun. For a complete chart of the five studies compared by
Steinmetz, see Steinmetz, supra note 6, at 502.
13. In three separate studies directed by Professor Steinmetz—one involving a ran-
don sample of couples in New Castle, Delaware; the second based upon a broad based non-
representative U.S. sample; and the third relying upon a sample of Canadian college stu-
dents—she found that husband and wives were roughly equal in their use of any form of
physical violence. In the Delaware study, 47% of husbands were reported to have used
some form of physical violence against their wives at some point during the marriage. 43%
of wives also were reported to have used violence (on at least one occasion) during the mar-
rriage. From the Canadian sample of fifty-two families, 23% of husbands resorted to vio-
lence, compared to 21% of wives within the sample. Similarly, in the broad-based, non-
representative U.S. sample of ninety-four couples, 32% of husbands and 28% of wives were
reported to use violence during their marriages. While the percentages vary between stud-
ies, the rates of husband and wife violence within each study are similar. Steinmetz, supra
note 6, at 501-03. For Steinmetz’s work dedicated to the Delaware study, see Suzanne K.
Steinmetz, The Cycle of Violence: Assertive, Aggressive, and Abusive Family
of 1975. Directed by sociologists Richard Gelles, Murray A. Straus and Suzanne Steinmetz through the support of the Family Violence Research Program of the University of New Hampshire, the team was responsible for conducting the first national survey dedicated to family violence. Focusing specifically on spousal violence and child abuse, a nationally representative sample of 2143 “intact couples” was selected. From this sample one adult in each couple was interviewed in person and questioned about family conflicts which had occurred over the last twelve months and conflicts which had occurred at any time during the marriage or parent-child relationship. The Family Violence Research Laboratory’s national survey proved consistent with Steinmetz’s other work. Reporting the rate of marital violence committed only in the year 1975, the data reflected gender parity—12% of husbands and 12% of wives had used violence during the year 1975 against their mates.

The final, fifth comparison study did suggest husbands used violence at a greater rate than wives. However, unlike the other four surveys, such results were recognized to be skewed as half of the

14. Steinmetz, supra note 6, at 501-03. Many books and papers devoted exclusively to the 1975 Family Violence Survey, as well as comparing it to the later surveys, have been published. For one of the first published papers dedicated to the 1975 survey, see Murray A. Straus, Wife Beating: How Common and Why?, 2 VICTIMOLOGY: AN INT’L J. 443 (1977). For the first book devoted to the 1975 survey and authored by its leading sociologists, see STRAUS ET AL., supra note 12. For later works comparing the 1975 survey with later, similar surveys conducted in 1985 and 1992 by the same individuals, see RICHARD J. GELLES, INTIMATE VIOLENCE IN FAMILIES (3d ed. 1997) (comparing results of the 1975, 1985 and 1992 surveys) [hereinafter INTIMATE VIOLENCE IN FAMILIES]; RICHARD J. GELLES & MURRAY A. STRAUS, INTIMATE VIOLENCE (1988) (comparing the 1975 and 1985 surveys for a broad, public audience) [hereinafter INTIMATE VIOLENCE]; MURRAY A. STRAUS & RICHARD J. GELLES, PHYSICAL VIOLENCE IN AMERICAN FAMILIES: RISK FACTORS AND ADAPTATIONS TO VIOLENCE IN 8,145 FAMILIES (1989) (comparing the 1975 & 1985 surveys for an academic audience)[hereinafter PHYSICAL VIOLENCE]. For further discussion of these three leading surveys, see infra notes 15-18 and accompanying text.

15. In order to qualify as an intact couple, the couple was required to consist of a man and woman currently living together. While marriage was not required, the survey characterizes all intercouple violence as spousal, without drawing any marital/non-marital distinctions. Of the 2143 families surveyed, 1146 families met the study’s criterion of having children between the ages of 3 and 17. STRAUS ET AL., supra note 12, at 24-26. For criticisms by the Family Violence Research Laboratory of their sampling methodology and efforts to improve it in later studies, see id. at 25-26, PHYSICAL VIOLENCE, supra note 14, at 9-14, 49-73.

16. Of the 2143 families, 960 men and 1183 women were interviewed. STRAUS ET AL., supra note 12, at 24.

17. While the 12% figure is drastically lower than the percentages provided in the other surveys reported in Steinmetz’s work, the 12% figure represents violence during a one-year reporting figure, while the other surveys report violence occurring at any time during the marriage. Steinmetz, supra note 6, at 502. For a summary of Steinmetz’s three projects, see discussion supra Part I.A.

18. Reporting on the use of any violence during the course of a marriage, the study found 47% of husbands, as compared to only 32% of wives, had used violence on at least one occasion. RICHARD J. GELLES, THE VIOLENT HOME 51-52 (1972) (reporting survey results).
survey participants were selected from police blotter or social service agency reports of domestic violence which were recognized to disproportionately reflect male-perpetrated domestic violence.¹⁹

2. Frequency

The comparable use of physical violence by spouses marks only the beginning of the similarities evidenced by these comprehensive first studies of intimate violence. Arguably, husband abuse can be discounted in comparison to wife abuse if women are found to utilize physical violence against their spouses in a much more sparing fashion than men. However, the various surveys consistently reported that women not only use violence at rates similar to men, but that women match, and often exceed, husbands in the frequency with which they engage in violent behavior.²⁰

Later studies confirm the findings of these early, controversial works. In 1985, the Family Research Laboratory conducted another national study.²¹ Reaching more than 6000 families through tele-

¹⁹. See Steinmetz, supra note 6, at 503. For further discussion of the ongoing bias of recognizing male abuse and ignoring female abuse evident in cultural norms and in the policies of legal and social service agencies, see infra note 90 and accompanying text. See also infra Part IV.B.

²⁰. Counting the number of reported acts by husbands and wives during the course of their marriages, each of Steinmetz’s three surveys reported wives engaging in more acts of physical violence than husbands. Looking only at the frequency of violence in couples in which one or more acts of violence had occurred, the reported frequency of violence for wives versus husbands was as follows: in the Delaware survey, 4.04 versus 3.52; in the Canadian survey, 7.82 versus 6.00; in the national non-representative survey, 7.00 versus 6.60. Steinmetz, supra note 6, at 503.

Similarly, the Family Violence Research Survey of 1975 also found that women on average committed more acts of violence in the reporting year of 1975 than husbands. Again looking only at the frequency of violence in couples in which one or more acts of violence had occurred during the year 1975, Straus’s report reflects wives committed an average of 10.3 physically violent acts against husbands, while husbands averaged 8.8 acts against their wives. Id. at 503. However, in order not to “overstate the case,” Straus states a preference for relying upon the median as better reflecting the typical frequency of violent behavior in couples for which violence had occurred during 1975. Straus, supra note 14, at 445. This median reflects husbands’ medium frequency of violent acts for the year was 2.5 acts, while wives’ median frequency of violent behavior was reported slightly higher, at 3.0 acts for 1975. Id. For a full narrative and statistical frequency discussion of the Family Violence Research Survey of 1975, see Straus et al., supra note 12, at 41-43; Straus, supra note 14.

While the fifth survey in Steinmetz’s study (which was completed by Richard J. Gelles) reported that men relied upon physical violence more frequently than women in their intimate relationships, such results need to be considered critically in light of the frequency rates consistently reported by the other survey and perhaps more importantly, the skewed nature of the sample gathered to be surveyed. In his 1974 survey, Gelles found, for example, that 11% of husbands and 5% of wives engaged in violent acts between two and six times a year. Steinmetz, supra note 6, at 503. For further discussion of the skewed nature of samples used in various domestic violence studies, see infra note 59 and accompanying text.

²¹. There is a wealth of literature which relies upon the 1985 survey and compares its findings to the earlier surveys (particularly its predecessor survey of 1975). Two critical
phone interviews, the research team sought to ensure its data was representative by creating a sample group over twice the size of the 1975 survey as well as “oversamples” of certain minority groups and populations in certain states who were at risk of underrepresentation. Relying again on the Conflict Tactics Scale, the 1985 study confirmed the use and frequency of physical violence by men and women.

3. Severity

While the similarity of rates of physical violence by wives and husbands presented by the various surveys is revealing, such data is not sufficient to make an accurate comparison of the violent nature of wives and husbands. As the definition of “physical violence” used in the various CTS-based studies ranges from “throwing something” to “using a knife or gun,” wives arguably could compare to husbands in use and frequency of violent behavior, but not in the severity of the type of violence employed.

Some differences per type of violence utilized by each sex are certainly evident. Women were found to be twice as likely to throw something at their husbands. Wives were also more likely than husbands to kick, bite and punch. They were also more likely to hit, or try to hit, their spouses with something and more likely to threaten their spouses with a knife or gun. Husbands, on the other hand, rated higher in the four categories of pushing, grabbing and

books were written by the surveys’ authors and are dedicated to the subject: INTIMATE VIOLENCE, supra note 14 (written for wide public distribution); and PHYSICAL VIOLENCE, supra note 14 (written for the academic community). For other discussions of these two leading surveys, see supra notes 14-18, 20 and accompanying text, and infra notes 66-68 and accompanying text.

22. In total, the 1985 “resurvey” by the Family Research Laboratory interviewed 6,002 households, with 4,032 households selected in proportion to household distribution per state and three “oversamples” consisting of: (1) an oversample of 958 households in twenty-five states in order to assure 100 interviews in thirty-six key states; (2) an oversample of 508 black households; and (3) an oversample of 516 Hispanic households. In order to accommodate the increase in interviews from the 1975 in-person survey, interviews in the 1985 study were completed via telephone. For further discussion of the methodologies used in the second national survey, see INTIMATE VIOLENCE, supra note 14, at 207-12. For a scholarly analysis and comparison of the 1975 and 1985 methodologies, see PHYSICAL VIOLENCE, supra note 14, at 17-28. For further discussion of the 1975 survey, see supra notes 14-18 and accompanying text.

23. PHYSICAL VIOLENCE, supra note 14, at 95-99.

24. During 1975, 5% of wives threw something at their husbands, compared to 3% of husbands. STRAUS ET AL., supra note 12, at 37-38; Steinmetz, supra note 6, at 502.

25. During 1975, 3% of wives kicked, bit or slapped, as compared to 2% of husbands. STRAUS ET AL., supra note 12, at 37-38; Steinmetz, supra note 6, at 502.

26. During 1975, 3% of wives hit, or tried to hit their spouse with something, as compared to 2% of husbands. Six percent of wives threatened their husbands with a knife or gun, as compared to 0.4% of husbands. STRAUS ET AL., supra note 12, at 37-38; Steinmetz, supra note 6, at 502.
shoving;27 slapping or hitting;28 beating up;29 and actually using a knife or gun.30 Yet, such per category differences did not evidence that men were unquestionably more prone to acts of severe domestic violence than women. Combining the data collected on the last five categories of physical violence to create a “Severe Violence Index,” wives were found to engage in more severe acts of violence than husbands.31 Taking the frequency of severely violent behavior into account does not mitigate these findings. Wives show a pattern of severely violent behavior statistically comparable to husbands.32 Consistent with this “over-all similarity” found in the 1975 survey,33 other early reports also found that husbands and wives show “equal potential” for intimate violence and that they “initiate[d] similar acts of violence.”34

B. Recent Findings

These staggering findings on the use, frequency and severity of violence similarly perpetrated by husbands and wives kept sociologists committed to the study of family violence. In 1992, members of the Family Violence Research Laboratory completed yet another nationally representative survey.35 The study reaffirmed that wives en-

27. During 1975, 11% of husbands pushed, grabbed or shoved as compared to 8% of wives. STRAUS ET AL., supra note 12, at 37-38; Steinmetz, supra note 6, at 502.
29. Id.
30. As reported in 1975, husbands used a knife or gun on their wives in 0.3% of cases, as compared to 0.2% of wives committing similar behavior. STRAUS ET AL., supra note 12, at 37-38; Steinmetz, supra note 6, at 502.
31. As reported in 1975, 3.8% of husbands engaged in “severe violence,” as compared to 4.6% of wives. STRAUS ET AL., supra note 12, at 39-41. The last five CTS categories are: kicking, biting or hitting with a fist; hitting or trying to hit with something; beating up; and threatening with or using a knife or gun. For further discussion of CTS, see supra note 12 and accompanying text.
32. Looking just at couples reporting spouse “beating,” the survey found that in about one-third of the cases there had been only one incident of severe violence and that wives and husbands were equal in committing such “one-time” severe violence. STRAUS ET AL., supra note 12, at 41. Comparing households where two acts of severe violence occurred, 20% of severely violent husbands and 12.5% of severely violent wives committed two severe acts. Id. Forty-seven percent of “wife-beating” husbands beat their wives three or more times in 1975 compared to 53% of wives who beat their husbands three or more times. Id. at 41-42. Arguably, ending the comparison at three times or more is not a sufficient basis for comparison because within such a category a spouse of one sex who engages in three acts of severe violence is measured equally against a spouse of the opposite sex who engages in many more acts of severe violence. For a discussion of such criticisms of the Severe Violence Index, see infra Part II.A.2(a).
33. STRAUS ET AL., supra note 12, at 37-38. The specific act comparison chart is reproduced in Steinmetz’s work without the “beating up” item. See Steinmetz, supra note 6, at 502.
34. Steinmetz, supra note 6, at 505.
35. The 1992 survey was completed with a nationally representative sample of 1970 families and was conducted through telephone interviews by Glenda Kaufman Kantor. The results of the 1992 survey and its comparison to the two National Family Violence Surveys
gaged in intimate violence at rates comparable to husbands. However, a comparison of the 1975, 1985 and 1992 studies also reveals an important trend. Despite the finding that husbands and wives were roughly equal in terms of the percentages of spouses who engaged in any act of violence, the gap in the use of severe violence by husbands and wives had widened. In comparing the 1975 and 1985 results, researchers observed that while the male use of severe physical violence had declined 21%, the female use of such violence remained virtually constant. In the 1992 results, researchers again found that while severe assaults by wives remained fairly steady, the rate of severe abuse perpetrated by husbands decreased between 1985 and 1992 by almost 37%. In overall comparison to the constant rate of husband abuse, the combination of such significant decreases in wife-beating represented a 50% drop between 1975 and 1992.

II. THE REACTIONS

While the study of husband abuse by the Family Violence Research Laboratory, its associates, and others engaged in similar projects has received some praise, such support has been completely overshadowed by the degree and extent of criticism levied against it.


36. For a discussion of such a study, see supra note 35.

37. For an earlier discussion of the lower reporting of severe violence by men than by women as reported in 1975, see supra notes 14-18 and accompanying text.

38. In reporting on the use of “severe violence” (kicking, hitting or trying to hit with something; beating up; and threatening or using a knife or gun) as measured by CTS, the researchers reported that husband-to-wife use of “severe violence” decreased from thirty-eight husbands per 1000 couples in 1975 to thirty husbands per 1000 couples in 1985. By comparison, wives’ use of “severe violence” remained constant, with women reported to use “severe violence” at the rate of forty-six women per 1000 couples in 1975 and forty-four women per 1000 couples in 1985. INTIMATE VIOLENCE, supra note 14, at 250-51. For a discussion of the 1985 statistics and their comparison to the 1975 statistics, see id. at 108-15; PHYSICAL VIOLENCE, supra note 14, at 118-21, 529-34. For discussion of the “Severe Violence Index” see supra note 31 and accompanying text.

39. While the 1985 rate of wife-beating was thirty per 1000 couples in 1985, the 1992 rate was nineteen per 1000 couples. FORD, supra note 35, at 11-13; Gelles, supra note 35, at 797.

40. The 1992 study also included a comparison of violence by gender through a “minor assault” index, which was comprised of the less severe pushing, grabbing, shoving and slapping categories. Minor assaults perpetrated by husbands appeared to decrease between 1975 and 1992, minor assaults perpetrated by wives increased. See FORD, supra note 35, at 13; see also Gelles, supra note 35, at 797. For a discussion of the gender reporting differences found in comparing the studies, see infra Part II.A.2.(b).
Criticisms have ranged from personally attacking the researchers, to more academic efforts directed at attacking the work itself by denying the validity of the reports, to an outright defense of the violent behavior of women or otherwise minimizing its significance. Yet, while the nature of the criticisms has differed, they have invariably all been vehement.

A. Denying Female Violence

1. By Woozles and Scare Tactics

In the public arena, the media’s reaction to the reports of husband abuse was an overreaction. Estimates that two million males in the United States were subject to domestic violence quickly became inflated by the newspapers to twelve million.\(^{41}\) As the Family Violence Research team quipped, “Woozles, it seems, tend to multiply in direct proportion to the degree of controversy associated with a story.”\(^{42}\)

Other reactions could not be responded to so lightly. Perhaps the most physically and personally intimidating behavior was directed at Suzanne Steinmetz, who had first brought the issue to the public’s attention.\(^{43}\) Steinmetz appeared on such shows as the Today Show and Phil Donahue.\(^{44}\) Her work was reported in various newspapers and magazines, including a full-page story in Time magazine.\(^{45}\) Yet, while Steinmetz’s work received some support, the public attack against Steinmetz and her family evidenced the public’s overwhelming rejection of her work.\(^{46}\) Verbal threats were launched against her and her children—at home and in public. Threatening phone calls were made to Steinmetz and the sponsors of her speaking engagements in order to prevent Steinmetz from further publicizing her work. On one occasion, a bomb threat was called into an ACLU meeting at which Steinmetz was scheduled to speak.\(^{47}\) Professionally, Steinmetz was also threatened. In an attempt to prevent her from receiving tenure, every female faculty member at the University of Delaware was lobbied by individuals calling on behalf of the women’s rights movement.\(^{48}\) Academicians also became involved in the per-

---

41. The “twelve million” figure is attributed to the reporting of Roger Langley, a journalist for the New York Daily News. For discussion of the early media reports, see INTIMATE VIOLENCE, supra note 14, at 105-06.
42. Id. at 105.
43. For a discussion of Steinmetz’s work, see supra Part I.A.
44. PHILIP W. COOK, ABUSED MEN: THE HIDDEN SIDE OF DOMESTIC VIOLENCE 109-14 (1997); INTIMATE VIOLENCE, supra note 14, at 105-06.
45. COOK, supra note 44, at 109-14; INTIMATE VIOLENCE, supra note 14, at 105-06.
46. For my discussion of why there has been such a strong reaction to the study of battered men, see infra Part III.
47. COOK, supra note 44, at 109-12.
48. Id. at 112.
sonal attack, deriding her work as anti-feminist and simply biased to its funding source.49

Other social scientists committed to the study of husband abuse and family violence were similarly mistreated.50 Such tactics seem to have proven effective. Both researchers who were involved in the early projects, and even those who might have become involved, admit that they now choose to give the topic of battered men “wide berth.”51 Such a commentary is tragic, not only for those interested in female violence, but for all of us committed to protecting academic research and intellectual freedom.

2. By Methodological Critique

(a) CTS Challenges

Beyond woozles and scare tactics, a more effective and facially neutral intellectual tactic used to silence the study of female violence has been an attack on the methodology. As an initial criticism, the CTS-based reports, by definition, allow only a focus on violence resulting from conflict situations. While acknowledging the value of CTS in other social studies, sociologists critique their use in the study of family.52 Because the scale’s focus upon “conflict” does not acknowledge the use of violence in a familial setting as a tactic of coercive control, such methodology fails to emphasize the use of violence by men to maintain power or the use of violence without provocation.53 The CTS reporting methodology is also criticized for its limited focus upon the acts of violence, not the consequences, or more specifically, the severity of the injuries resulting from such acts.54 For example, the ordering of the violence with such acts as “trying to hit with something” regarded as more severe than “slapping” is deemed inappropriate given the potential of severe physical injury which can result from a slap, while no injury could ever result from throwing

49. Id. at 109-12.
50. Many years after the initial reporting of the family violence research team, a seemingly embittered Straus charged that although his early work on violence against women had been commonly relied upon by feminists in their arguments of institutional male violence, he was simply “excommunicated” from the feminist ranks after the family violence surveys’ publications. Murray A. Straus, Physical Assaults by Wives: A Major Social Problem, in CURRENT CONTROVERSIES ON FAMILY VIOLENCE 82 (Richard Gelles & Donileen R. Loseke eds., 1993). For earlier reaction of the family researchers criticized for their findings, see INTIMATE VIOLENCE, supra note 14, at 105-06.
51. INTIMATE VIOLENCE, supra note 14, at 106.
52. Kersti A. Yiilö, Through a Feminist Lens: Gender, Power and Violence, in CURRENT CONTROVERSIES ON FAMILY VIOLENCE, supra note 50, at 47, 52-53.
53. Demie Kurz, Physical Assaults by Husbands: A Major Social Problem, in CURRENT CONTROVERSIES ON FAMILY VIOLENCE, supra note 50, at 88, 94-96; Yiilö, supra note 52, at 53. For the related critique that the CTS’ noncontextual approach fails to understand the gendered nature of domestic violence, see infra Part II.B.2.
54. Yiilö, supra note 52, at 52-53.
something at someone but failing to strike. Criticism is also levied at separate groupings of various types of violence which are instead seen as overlapping.

Recognizing such concerns, the Family Violence Research Laboratory addressed them in their initial studies. However, rather than rejecting the methodology and therefore any information it produced outright, the researchers noted that CTS was previously accepted as a methodology in family studies limited to wife abuse. They therefore rationalized that the research methodology remained valuable in a combined study of husband abuse and wife abuse. Moreover, despite being collected through CTS methodology, the research of the Family Violence Researchers yields similar results to numerous other studies of husband and wife abuse, including those which rely upon non-CTS methodology.


56. For example, there is criticism over separating “kick[ing], bit[ing], or hit[ting] with a fist” from “[b]eat[ing] up the other one.” Id. at 59. For a review of the CTS categories, see supra note 12 and accompanying text.


58. Straus, supra note 50, at 84.

59. By 1999, over 100 studies are counted as consistently finding that men and women engage in domestic violence at similar rates. See Murray A. Straus, The Controversy Over Domestic Violence by Women: A Methodological, Theoretical, and Sociology of Science Analysis, in VIOLENCE IN INTIMATE RELATIONSHIPS 17, 17-18 (Ximena B. Arriaga & Stuart Oskamp eds., 1999). For a review of a number of studies of family and partner abuse yielding similar rates of abuse by men and women, see id. at 25-27. Straus, supra note 50, at 68-72 (discussing the results obtained from a study of the assault rate of partners in married and cohabitating relations and a study of the assault rate of individuals in dating relationships); see also PHYSICAL VIOLENCE, supra note 14, at 162 (comparing the similarity in data on female and male abuse to earlier studies and concluding that “[t]hese findings are so consistent that they leave little doubt about the high rate of assaults by women”); Terrie E. Moffitt, Partner Violence Among Young Adults, NAT’L INST. JUST.: RESEARCH PREVIEW (Apr. 1999) (discussing preliminary results of an over twenty-year ongoing longitudinal study in New Zealand finding women reporting greater acts of domestic violence against their partners than men); Daniel G. Saunders, Wife Abuse, Husband Abuse, or Mutual Combat? A Feminist Perspective on the Empirical Findings, in FEMINIST PERSPECTIVES ON WIFE ABUSE, supra note 55, at 96 (noting numerous studies supporting the work of the Family Violence Research team through findings which reported “23% to 71% of battered women used violence ... against their abusers”).

For discussion of other non-CTS studies yielding comparable results regarding the similar usage of violence by men and women, see FORD, supra note 35, at 7-14 (providing an overview of numerous critical studies including the National Crime Victimization Survey, National Family Violence Surveys, National Youth Survey and homicide and other intimate violence statistics collected by the U.S. Bureau of Statistics); WILLIAM A. STACEY ET AL., THE VIOLENT COUPLE (1994) (conducting a survey of eighty-six couples in Austin, Texas who were participating in a counseling program, but noting at the outset the skewed nature of the sample, largely because in two-thirds of the cases the couples were participating as a result of a court referral to the counseling program in lieu of the male member of the couple being convicted of domestic violence); Barbara J. Morse, Beyond the Conflict Tactics Scale: Assessing Gender Differences in Partner Violence, 10 VIOLENCE & VICTIMS 251 (1995) (reporting the results of the National Youth Survey, a nationally representative
(b) Gender Concerns

Other important methodological issues may also be accounted for. Of significant concern is the belief that men tend to underreport their acts of domestic violence, and that any study relying upon surveying men would therefore yield inaccurate results.\(^{60}\) However, in a variety of ways, the husband-abuse studies minimized this threat. First, while the CTS surveys collected data on households, typically only one spouse completed the questionnaire or interview. Since, in the majority of instances, the participating spouse was the wife, the perceived threat of unreliable male reported data was reduced at the outset.\(^{61}\) In the 1985 CTS study, the Family Violence Research team went further to address the risk of male underreporting by isolating the information reported by men from the information reported by women. Even when the researchers relied solely on the data reported by women, the data revealed that the use of violence by men and

seventeen-year longitudinal survey of intimate violence among married and cohabitating partners and finding trends and patterns of use of minor and severe violence by men and women similar to those reported by the National Family Violence Survey).

It must be noted that a number of other surveys have found significantly higher rates of violence by men than women. In 1997, the results of the National Violence Against Women in America Survey (NVAW) were released. NVAW surveyed 16,000 households, equally interviewing 8000 men and 8000 women. It concluded that men physically assaulted their female partners at three times the rate women engaged in such behavior. Such conclusions regarding disparate rates of assault, however, must be interpreted in light of the survey questions which were geared to focusing on injury-producing violence and types of violence more likely to be considered criminal. Straus, \textit{supra}, at 26-27.

Similarly, several other commonly cited surveys are the “battered women’s studies” which generally rely on data collected from women residing in battered women’s shelters; “police call studies,” which are based on statistics gathered from calls made to the police by victims of domestic violence, who are typically women; and the National Crime Victims Survey, a government survey based upon criminal statistics regarding domestic violence. These surveys have been criticized as based on non-representative, skewed samples with disproportional numbers of abused women. The battered women’s shelter studies are non-representative because they rely only upon interviews with female victims who are residing in shelters. Similarly, because it is acknowledged that male victims of violence are less likely to call the police than female victims, the police call surveys also are based on samples with a disproportionate number of female victims. Likewise, the National Crimes Victims Survey is seen as non-representative because acts of domestic violence committed by men are more commonly reported and treated as a “crime” than domestic violence perpetrated by women. Straus, \textit{supra} note 50, at 68-72; see also, STACEY ET AL., \textit{supra}, at 45 (acknowledging the unreliability of official police and court statistics on domestic violence and quoting one critic as declaring such data to be “practically worthless for purposes of criminological research”); Steinmetz, \textit{supra} note 6, at 503-04 (critiquing earlier studies of female violence due to the skewed sample). For a further general discussion of the methodological problems and non-representative concerns in both clinical studies (such as those based upon criminal and shelter data) and representative community samples (such as the National Family Violence Survey), see Straus, \textit{supra} note 50, at 77-80.

60. Straus, \textit{supra} note 50, at 68 (acknowledging the concern of male underreporting).

61. STEINMETZ, \textit{supra} note 13, at 14-16 (noting that of the fifty-seven families studied in her survey, 35% of the husbands-fathers participated in the questionnaire component of the survey, while only 9% of the personal interviews relied upon male participation). For further discussion of Steinmetz’s work, see \textit{supra} notes 6-34 and accompanying text.
women remained at similar rates. Moreover, the parity between the acts of violence committed by men and women existed when the data was separated into the CTS indices of “minor” and “severe” acts of violence.62

B. Defending Female Violence

1. Quantitative Criticisms

If the methodological critiques are accurate, the assertion that men and women engage in similar patterns of domestic violence can be denied fairly. Yet even effectively raising the shortcomings of these criticisms (as I have just attempted to do) does not sufficiently respond to the critics. For those interested in discrediting the assertion that men and women both act violently, a bolder move is to not only accept the female use of violence, but to defend it. In conceding that women do engage in acts of domestic violence, female use of vio-

<table>
<thead>
<tr>
<th>Perpetrator of Violence</th>
<th>“Minor”</th>
<th>“Severe”</th>
<th>“None”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>6.9%</td>
<td>4.9%</td>
<td>88.1%</td>
</tr>
<tr>
<td>Female</td>
<td>7.7%</td>
<td>4.4%</td>
<td>87.9%</td>
</tr>
</tbody>
</table>

As reported by men, the 1985 Family Violence Research Survey data indicated the following percentage of spouses who engaged in domestic violence toward their mates:

<table>
<thead>
<tr>
<th>Perpetrator of Violence</th>
<th>“Minor”</th>
<th>“Severe”</th>
<th>“None”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>9.2%</td>
<td>1.3%</td>
<td>89.5%</td>
</tr>
<tr>
<td>Female</td>
<td>7.5%</td>
<td>4.7%</td>
<td>87.8%</td>
</tr>
</tbody>
</table>

PHYSICAL VIOLENCE, supra note 14, at 156-57, 162.
There appears to be some dispute on the gender honesty question. In comparing the 1975, 1985 and 1992 data, other statisticians have found that men tended to report greater rates of male violence than women, and that women tended to report greater rates of female violence than men. For a breakdown by gender differences per reporting period, see Ford, supra note 35, at 13. See also Gelles, supra note 35, at 797.
Vio-
elence is justified as self-defense—a lifesaving reaction of women who are being physically attacked by their male partners.63

The development of the battered woman syndrome as a defense for crimes committed against abusive male partners, including homicide, evidences the wide acceptance of a woman’s use of violence as self-defense.64 The self-defense theory of female domestic violence is not, however, fully supported by the statistics. According to the statistics on intimate violence, while in approximately 50% of cases both spouses are reported to act violently, in the remaining 50% only one spouse is reported to ever use domestic violence.65 Admittedly, the finding that in approximately 50% of cases both spouses engage in violence does not help in determining who is initiating the violence.66

63. Dobash & Dobash, supra note 55, at 59-60. For the self-defense characterization of the statistical reporting of violence by women, see Daniel Saunders, When Battered Women Use Violence: Husband-Abuse or Self-Defense?, 1 VIOLENCE & VICTIMS 47 (1986). For further discussion of the battered woman’s defense in employing violence, see infra notes 64, 69, 137 and accompanying text.

64. Lenore Walker, a pioneer in the study of domestic violence, is credited with the development of the battered woman’s syndrome. See, e.g., LENORE E. WALKER, TERRIFYING LOVE: WHY BATTERED WOMEN KILL AND HOW SOCIETY RESPONDS (1989); Lenore E. Walker, Battered Women Syndrome and Self-Defense, 6 NOTRE DAME J.L. ETHICS & PUB. POL’Y 321 (1992). For a sampling from the tremendous body of literature discussing the creation, use, and debate surrounding the battered woman’s syndrome, see ANGELA BROWNE, WHEN BATTERED WOMEN KILL (1987); AMY LOU BUSCH, FINDING THEIR VOICES: LISTENING TO BATTERED WOMEN WHO’VE KILLED (1999); INTIMATE VIOLENCE, supra note 14, at 141-59 (challenging the notion of “learned helplessness” by reporting on the variety of female responses to domestic violence); Anne M. Coughlin, Excusing Women, 82 CAL. L. REV. 1 (1994) (critiquing the battered women’s syndrome defense as a reinforcement of negative female stereotypes). For further discussion of battered women’s syndrome and its relation to the patriarchal definition of domestic violence, see infra notes 109, 116, 254 and accompanying text. For further discussion of learned helplessness, see infra note 130 and accompanying text.

There are few efforts to extend the battered women’s syndrome beyond women to include other intimates who are abused and defend themselves with violence. This one-sided application of a self-defense theory for intimates supports this Article’s argument regarding the unquestioning acceptance of a feminist-oriented definition of domestic violence. See infra notes 120, 256 and accompanying text (on the pervasive impact of the feminist definition of domestic violence on legal and social institutions). For one author’s effort to extend the battered women’s syndrome to all victims of domestic violence, see Hope Toffel, Note, Crazy Women, Unharmed Men, and Evil Children: Confronting the Myths About Battered People Who Kill Their Abusers, and the Argument for Extending Battering Syndrome Self-Defenses to All Victims of Domestic Violence, 70 S. CAL. L. REV. 337 (1996).

65. STRAUS ET AL., supra note 12, at 36-37.

66. It is interesting to note, however, that the 1985 Family Violence Survey found that women were more likely than men to cite using violence as their common response to the violence of their partners. PHYSICAL VIOLENCE, supra note 14, at 155-56. Based upon a given range of possible responses (from crying, yelling, running to another room, hitting back, running out of the house, calling a friend or relative, or calling the police), 24% of women and 15% of men cited violence as their common response. Id. at 155. The most common response of women was to cry (54%), while the most common response of men was “other” (32%). Id. This high report of “other” by men has suggested to researchers that men are more likely to ignore the violence of women than to respond in any affirmative manner. Id. at 155-56; see also INTIMATE VIOLENCE, supra note 14, at 149-50. For a consistent discussion of the minimization of female violence by men, see infra note 90.
Consequently, the battered woman/self-defense theory is a plausible explanation for the female use of violence in those cases. However, the remaining 50% of couples who report violence by only one spouse further breaks down to reveal that while the husband is the sole perpetrator in one half of such cases, the wife is the sole perpetrator in the remaining half. Moreover, when questioned specifically as to initial abuse, men and women report initiating violence at similar rates. These virtually identical rates of violence by men and women as sole perpetrators call into question the assertion that women’s use of domestic violence is always defensive; they also suggest that women may be the only physical aggressor in violent relationships as often as men.

67. Based on the 1985 National Family Violence Survey results of 495 couples admitting to violence, in 48.8% of the cases both spouses were violent, in 25.3% only the male was violent and in 25.5% only the female was violent. Straus, supra note 50, at 74. Similarly, in the 1975 study, 49% of the couples who engaged in violence did so on a mutual basis, whereas of the remaining 51%, in 27% of the cases only the husband was violent and in 24% of the cases only the wife was violent. Straus et al., supra note 16, at 36-37; see also Intimate Violence, supra note 14, at 146-50, 258 (discussing female and male reaction to violence yet providing graph and detailed statistics only for female reaction).

68. This statistic did not significantly vary based upon the sex of the survey respondent. Broken down by gender, the 1985 Family Violence Survey reported the following results regarding who initiated the violence:

<table>
<thead>
<tr>
<th>Gender Reporting</th>
<th>Male Initiator</th>
<th>Female Initiator</th>
<th>No Memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>43.7%</td>
<td>44.1%</td>
<td>12.2%</td>
</tr>
<tr>
<td>Female</td>
<td>42.6%</td>
<td>52.7%</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

Physical Violence, supra note 14, at 154-55.

69. Various quantitative studies have been offered to bolster the self-defense motivation for a woman’s use of violence. In a study first reported in 1986 aimed at determining the motivations for a woman’s use of violence, Daniel Saunders, a self-described “feminist researcher,” interviewed fifty-two battered women and found that “self-defense” was the most frequently cited reason given for the use of violence, with 40% of women using severe violence and 30% of women using nonsevere violence as defined by a “modified Conflict Tactics Scale.” Saunders, supra note 59, at 102-05. “Fighting Back” was also commonly cited, with 33% of women using severe violence and 23% of those using nonsevere violence ascribing to such motivation. Id. at 105. By contrast, only one woman (representing 3% of the study) who used severe violence and four women (11%) who used nonsevere violence explained their motivation as self-initiated. Id. As acknowledged by Saunders, however, interviewing only battered women created a skewed sample which fails to include nonbattered women who rely upon violence. Id. at 108. Without a representative sample of battered and nonbattered women, Saunders’ study fails to accurately reflect when women’s use of violence is self-initiated or in response to male violence. For an earlier, exclusive discussion of Saunders’ study, see Saunders, supra note 63. Other studies purporting to support the self-defense theory have also been acknowledged to be inconclusive. See Saunders, supra note 59, at 97-101.
2. Qualitative Criticisms

Yet, even such quantitative statistics on the similar rates of violence initiation does not end the controversy surrounding female use of domestic violence. Qualitative defenses are also offered to explain the female use of violence. An appreciation of the “power and control” dynamics of domestic violence cautions against relying solely upon such statistics to reject outright the female use of violence as self-defense theory.70 When the severity of a woman’s violent behavior is greater than her spouse’s, it is suggested that a woman’s lack of training in less violent reactions may excuse her behavior.71 A woman’s anger toward her mate, rather than simply her fear of violence, is also raised as a defense.72 Such a defense is illustrated, for example, in the characterization of the female use, perhaps even initiation, of domestic violence as simply a “slap the cad” response to offensive, even if not physically violent, male behavior.73 Others have gone further in defending the female use of violence. Warning against describing male behavior as simply caddish, the defense raised instead is that women’s use of violence is warranted when men engage in “unwanted sexual advances, belittling of . . . [women], verbal intimidation, [and] drunken frenzy.”74 In this vein, the question changes from, “[W]ho began [the] hitting?” to, “[W]ho began the argument?”75 Such a change is more dangerous than defending female violence as a response to caddish behavior, for it risks legitimizing female violence whenever the argument is male initiated. However, this focus on the argument, rather than on the use of violence, is significant in that it exposes the core justification for disregarding the violence perpetrated by women.

By focusing on the question of who initiates the argument, the implication is that because male anger carries the threat of greater harm, acts of female violence cannot only be condoned as a preventive defense, but can also be overlooked. In choosing the latter, we

70. In reviewing his findings, Straus suggested that statistics on the initiation of violence by women needed to be interpreted in consideration of such factors as: (1) a woman being more likely to take the blame for violence initiation; (2) a woman’s likelihood of understanding the question “Who initiated the violence?” as “Who initiated the fight?”; (3) a woman perhaps being the “first hitter” in an isolated incident, but the act is defensive in contrast to the man’s initiation of violence in the relationship; and (4) a woman perhaps reporting to have initiated violence in response to a question directed at solely examining the most severe incidents of violence and framing a question in such a manner which prevents recognizing that the man may have initiated the use of violence in the relationship. See Straus, supra note 50, at 75-76.

71. Saunders, supra note 59, at 99 (discussing State v. Wanrow, 559 P.2d 548, 558 (Wash. 1977)).

72. Id. at 107.

73. Straus, supra note 50, at 79.

74. Kurz, supra note 53, at 96.

75. PHYSICAL VIOLENCE, supra note 14, at 155.
shift from a position of defending female violence to minimizing its consequences. Simply put, female violence can be legitimately ignored because male violence causes greater injury.

C. Minimizing Female Violence

Since the first controversial exposure of husband battering, there has been little debate on the assertion that male domestic violence is more likely to produce injury. In attempting to understand why husband battering was given such “selective inattention,” pioneer Suzanne Steinmetz determined that the realization of the greater physical strength of men as compared to women provides the most plausible explanation for disregarding female violence.76 Male violence produces injury at six times the rate of female violence.77 Comparing the type of injuries also shows that women suffer greater physical and psychological harm when physically assaulted. When measuring physical injury via the three categories of the need for medical care, time off from work, and time spent bedridden, women rank higher in every category.78 In terms of psychological injuries, while abused men and women consistently display psychosomatic symptoms, abused women suffer greater depression and stress levels than abused men.79

76. Steinmetz, supra note 6, at 504-06. For further discussion of Professor Steinmetz’s work on female and male domestic violence, see supra notes 10-34 and accompanying text.
77. Relying upon the 1985 Family Violence Research Survey, it was concluded that 3% of male assaults caused injuries compared to injuries produced in only 0.4% of the cases involving female assault. Straus, supra note 50, at 69; see also Steinmetz, supra note 6, at 505.
78. These statistics were based upon the 1985 Family Violence Research Survey. See PHYSICAL VIOLENCE, supra note 14, at 152. According to the results regarding the need for medical care, 7.3% of women versus 1% of men who were the victims of “severe violence” as defined by the CTS scale required such attention. Id. at 157. Neither sex reported the need for medical care when victim to minor violence, as defined by CTS. Id. In response to the need to take time off from work, 19% of women versus 10% of men resorted to such a measure when victim to “severe violence.” Id. When victim to minor violence, both sexes requested time off at the rate of 4%. Id. In regard to having to resort to bed rest, of “severe violence” victims, 22.8% of women required one or more days in bed, compared to 13.8% of male victims. Id. at 158. The need for bed rest was also broken down by minor and severe violence to reveal that when the violence was minor, 15.2% of women and 12.8% of men needed bed rest. Id. at 159. Commenting on the results, the researchers observed that the higher rate of female reaction did not conclusively indicate that women suffered greater injury. Id. at 158. Researchers cautioned that the data should be considered in conjunction with such factors as the lesser amount of humiliation suffered by women as opposed to men who are victims of domestic violence, as well as the relative ease in which women might be able to take time off from jobs which typically provided the family’s second, lower wage and a more flexible schedule. Id. at 154-64. On the stigma associated with being a male victim of domestic violence, see infra notes 183, 285 and accompanying text.
79. The study initially found that of non-victimized individuals, men and women disproportionately displayed psychosomatic symptoms (26.8% of non-victimized women and 15.9% of non-victimized men displaying psychosomatic symptoms). PHYSICAL VIOLENCE, supra note 14, at 158-59. When studying abused men and women, psychosomatic symptoms were positively correlated to increased levels of domestic abuse. Id. However, abused
In addition to the higher risk of injury faced by abused women, the differing rates of injury also supply an argument for dismissing the statistical parity between the genders in their use of violence. Men are seen to possess the “single beating” advantage, as the mere threat of causing injury allows men to control women without physically having to raise a hand after the first beating.80

D. Responding to The Critics: Why Female Violence Must Be Examined

Acknowledging the damage differential in the domestic violence used by men and women may seem to end the need to study female violence.81 Yet does it? A number of important practical and theoretical justifications militate against ignoring female violence. First, notwithstanding the “damage differential,” some important normative observations about men and women can be drawn from an acknowledgment of male and female violence. Given the statistical parity in the use of domestic violence,82 there appears to be no basis for the traditional belief that women are either born or bred to be less physically aggressive than men. Likewise, the statistics do not bear out the “nagging” wife stereotype. Women are not more prone to engage in verbal abuse than men.83 Moreover, the recognition of the difference in consequences between male and female violence does not diminish the fact that men and women bear similar intentions in regard to their inclination to engage in intimate violence.84 In fact, their comparable intent leads to similar results when the physical strength difference between men and women is taken into account. Controlling for the “hand-to-hand” combat advantage of men by relying solely upon statistics measuring injury produced by domestic violence involving a weapon, the rate at which men are injured by

---

80. Strauss et al., supra note 12, at 42; see also Straus, supra note 14, at 446.
81. Steinmetz, supra note 6, at 505 (discussing the idea of damage differential).
82. Id.
83. Ironically, statistical data reveals that men are as likely or more likely than women to engage in the “nagging” type verbal abuse. Id.
84. Id.
women is similar or greater than the rate at which women are in- 
jured by men.\footnote{85} Put succinctly by one commentator, “[a]pparently, it’s just a matter of style.”\footnote{86}

Second, focusing on the injury, rather than the assault, contra-
dicts the understood campaign against wife-beating which has been
to end wife abuse per se, not just the violence which produces in-
jury.\footnote{87} To support a domestic violence policy restricted only to injury-
produced violence would end the expectation of any legal response or
social protection to noninjured victims.\footnote{88}

A third argument for resisting the effort to deny, defend, or mini-
mize female abuse of men lies in recognizing that the decrease in
male abuse of women is largely credited to the attention which has
been given to male abuse of women since the earliest explorations of
domestic violence. Increased treatment and prevention programs,
counseling for male abusers, and shelters for abused women are
amongst other important items credited for the significant reductions
in male abuse of women.\footnote{89} Defining domestic violence as the abuse of
women by men has brought a growing cultural intolerance for wife
abuse, while there is reported to be little change in the tolerance of
female violence.\footnote{90} In very real terms then, the failure to stigmatize or

\footnote{85. Philip W. Cook, Female Violence Against Men Is a Serious Problem, in DOMESTIC
86. Id.
87. See Straus, supra note 50, at 69-70.
88. See Ford, supra note 35, at 23.
89. Other factors raised to explain reductions in male violence on women include
changing family roles and improved economies. INTIMATE VIOLENCE IN FAMILIES, supra
note 14, at 79. For further discussion of the history of funding and attention given to wife
and child abuse and its role in reducing the abuse of women and children, see id. at 19-39.
For further discussion of treatment programs, see infra notes 234-35, 238, 241, 243, 258
and accompanying text. For further discussion of shelters and victim services, see infra
notes 235, 261, 263-76 and accompanying text.
90. The reported decline in the rates of male abuse of women and the lack of change
in the rates of female abuse of men is matched by findings in attitudinal studies. In one
study, data collected from four surveys over a twenty-six year period (from 1968 through
1994) was compared to measure the change in cultural norms toward domestic violence.
Each survey included two critical questions: (1) whether it was acceptable for a husband to
hit a wife, and (2) whether it was acceptable for a wife to hit a husband. Murray A. Straus
et al., Change in Cultural Norms Approving Marital Violence from 1968 to 1994, in OUT OF
THE DARKNESS: CONTEMPORARY PERSPECTIVES ON FAMILY VIOLENCE 3, 6 (Glenda Kauf-
man Kantor & Jana L. Jasinski eds., 1997). A comparison of the responses for the twenty-
six year survey period demonstrated that approval for a husband slapping a wife had “de-
creased sharply”—from 20% to 10% of all individuals (both men and women) approving
such behavior. Id. at 7. By contrast, the approval given by both sexes for a wife slapping a
husband remained “almost identical over the 26 years,” with little variance from the 20%
approval figure measured in 1968 and again in 1994. Id. A breakdown of the approval
rates by gender showed other similarities. Id. at 7-12. While approval rates for men slapping
women had decreased with both sexes, men were consistently found to be more ac-
cepting of such behavior. Id. at 12-13. Similarly, while approval rates for women slapping
men remained constant with both sexes, men were consistently found to be more accepting
of such behavior. Id.
even acknowledge the female abuse of men allows and encourages its continuation.°

The significant decreases in the use and approval of male violence, in drastic contrast to the lack of change in use or approval of female violence, has led researchers to what appears to be a self-evident conclusion. “[S]ocial movements condemning violence against women, legal and institutional reforms, and systemic antiviolence educational efforts can produce major changes in public attitudes about violence and should therefore be expanded.”°° However, even recommending “zero tolerance of violence by both men and women” does not necessitate complete insensitivity to the greater risk of injury which female victims of domestic violence face.°° Yet an emphasis on the abuse of women by men is far different from demanding an exclusive focus on such abuse to the preclusion of the abuse of men by women. Indeed, even if motivated solely by an interest in ending wife or child abuse, it is still necessary to address the various forms family violence can take—including female violence. The abuse of men by women and the abuse of children by either parent are two forms of family violence which are directly related to any effort to systematically address wife abuse.°° Consequently, a fourth argument for acknowledging and addressing the abuse of men by women is that it will ultimately work to end the abuse of women by men. Put in blunt utilitarian terms, female violence must be addressed in order to protect women as a man provoked by a violent female has the potential to inflict greater injury.°°

This argument is somewhat controversial because the demand for self-control is placed solely on the female and seems tantamount to victim-blaming.°° Such an objection may legitimately refute any ar-

---

91. See supra notes 20-22, 24-27, 30-32, 35, 37-41 and accompanying text (discussing the changing rates of violence per gender over time).
92. Straus et al., supra note 90, at 14.
93. Id.
94. For a discussion of the relation of wife abuse to other family violence, see Straus et al., supra note 12, at 97-122. For a recognition of the many other forms of family violence, see infra notes 297-300 and accompanying text.
95. Straus, supra note 50, at 79. For a discussion of the differences in the injury caused by abusive men and women, see supra notes 76-79 and accompanying text.
96. See Morse, supra note 59, at 252. In this respect, addressing the abuse of men by women in order to protect women from further abuse is similar to the “victim-blaming” criticisms which were levied at the early efforts to address the abuse of women by targeting the (non-violent) actions taken by female victims believed to trigger male violence. For various discussions of victim-blaming, see, for example, Dobash & Dobash, supra note 8, at 159-60 (cautioning against allowing the public to blame victims of violence); Donald G. Dutton, The Batterer: A Psychological Profile 26 (1995); V. Michael McKenzie, Domestic Violence in America (1995) (acknowledging domestic violence victim self-blaming ability); Barbara Hart, Battered Women and the Criminal Justice System, in Do Arrests and Restraining Orders Work? 98, 101 (Eve S. Buzawa & Carl G. Buzawa eds., 1996) (recognizing victims’ assumption of guilt for beatings suffered and batterers’ contribution to such feelings).
argument raised to look at female violence in order to protect women. However, this type of controversy does not prevent recognizing the effect of husband-beating on children. Regardless of the gender of the child or the violent parent, children who witness the violence of one parent on another are more likely to be violent in their adult relationships.97 These findings therefore provide the basis for a fifth reason for supporting a gender-neutral effort to address spousal violence. Witnessing either husband-beating or wife-beating as a child is equally likely to breed a predisposition toward intimate violence.98

Finally, a wider look at family violence that includes an awareness of husband-beating will also allow a greater emphasis to be placed on the socio-cultural factors which teach violent tendencies.99 Male-blaming can no longer be relied upon as the single explanation for the ills of society.

From such an array of reasons for addressing female violence, it may seem that no one should be against such an endeavor. Some reason should appeal to everyone. Acknowledging female violence arguably not only will protect men, but it will ultimately work to protect women and children. Social causes of family violence also seem more likely to be given serious attention. Moreover, an effort to address actions, rather than injury, maintains the integrity of traditional domestic violence policy. What then is the harm in targeting, discussing, or even revealing the abuse of men by women? Why has such abuse not been “explode[d]” in the manner Catherine MacKinnon demands the abuse of women by men must be treated?100 Why does the denial continue?

Although the reactions against examining female violence may vary, the critics share key theoretical and practical motives which are often intertwined. It is these motives that not only account for the vehement rejection of efforts to study female violence (unless it is couched in the language of self-defense), but are also responsible for shaping our legal, cultural and social policies and norms on domestic violence. Looking at the development of domestic violence theory is

98. While such discussion is beyond the context of this Article’s focus on husband-beating, it must be noted that the direct abuse of children is also directly linked to spousal abuse. For example, people who experience the most physical punishment as teenagers are four times more likely to be wife-beaters or husband-beaters than those not physically abused. Id. at 110. For further discussion of the childhood victimization of adult abusers and its creating a predisposition toward family violence, see Lee H. Bowker et al., On the Relationship Between Wife Beating and Child Abuse, in Feminist Perspectives on Wife Abuse, supra note 55, at 158.
99. Steinmetz, supra note 6, at 507. For discussion of the need to move away from male-bashing explanations in order to both further social progress and legitimize feminism, see infra Part V.
the foundation for exploring such explanations and their implications.

III. THE EXPLANATIONS

A. Female Violence: A Theoretical Threat

1. The Development of the Patriarchal Definition of Domestic Violence

The “discovery” of domestic violence is credited to the battered women who came forward in the 1970s and began telling their stories in the new female focused community centers of England.101 Of course, violence against women was certainly not a new phenomenon. It had not only been previously recognized, but also, on a sporadic and brief occasion, been delegitimized.102 However, it is the 1970s’ attention to the domestic violence suffered by women which marked the beginning of the current effort to eradicate domestic violence. Consequently, it is this identification of domestic violence as a woman’s issue that shapes today’s understanding of domestic violence. Because of this background, the definition of domestic violence has developed as the use of physical power by men against women not motivated simply by a desire to inflict physical pain or even emotional suffering but rather as part of a larger effort by men to gain and maintain control over women.

While such a definition may now be accepted without question, the characterization of the male as sole user of physical force and the female as sole recipient was revolutionary in several important respects.103 Recognizing domestic violence as a social phenomenon, the male as batterer/female as victim perspective largely dispelled earlier understandings of domestic violence as an illness suffered by both the batterer and the victim.104 From this medical perspective, a

101. DOBASH & DOBASH, supra note 8, at 1-3; see also ELLEN PENCE & MICHAEL PAYMAR, EDUCATION GROUPS FOR MEN WHO BATTER: THE DULUTH MODEL 173 (1993).
102. DOBASH & DOBASH, supra note 8, at 3 (noting the “fleeting” focus on domestic violence in the late nineteenth and early twentieth centuries).
103. As Lenore Walker described:
A battered woman is a woman who is repeatedly subjected to any forceful physical or psychological behavior by a man in order to coerce her to do something he wants her to do without any concern for her rights. Battered women include wives or women in any form of intimate relationships with men. Furthermore, in order to be classified as a battered woman, the couple must go through the battering cycle at least twice. Any woman may find herself in an abusive relationship with a man once. If it occurs a second time, and she remains in the situation, she is defined as a battered woman.
104. This sociological treatment of domestic violence is recognized to have coincided with the general growth of sociology and the civil rights movement of the 1960s. Yllö, supra note 52, at 50.
batterer’s use of domestic violence had been explained as the product of illness.\(^\text{105}\) Likewise, a victim’s inability to leave a battering relationship was understood as a manifestation of her own masochistic or pathological nature.\(^\text{106}\) Reacting to this limiting, medical jargon, an alternative definition of domestic violence was eventually found in Lenore Walker’s “cycle of violence,” with its ongoing pattern of tension building, acute battering and batterer contrition.\(^\text{107}\) The recognition that a battered woman may stay in an abusive relationship for legitimate reasons including love, concern for her children, and fear of heightened violence upon separation was also welcomed as it further encouraged an end to victim blaming.\(^\text{108}\)

\(^{105}\) For a discussion of the early treatment of male battering as an illness and the gradual move from medical definitions to more sociological understandings of domestic violence, see DUTTON, supra note 96, at 61-77; DANIEL JAY SONKIN ET AL., THE MALE BATTERER: A TREATMENT APPROACH (1985); WALKER, supra note 103, at 205-50; see also infra note 106.

\(^{106}\) While much work has been done to correct these medical misconceptions, battered women remain subject to a host of overly simplified characterizations ranging from helpless victim to violence provo- cateur. For a recognition of the early medical stereotyping of battered women and its persistence today, see Naomi Cahn & Joan Meier, Domestic Violence and Feminist Jurisprudence: Towards a New Agenda, 4 B.U. PUB. INT. L.J. 339, 343-44 (1995); Cheryl Hanna, No Right to Choose: Mandated Victim Participation in Domestic Violence Prosecutions, 109 HARV. L. REV. 1849, 1882-85 (1996); Joan S. Meier, Notes from the Underground: Integrating Psychological and Legal Perspectives on Domestic Violence in Theory and Practice, 21 HOFSTRA L. REV. 1295, 1301-02 (1993).

\(^{107}\) For the gradual development in understanding battered women from pathological and sadistic to more sociological definitions, see OLA W. BARNETT & ALYCE D. LAVIOLETTE, IT COULD HAPPEN TO ANYONE: WHY BATTERED WOMEN STAY 65-92 (2d ed. 2000); WALKER, supra note 103, at 18-54.

\(^{108}\) The myriad of reasons a domestic violence victim remains in the relationship has been widely discussed. For a discussion of the emotional difficulty in leaving due to an ongoing love for the batterer as well as the shame associated with leaving and publicly disclosing the relationship’s violent nature, see, for example, DOBASH & DOBASH, supra note 8, at 145-46; WALKER, supra note 103, at 27 (recognizing the love which is restored during periods of contrition in the cycle of violence); Kimberlé Williams Crenshaw, Panel Presentation on Cultural Battery, 25 U. TOL. L. REV. 891, 893 (1995) (acknowledging that a victim of abuse does not want to leave, but simply wants the violence to stop); Martha R. Mahoney, Legal Images of Battered Women: Redefining the Issue of Separation, 90 MICH. L. REV. 1, 8 (1991) (acknowledging her reluctance to tell her own story).

On the fear of separation assault, see, for example, Mahoney, supra addressing the fear of separation assault. For further discussion of separation assault, see infra notes 138, 169 and accompanying text.

On the victim’s concerns regarding the children’s physical safety and emotional well-being if removed from an abusive home, see DOBASH & DOBASH, supra note 8, at 148 (recognizing children are cited as the most common reason a battered woman stays in a relationship); Mahoney, supra, at 19 (recognizing that a mother’s decision-making process includes calculating her spouse’s and children’s interests); Elizabeth M. Schneider, The Dialectic of Rights and Politics: Perspectives from the Women’s Movement, 61 N.Y.U. L. REV. 589, 614-15 (1986) (recognizing that a woman’s connected nature tends to result in her making decisions regarding her own welfare in connection with consideration of her children’s and spouse’s interests).

On the economic issues associated with leaving, see CATHERINE T. KENNEY & KAREN R. BROWN, NOW LEGAL DEF. & EDUC. FUND, REPORT FROM THE FRONT LINES: THE IMPACT OF
While such work appreciably advanced our understanding of domestic violence, its ongoing focus on the physical aspect of domestic violence was soon criticized as secondary to the need to focus on the patriarchal dynamics surrounding the use of violence. 109 Such a perspective would properly emphasize that domestic violence “is not gender neutral any more than the economic division of labor or the institution of marriage is gender neutral.” 110 Replacing the cycle of violence, the “power and control” wheel provided a better means of depicting the gender driven nature of domestic violence and emphasized that physical violence formed only a part, albeit an important one, in the patriarchal scheme. 111 From a central hub of power and control, the wheel’s outer rim is formed by a circle of physical and sexual violence. This central power is then connected to the outer use of violence through spokes identified with such additional forces as using children, minimizing, denying, blaming, isolating, relying upon male privilege, coercing, threatening, intimidating, and abusing both


On the cultural and racial concerns which complicate a decision to leave, see Kimberlé Crenshaw, Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color, 43 STAN. L. REV. 1241 (1991). That similar reasoning may also explain the decision of male victims of violence to remain has also been specifically recognized. Cook, supra note 85, at 30 (arguing the male victims’ concerns regarding child custody); Steinmetz, supra note 6, at 507. Instead of explaining the victim’s reaction, others have emphasized the need to focus on the perpetrator. See Hanna, supra note 106, at 1879-80; Linda Kelly, Domestic Violence Survivors: Surviving the Beatings of 1996, 11 GEO. IMMIGR. L.J. 303, 305 (1997).

109. Mahoney, supra note 108, at 28-34 (criticizing the physical definition of physical violence given by such individuals as Lenore Walker, Mary Ann Douglas, Angela Browne, and Mildred Pagelow); id. at 53-55 (criticizing legal literature’s failure to focus on the “power and control, domination and subordination” dimensions of domestic violence as a result of the “traditional” acceptance of male domination); Meier, supra note 106, at 1317-22 (discussing the development of the patriarchal dynamics of domestic violence); see also G. Chezia Carraway, Violence Against Women of Color, 43 STAN. L. REV. 1301, 1305-06 (1991) (advancing the patriarchal definition of domestic violence as part of an overall effort to redefine violence against women of color in order to include “economic violence, cultural violence, legislative violence, medical violence, spiritual violence, emotional violence, and educational violence”).

The importance of social forces is likewise recognized by the family researchers in their study of both male and family violence. However, while acknowledging the patriarchal condition, their thrust is upon how socioeconomic forces can cause both men and women to behave violently. The external factors affecting both sexes include poor economic and employment conditions, as well as the social and media condonation of violence, racism, and sexism. See, e.g., INTIMATE VIOLENCE, supra note 14, at 194-206; THE SOCIAL CAUSES OF HUSBAND-WIFE VIOLENCE (Murray A. Straus & Gerald T. Hotaling eds., 1980); STRAUS ET AL., supra note 12, at 12-52.

110. Yllo, supra note 52, at 54.

111. The “power and control wheel” is part of the “Duluth Model” developed by the Duluth Domestic Abuse Intervention Project (DAIP) in Duluth, Minnesota. PENCE & PAYMAR, supra note 101, at 1-15. For further discussion of the critical impact of the “power and control wheel” on the treatment of domestic violence, see infra notes 247-55 and accompanying text.
emotionally and economically.\textsuperscript{112} In combination, such tactics allow the wheel of power and control to be spun by men in their efforts to control women.\textsuperscript{113} Violence, then, is not simply the male use of physical violence against women. Through the power and control wheel, domestic violence can more broadly be described as the male “way of ‘doing power’ in a relationship; battering is power and control marked by violence and coercion.”\textsuperscript{114} Such a framework also allows for the controlling effects of the violence on women and the pervasive use of violence as a patriarchal tool to be highlighted. “A battered woman is a woman who experiences the violence against her as determining or controlling her thoughts, emotions, or actions, including her efforts to cope with the violence itself. Many, many women experience such violence in our society.”\textsuperscript{115}

2. Challenging Domestic Violence, Challenging Feminism

(a) Through Theory

Beyond serving as merely an academic change, the widespread acceptance of this association of domestic violence with male power and control has critical consequences for society’s treatment of domestic violence.\textsuperscript{116} These very tangible results form an important part of this

\textsuperscript{112} Pence & Paymar, supra note 101, at 3.
\textsuperscript{113} Id. at 1-15; see also Yllo, supra note 52, at 54-55.
\textsuperscript{114} Mahoney, supra note 108, at 93.
\textsuperscript{115} Id.
discussion. However, before discussing the impact of the definition, it is important to remain at the theoretical level for a moment longer in order to fully grasp why there is such resistance to acknowledging female violence. Domestic violence is not viewed as just another tool used by men in the subordination of women. Rather, it is considered “one of the most brutal and explicit expressions of patriarchal domination.” Such strong roots in patriarchy have produced an equally strong force against accepting female violence. Acknowledging female violence risks negating the very basis of the existing domestic violence definition.

The consequences for domestic violence theory, however, are only a small part of a much larger threat. Domestic violence represents the prized gemstone of feminist theory’s fundamental message that our legal, social, and cultural norms are fashioned in a manner which permit men to engage in a constant and pervasive effort to oppress women by any and every available means. A successful challenge to the patriarchal definition of domestic violence may thus undermine feminism itself. To remain true to feminist theory, no aspect of male-female relations can be considered without first accepting the male as all powerful and the female as powerless. The gender hierarchy is omnipresent.

(b) Through Methodology

Given this dynamic, the suggestion that women may rely upon physical violence for anything other than self-defense must be rejected. However, the implications of grounding domestic violence in feminist theory go beyond prohibiting any consideration of female violence. Feminist theory also provides the means for discrediting


117. While the definition permeates our legal, social, and medical institutions, I have focused within this Article on discussing the practical effects such definition has upon the criminal justice system (particularly on arrest, prosecution, punishment, and victim services). See infra notes 144-290 and accompanying text.
118. DOBASH & DOBASH, supra note 8, at ix.
119. MACKINNON, supra note 100, at 170.
120. As one feminist acknowledged:
[From a feminist perspective, sexism is not just a factor in domestic violence. For feminists, gender is one of the fundamental organizing principles of society. It is a social relation that enters into and partially constitutes all other social relations and activities, and pervades the entire social context in which a person lives.]
Kurz, supra note 53, at 97.
any methodology employed to demonstrate the existence of female violence. By relying upon the definition of domestic violence as a patriarchal tool of control, any methodology which is not similarly grounded within this contextual framework can be rejected outright. Specifically, studies categorized as quantitative are deemed “inherently patriarchal” and therefore invalid because their dependence upon scientific, empirical data does not account for the history and context of male domination over women.121 By contrast, qualitative studies that depend upon a more clinical approach (which include the personal, in-depth interviewing of battered women) are considered to yield more accurate findings because of their fundamental commitment to a feminist perspective of domestic violence.122

At the risk of being an overly-crude distinction, the qualitative/quantitative dichotomy in the sociological study of domestic violence explains the criticisms of various CTS and non-CTS empirical studies which have reported on the violence of women.123 However, at a deeper level, the divide can also be appreciated as a predictable result of the liberal-feminist tension which cuts across all fields of study. Qualitative research easily links with liberalism and its commitment to neutral principles and abstract rules on the one hand, while quantitative research adheres to feminism and its emphasis upon personal relations and context on the other.124 The feminist sociologist’s critique of objective, positivist research as providing

121. Kersti Yllö, Political and Methodological Debates in Wife Abuse Research, in FEMINIST PERSPECTIVES ON WIFE ABUSE, supra note 55, at 31.

122. Having relied upon both qualitative and quantitative approaches in the study of domestic violence, sociologist Kersti Yllö speaks authoritatively on the philosophical division and argues for extracting and combining the strengths of each method. Her work includes, for example, a quantitative study demonstrating the relation between domestic violence and the social, economic, political and legal status of women. Id. at 30-36.

123. Like any simplistic distinction, a certain overlap between the quantitative and qualitative methodologies already exists through, for example, the quantitative study’s reliance upon personal interviews and existence of researcher biases and the qualitative study’s inevitable dependence on numeric data and objectifying of interviewees. See id. at 39-48 (discussing the strengths and weaknesses of the two approaches in domestic violence work).

“methodological tickets to scientific respectability but deliver[ing] intellectual blinkers and mindless adherence to sterile sophistication” while failing to account for the experiences of women125 resonates in the charges of radical legal feminist Catherine MacKinnon. “Liberal legalism is . . . a medium for making male dominance both invisible and legitimate by adopting the male point of view in law at the same time as it enforces that view on society.”126

Once this connection between research methods, definitions of domestic violence, and philosophical perspectives is made, the escalating costs to feminist theory can be well understood. However, the costs become even more real when the practical impact is included in the calculus. Defining domestic violence beyond the threat to women is seen in and of itself as a threat to women’s lives.

B. Female Violence: A Practical Threat

1. Financial Concerns

For real-world domestic violence advocates, defining domestic violence as a woman’s problem is a practical, not an academic, decision. In a world of infinite problems but finite resources, competition for adequate attention and funding is terrific. Recognition demands prioritization. By limiting the definition of domestic violence to male violence, domestic violence advocates have been able to frame the issue in a manner narrow and sympathetic enough for it to remain high on the public agenda. Broadening the definition to include female violence risks diluting the effectiveness of domestic violence funding campaigns, as female violence as well as male violence would then have to be targeted with, presumably, the same fixed amount of money.127 As the commentators honestly explain, given the “fierce competition” for funding,

[i]f we acknowledge the existence of battered husbands, then the funding designated for programs to assist battered women will be cut further because monies will be directed at programs for battered men. Thus, many radical feminists have fought for years to keep battered husbands closeted so that the small amount of money that was available for wife abuse would not be jeopardized.128

127. Kurz, supra note 53, at 99 (recognizing that perceiving women as batterers would require diverting money from battered women’s programs in order to assist battered men).
128. INTIMATE VIOLENCE, supra note 14, at 188.
Arguably, such practical consequences may be so great that even those willing to accept the violent tendencies of women may legitimately favor concentrating on the battered woman’s cause. The severity of the injuries suffered by battered women and the subordinate position traditionally held by women in the family are justifiable priorities.129 However, emphasizing the battered woman’s cause over the issue of battered men is far different than the current strategy that, instead, focuses on battered women by denying the existence of battered men. Practical motivations for denying the possibility of female violence altogether become more complicated.

2. Marketing Dilemmas

In domestic violence circles, the image of a battered woman as a helpless victim is routinely disavowed as overly simplistic and failing to account for the multitude of strengths shown by victims of violence.130 However, notwithstanding such theoretical posturing, the “helpless victim” stereotype remains very much a part of the common understanding of domestic violence and continues to fit well within the feminists’ overriding message of woman as subordinate.131 A woman’s ability to act aggressively challenges this simplicity.132 Consequently, when it comes to funding, battered women advocates are

129. As one of the authors of the CTS studies remarked:
Most people feel that social policy should be aimed at helping those who are in the weakest position. Even though wives are also violent, they are in the weaker, more vulnerable position in respect to violence in the family. This applies to both the physical, psychological, and economic aspects of things.
Straus et al., supra note 12, at 44.

For a comparison of the injuries suffered by men and women and the general acknowledgment that the injuries suffered by battered women are greater than those suffered by battered men, see supra Part II.C.


The helpless victim stereotype of battered women is built upon the notion of learned helplessness which suggests that a battered woman is unable to leave a violent relationship because she has been conditioned, over time, to be weak and passive in order to survive.


131. As part of her critique of the CTS studies, Yllö holds the studies’ researchers accountable for damaging the feminist agenda by releasing their findings. She openly argues that, as sociologists committed to ending domestic violence, they should have foreseen the controversy such statistics would unleash. In contrast to their non-contextual approach, Yllö openly praises feminists whose commitment to helping women determines what ideas and arguments are released into the “marketplace of ideas.” Yllö, supra note 121, at 42.

132. Recognizing that a woman’s ability to act violently does not refute the reality that women are battered is central to the challenge of acknowledging the violent capacities of women. For a discussion of how both the violent behaviors of men and women should be acknowledged and addressed, see infra Part V.
willing to promote a stereotype which, in other contexts, is readily recognized as inaccurate.

3. Vindictive Possibilities

Sympathetic to the interest in protecting women who are more likely than men to suffer injury, perhaps this strategic depiction of “woman as victim” has justification. However, not every motivation may be understood as so sincere. At least one author has argued that the treatment of female violence by feminists—be it to deny, defend, or minimize—is not a complicated matter at all. Rather, it is purely an act of revenge. Angry over a history of domination, feminists have discredited female violence in order to give women a secret way to strike back. From this perspective, the critiques of female violence are “not feminist critiques, but justifications of violence by women in the guise of feminism.” 133

IV. LEGAL IMPLICATIONS

A. Academic

From theoretical concerns, to financial and marketing strategies, to perhaps even vindictive desires, the treatment of domestic violence as yet another, if not the central, manifestation of our patriarchal society, has had definite social implications. 134 Given the natural impact of social norms on legal institutions, it is not surprising to find that the patriarchal definition of domestic violence has had a pervasive influence upon our legal system. Feminist scholars in the legal arena, like their counterparts in the social arena, have labored to identify and eradicate the law’s perpetuation of female oppression. Defining the law’s approach to domestic violence, as well as such issues as pornography, sexual harassment, and rape within the patriarchal framework, legal feminists share the more socially-oriented feminists’ objective of addressing each issue not simply as an isolated, unique phenomenon, but rather as part of an overarching scheme to eradicate the gender hierarchy and to empower women. 135 The patriarchal definition of domestic violence is integral to such a vision.

The domestic violence as gender violence characterization underlies the full array of legal scholarship on domestic violence. In practice-oriented scholarship, the patriarchal understanding of domestic violence is championed as a means of sharpening lawyering skills. 136

133. Straus, supra note 50, at 83.
134. See supra Part III.B (discussing such social motivations).
135. MACKINNON, supra note 100, at 22.
136. See, e.g., Meier, supra note 106, at 1325 (encouraging the adoption of an interdisciplinary approach in law school domestic violence clinics through the integration of psy-
In more theoretical scholarship, the patriarchal definition is urged as a means of shaping substantive legal issues. For example, the patriarchal definition is critical to distinguish criminal violence from the self-defense theory of battered women’s syndrome. Building upon the self-defense theory for women who kill, Professor Mahoney’s work on separation assault illustrates the fundamental male-power definition of domestic violence. Relying upon the separation assault theory and its realization of the greater risk of violence associated with leaving one’s abuser, Mahoney explains a woman’s use of violence not as a choice, but rather as the only means of escape. Mahoney’s separation assault theory defends women who resort to violence while providing a strong response to the ever-persistent question, “Why doesn’t she just leave?” Understanding domestic violence as simply another patriarchal means of “doing power” with the family, battered women are without the power to leave. Consequently, separation assault and its use in explaining battered women self-defense theories fits well within the subordination continuum experienced by women on account of our patriarchal familial and social structures.

Yet legal literature and its influence on such issues as battered woman’s syndrome is only one small example of the influence of feminist domestic violence theory on the law. Genderizing the use of domestic violence is openly and widely promoted as the means of achieving greater legal redress for battered wives from a full range of legal contexts. A critical examination of the criminal response to psychological and legal perspectives of domestic violence as an instrument of male power); Jane C. Murphy, Lawyering for Social Change: The Power of the Narrative in Domestic Violence Law Reform, 21 Hofstra L. Rev. 1243 (1993) (arguing for the greater use of narratives about women and people of color in the domestic violence arena in order to bring legal and social reform).

137. For a sampling of the literature linking the battered woman’s syndrome and the male use of domestic violence, see Walker, supra note 64, at 55-70; Meier, supra note 106, at 1316-17; Walker, supra note 64. For an earlier discussion of battered woman’s syndrome and its lack of statistical support from domestic violence studies see, supra notes 64-69 and accompanying text.

138. Mahoney, supra note 108, at 65-66. Separation assault is the attack on the woman’s body and volition in which her partner seeks to prevent her from leaving, retaliate for the separation, or force her to return. It aims at overbearing her will as to where and with whom she will live, and coercing her in order to enforce connection in a relationship. It is an attempt to gain, retain, or regain power in a relationship, or to punish the woman for ending the relationship. It often takes place over time.

139. For further discussion of the reasons women are traditionally seen as unable to leave a violent relationship, see McKenzie, supra note 96, at 53; Meier, supra note 106, at 1317-18.

140. Mahoney, supra note 108, at 93-94.

141. Id.

142. Such goals run the gamut—from the development of pro-arrest policies for male offenders, to increased criminal prosecutions of wife-beaters, to the provision of civil protective orders for battered wives, to criminal sanctions for violations of such orders, and to
domestic violence quickly reveals the success of such efforts. From the arrest, prosecution, and punishment of batterers, to the services offered to victims, all aspects of our legal response to domestic violence attest that domestic violence is controlled by a feminist state.143

B. The Feminist State Of Domestic Violence

1. Arrest


When domestic violence first began to receive public attention, mediation prevailed as the standard police response.144 Arrest, at least in the misdemeanor context, was the tool of “last resort.”145 Yet, recognizing that the power differential typically associated with domestic violence made mediation completely ineffective, by the late 1970s, women’s advocacy groups began lobbying for greater use of arrest.146 Today, while arrest is still not the predominant response, every state has adopted laws and policies which endorse the greater use of arrest.147
While the lobbying efforts of women's groups clearly influenced this development, the occurrence of three events in 1984 are often credited as being directly responsible for producing the dramatic turnaround in police response to domestic violence. The year was marked by the publication of the Minneapolis Experiment findings, the release of the U.S. Attorney General’s Family Violence Task Force recommendations, and the decision in Thurman v. City of Torrington. None of these three events conclusively found that arrest deters or prevents domestic violence. However, in unique ways, each event was nevertheless relied upon to promote arrest as the least costly response to domestic violence.

In studying the effectiveness of arrest as used by police in Minneapolis, the Minneapolis Experiment suggested that arrest served as a real deterrent to domestic violence. Based on their findings, the researchers recommended the adoption of warrantless arrest for the event of misdemeanor domestic violence such as simple assault and battery. David Hirschel & Ira W. Hutchison, The Relative Effects of Offense, Offender, and Victim Variables on the Decision to Prosecute Domestic Violence Cases, 7 Violence Against Women 46, 47 (2001); Sherman, supra note 144, at 15. By 1998, all fifty states authorized warrantless arrest for when probable cause exists to show that a misdemeanor domestic violence act has occurred or that a restraining order has been violated. Cheryl Hanna, The Paradox of Hope: The Crime and Punishment of Domestic Violence, 39 WM. & MARY L. REV. 1505, 1518 n.46 (1998). For a general discussion of changing arrest policies from 1970 to 1990, see Joan Zorza, The Criminal Law of Misdemeanor Domestic Violence, 1970-1990, 83 J. CRM. L. & CRIMINOLOGY 46 (1992).


149. ATT'Y GEN. TASK FORCE ON FAM. VIOLENCE: FINAL REP. (1984). For further discussion, see infra note 155 and accompanying text.

150. 595 F. Supp. 1521 (D. Conn. 1984). For a recognition of the combined influence of these events, see, for example, Nancy Egan, The Police Response to Spouse Abuse: A Selective, Annotated Bibliography, 91 LAW LIBR. J. 499, 502 (1999); Hanna, supra note 106, at 1859.

151. The study was based upon 314 cases in which the Minneapolis police had probable cause to believe that an incident of misdemeanor domestic violence had occurred. Sherman, supra note 144, at 18. However, rather than arresting in all cases, the study relied upon a lottery system in which police were randomly required to respond by either arresting the suspect, advising the parties, or sending the suspect away from the home upon threat of arrest. Id. at 16. In a six month follow up period, about 10% of the arrested suspects as compared to roughly 20% of the non-arrested suspects were reported to the police as having committed another act of domestic violence. Id. at 19. While not statistically significant, the roughly 20% of the non-arrested suspects who committed another act of domestic violence broke down to 19% of the advised suspects and 24% of the suspects who were sent away from the home. Id.

For the original report on the Minneapolis Experiment, see Sherman & Berk, supra note 148. For other later discussions, see Lawrence W. Sherman & Ellen G. Cohn, The Impact of Research on Legal Policy: The Minneapolis Domestic Violence Experiment, 23 L. & SOC'Y REV. 117 (1989).
policy. However, the researchers also recognized the limited nature of their study. They cautioned against mandatory arrest and encouraged replication studies. Five replication studies ultimately followed and were much more equivocal on the effectiveness of arrest as a domestic violence deterrent. Yet the replication studies had little practical impact. Ignoring the skepticism of the original experiment’s authors and without waiting for the further studies, the U.S. Attorney General released his recommendation in favor of warrantless arrest and arrest as the “preferred response” within four months of the Minneapolis Experiment.

The Minneapolis Experiment and the replication studies were evenly divided in their results—one half finding arrest to be a deterrent and the other half finding arrest to be a promoter of domestic violence. However, even the studies’ full release and review did little to sway many domestic violence advocates from their seemingly intuitive commitment to some policy of mandatory, presumptive, or

152. Sherman, supra note 144, at 21-22. For discussion of the eventual adoption of warrantless arrest by all the states, see Hanna, supra note 147, at 1518 n.46. See also supra note 147 and accompanying text (discussing state developments in arrest policies).

153. Sherman, supra note 144, at 21-22.

154. Like the Minneapolis Experiment, the replication studies were also sponsored by the National Institute of Justice. Id. at 25. The five studies were conducted in Milwaukee, Omaha, Charlotte, Colorado Springs, and Miami (Metro-Dade). Only the two experiments undertaken in Colorado Springs and Miami concluded that arrest had a deterrent effect as the original experiment had found. Id. at 25, 32. The Milwaukee, Omaha, and Charlotte studies found that arrest correlated with an escalation in violence. Id. at 25. For a brief comparison of the initial Minneapolis Experiment, with the five replication studies and their varying methodologies, see JEFFREY FAGAN, THE CRIMINALIZATION OF DOMESTIC VIOLENCE: PROMISES AND LIMITS 13-15 (1996); Hirschel & Hutchison, supra note 147, at 46; Lisa G. Lerman, The Decontextualization of Domestic Violence, 83 J. CRIM. L. & CRIMINOLOGY 217, 222-38 (1992); Schmidt & Sherman, supra note 147, at 44-52; Sherman, supra note 144, at 25.


156. See supra note 151 and accompanying text.
preferred arrest.\textsuperscript{157} Today, the support for arrest remains firm despite the recognition that mediation between the parties continues to be the more likely police response.\textsuperscript{158} Rather than simply relying upon arrest to end domestic violence, arrest is argued to be the first critical link in a “coordinated response” of various criminal and social agencies which must respond collectively and aggressively against domestic violence.\textsuperscript{159} Emphasizing the need to send a message of community intolerance, arrest supporters promote the importance of arrest as a general deterrent to domestic violence despite its failure as a deterrent in specific instances.\textsuperscript{160} Objectives other than deterrence are also advanced. Just as the effectiveness of arrest is not considered in determining whether or not to arrest in non-domestic situations, pro-arrest advocates argue that the effectiveness of arrest is not a relevant consideration in the domestic violence context.\textsuperscript{161} Arrest is favored because it marks a “get tough” attitude toward domestic violence.\textsuperscript{162} Pro-arrest policies consequently take on a symbolic

\textsuperscript{157} In commenting on reactions to the various studies, a principal author of the Minneapolis experiment found that, “[t]he Minneapolis findings stirred enormous interest by a wide range of writers and editorialists, who hailed the results as a breakthrough. The replication results received grudging acceptance in some of those quarters, and complete silence in most others.” Sherman, supra note 144, at 44-45 (citations omitted). While advocates in support of arrest have held this position despite the studies’ conflicting findings, this belief seems to be supported by more recent statistical analysis of all six studies which found that arrest was associated with lower rates of repeat violence through a variety of measures. See Christopher D. Maxwell et al., \textit{The Effects of Arrest on Intimate Partner Violence: New Evidence from the Spouse Assault Replication Program}, NAT’L INST. JUST.: RES. BRIEF, July 2001, at 1, 9.

For discussion of the opposition to arrest, see infra notes 164-69 and accompanying text.

\textsuperscript{158} A more recent national study based upon data drawn from the 1992 through 1994 National Crime Victimization Surveys (NCVS) on personal and household victimization similarly concluded that police are less likely to arrest in cases of intimate assault than in other assault cases. While the study was limited to male-on-female violence, it also reported that in cases of intimate violence the probability of arrest increased: (1) with the increased age of the victims or offenders; (2) when the victim is white; (3) when the offender is black; (4) when the offender is under the influence of drugs or alcohol; (5) when weapons are involved; and (6) with the increased wealth of the victim. Edem F. Avakame & James J. Fyfe, \textit{Differential Police Treatment of Male-on-Female Spousal Violence: Additional Evidence on the Leniency Thesis}, 7 VIOLENCE AGAINST WOMEN 22, 29-34 (2001); see also Hanna, supra note 147, at 1516-19 (noting that the inclination against arrest in the domestic violence context appears to continue although pro-arrest policies are increasing).


\textsuperscript{160} Bowman, supra note 116, at 206; Frisch, supra note 159, at 209; J. David Hirschen & Ira W. Hutchison, \textit{Realities and Implications of the Charlotte Spousal Abuse Experiment, in DO ARRESTS AND RESTRAINING ORDERS WORK?}, supra note 147, at 54, 80; Lerman, supra note 154, at 225; Daniel D. Polsby, \textit{Suppressing Domestic Violence with Law Reforms}, 83 J. CRIM. L. & CRIMINOLOGY 250, 252 (1992); Zorza, supra note 147, at 72.

\textsuperscript{161} Frisch, supra note 159, at 213; Lerman, supra note 154, at 224.

\textsuperscript{162} Sherman, supra note 144, at 45. For a direct criticism of criminal efforts in the domestic violence context which may have only symbolic value, see Hanna, supra note 147.
importance, for they reject the traditional treatment of domestic violence as a private matter and openly acknowledge domestic violence as a public crime. In this respect, the pro-arrest position of many feminists can be easily understood as it is consistent with the broader campaign of feminists to “explode” the private/public dichotomy which has historically been used to relegate women’s issues to the private, thereby preventing any legal response.\textsuperscript{163}

While support for arrest is strong, the inconsistent nature of the studies on arrest has resulted in some opposition to arrest. Not all domestic violence scholars unequivocally support arrest. Ironically, today’s most notable critics of arrest are the authors of the original Minneapolis Experiment. While still in favor of warrantless arrest, the researchers are strongly against mandatory arrest laws. Believing mandatory arrest policies only decrease violence when implemented against persons without strong social connections that could be jeopardized by arrest, the researchers go so far as to recommend the repeal of mandatory arrest laws in communities with “substantial ghetto poverty populations with high unemployment rates,” and to suggest returning to greater police discretion combined with increased reliance on such non-criminal responses as shelters and treatment programs for victims and offenders.\textsuperscript{164} Feminists concerned with the limited value of arrest can also be found. Some feminists critique the emphasis on arrest for its failure to address the “sexual inequality, coercive control, and entrapment” dynamics surrounding

\textsuperscript{163} MACKINNON, supra note 100, at 100. The private/public domestic violence dichotomy and the effort to treat domestic violence as a public act has received terrific attention. See, e.g., Honorable Karen Bunstein, Naming the Violence: Destroying the Myth, 58 ALB. L. REV. 961, 964-65 (1995) (arguing the failure to recognize domestic violence as a public issue allows for its tolerance); Hanna, supra note 106, at 1869-77, 1907 (discussing the feminist public/private dichotomy and aggressive prosecution of batterers as a means of recognizing domestic violence as a public crime rather than a private matter); Hanna, supra note 147, at 1509 (furthering arguments for aggressive action against batterers through a discussion of stiffer sentencing standards); Mahoney, supra note 108, at 11-13 (acknowledging society’s denial of domestic violence); Schneider, supra note 116, at 1250-51 (discussing the need to reconceptualize domestic violence as a social problem in need of a public solution); Schneider, supra note 108, at 645-48 (outlining the growing treatment of domestic violence as a public harm); Seymore, supra note 116, at 1070-73 (discussing the spousal immunity doctrine in light of the private versus public domestic violence debate); Siegel, supra note 116, at 2154 (discussing the struggle against marital violence being regarded as a right of privacy and the “modernization” of arguments which allow for the perpetuation of domestic violence); Laura W. Stein, Living with the Risk of Backfire: A Response to the Feminist Critiques of Privacy and Equality, 77 MINN. L. REV. 1153 (1993) (arguing in favor of a limited sphere of privacy in order to enhance autonomy and other values).

\textsuperscript{164} Schmidt & Sherman, supra note 147, at 51. For a full discussion of such class concerns, see Lawrence W. Sherman et al., From Initial Deterrence to Long-Term Escalation: Short-Custody Arrest for Poverty Ghetto Domestic Violence, 29 CRIMINOLOGY 821 (1991).
domestic violence.165 In an effort to maintain domestic violence characterized as a crime of patriarchy, feminist critics also oppose more stringent arrest policies as they believe female victims of domestic violence will end up the ones at greater risk of arrest.166 Other victim-related concerns are also raised. The use of aggressive arrest policies are feared likely to backfire in communities where cultural and social norms dictate against taking any highly visible public action against one’s abuser, thus discouraging a victim who might otherwise call the police for limited counseling and assistance from calling at all.167 Similarly, aggressive arrest policies are criticized for their risk of discouraging victims from seeking police assistance when they believe arrest could threaten the batterer’s employment, thereby threatening the victim’s and family’s economic stability.168 In the most general sense, by ignoring the individual victim’s unique situation and feelings regarding personal safety, and the effectiveness of arrest, “forceful” legal responses such as mandatory arrest may legitimately be characterized as “revictimizing” the victim.169


166. In raising such a concern, Linda Mills relied upon a Los Angeles study, which found that while the number of men arrested doubled, the number of women arrested quadrupled with the implementation of mandatory arrest. LINDA G. MILLS, THE HEART OF INTIMATE ABUSE: NEW INTERVENTIONS IN CHILD WELFARE, CRIMINAL JUSTICE, AND HEALTH SETTINGS 50 (Albert R. Roberts ed., 1998). Interestingly, the statistics were relied upon without any effort by Mills to determine whether the women might have been properly arrested for abusive behavior. See id. A similar study noting the increase in dual arrest and the risk such arrest presents to battered women was also conducted in Connecticut. See Margaret E. Martin, Double Your Trouble: Dual Arrest in Family Violence, 12 J. FAM. VIOLENCE 139 (1997).

The failure of such criticisms to take into account the possibility that the women arrested for violence may not always simply be victims is consistent with this Article’s general premise that women are implicitly believed not capable of committing domestic violence.


168. On the acknowledgment of the economic considerations that may deter a victim from seeking assistance, see, for example, Hirschel & Hutchison, supra note 147, at 47. As a corollary, it is also acknowledged that victims may have a difficult time finding or maintaining employment due to such factors as having limited child care alternatives, showing poor employment histories due to prolonged domestic violence, and suffering from low self-esteem or post traumatic stress disorder. Davis & Kraham, supra note 108, at 1150-55.

169. For revictimization concerns brought on by a state’s unilateral decision to prosecute, see Bowman, supra note 116, at 203-05; Hanna, supra note 106, at 1884. For related discussions, see Kelly, supra note 108, at 321 (discussing revictimization in federal government’s decision to deport victim’s batterer); Meier, supra note 106, at 1333-34 (discussing the need for domestic violence advocates to take care to avoid “usurping [the battered woman’s] autonomy and decisionmaking”).

While the fear of increased violence may be specifically associated with arrest, it is more generally associated with any efforts made by the victim to leave or separate from the abuser. For discussion of separation assault, see supra note 138 and accompanying text.
In setting out this variety of opinions, the effort here is not to resolve the debate over the efficacy of arrest. Rather, regardless of the position on arrest one may ultimately adopt, the objective here is simply to reveal the gender biases running throughout police practices and legal standards of liability, and to argue that such biases must be acknowledged in any sincere effort to effectively utilize arrest.

(b) The Gendered Promotion of Arrest

When it comes to arrest, female domestic violence offenders are simply not considered. As one of the arrest replication experiments candidly noted, female offenders were not included in the study because they were considered not to pose any danger. Their exclusion from the study clearly implied that arrest could be used in a more discretionary manner by the police when it came to female offenders than it could be in the case of male offenders. Such an assertion seems glaringly inconsistent with the experiment’s authors’ later conclusion that despite any inconclusive findings on the effectiveness of arrest, arrest remains a “more conscionable choice than non-arrest” because of the desire to communicate to men, women and children that domestic violence is unacceptable. While such remarks and oversights might be disregarded as mere political commentary, other evidence clearly demonstrates that police ignore the violence perpetrated by women and that such inaction is judicially endorsed.

(c) The Statistics

As part of the second National Family Violence Survey, researchers included a series of questions directed at men and women who had been involved in some form of intimate violence in which they had called the police for assistance. While arrest was gaining ac-

---

On concerns regarding the safety of a victim’s children upon separation, see supra note 108 and accompanying text.


171. Id.

172. These remarks were made in conjunction with the Charlotte replication experiment which found that arrest did not deter domestic violence. Id. at 80 (quoting J.D. HIRSCHEL ET AL., CHARLOTTE SPOUSE ASSAULT REPLICATION PROJECT: FINAL REPORT, 159-69 (1991). For a discussion of the Charlotte replication study and the other arrest experiments, see supra note 154 and accompanying text. For other similar support for arrest despite the inconclusive nature of the arrest studies, see supra notes 155-57 and accompanying text.

173. Of the over 6,000 individuals and over 4,000 families interviewed as part of this survey in 1985, approximately 800 individuals had been involved in some form of intimate violence. INTIMATE VIOLENCE, supra note 14, at 170. Of that number approximately one in twenty households (4.8%) had called the police. Id. Of these callers, seventeen were men.
ceptance during the time of the study (as remains true today), mediation efforts such as trying to calm everyone or taking time to listen to the story remained the “typical form of intervention used by [the] police” in the cases studied. Yet, in comparing the use of more forceful police action when women called as opposed to when men called, police were much more likely to respond aggressively when the caller was a woman. When women called, the male spouse was ordered out of the home 41.4% of the time. However, in no cases were female spouses ordered to leave the home when men called. Similarly, when women called, the men were threatened with arrest at the time of the incident in 10.7% of the cases, and threatened with arrest upon the next call in 28.2% of the cases. By contrast, women were never threatened with arrest when men called. Finally, in comparing the actual use of arrest, while in 15.2% of the cases the man was arrested upon the woman’s call, no woman was ever arrested when a man called. In fact, it was three times more likely that a man would be arrested if he called as opposed to the female caller being arrested. Not surprisingly, researchers found male callers to be generally “less pleased” with the police response than the female callers, who described the police intervention to be “about right.”

Given that the same survey found a striking similarity in the use of intimate violence by men and women, such statistics cannot easily be explained by the suggestion that the male callers were likely to have suffered more minimal acts of violence. Indeed, given society’s denial of the reality of female violence and the unmanly stigma attached to any man who claims to be a victim of such violence, men are much less likely than women to call for help, choosing instead to react “in private.”

and twenty-four were women. Id. For a discussion of the 1985 and other National Family Violence Surveys, see supra notes 21-23 and accompanying text.

174. INTIMATE VIOLENCE, supra note 14, at 170; see also supra note 158 and accompanying text (noting mediation remains the predominant response to domestic violence).

175. INTIMATE VIOLENCE, supra note 14, at 262.

176. Id.

177. Id.

178. Id.

179. Id.

180. As opposed to 4.2% of the women callers being arrested, 12.1% of the male callers were arrested. Id.

181. Id. at 170-71. For a recent study of female victims of intimate violence and the factors which affect their satisfaction with various components of the legal system when their batterers have been arrested, see Fleury, supra note 116, at 191.

182. For a comparison of the use of violence by men and women as uncovered by the second National Survey of Family Violence and other studies, see supra notes 173-80 and accompanying text.

183. In the 1985 National Survey of Family Violence, researchers found the most typical male reaction to being hit was to yell, followed by running out of the room. INTIMATE VIOLENCE, supra note 14, at 150. Less than fifteen in one hundred men said they reacted by striking back. Id. Female victims’ most common reaction was to cry, followed by yelling.
ers surveyed were victims of greater violence than the female callers as only severe violence would be threatening enough for a man to call. Taking such considerations into account, what such statistics do seem to reflect is that police share the accepted view about female violence: it does not exist.

(d)  
Thurman v. City of Torrington

In addition to the police, the feminist understanding of domestic violence is also adhered to and reinforced by the courts. Based upon Equal Protection claims, arguments for police liability have successfully been made when the police fail to adequately respond in the case of intimate violence. However, a review of the court decisions in this area suggests that the standard of liability extends only to cases involving female victims of intimate violence. The critical 1984 decision in Thurman v. City of Torrington set the tone, adding the judicial branch to the increasingly pervasive number of institutions that ignore female violence.

The facts of Tracey Thurman’s case against the City of Torrington and its police department are equally horrific and sympathetic. For nearly nine months, Tracey was “attacked” and “threatened” by her estranged husband, Charles, on numerous occasions. Tracey repeatedly requested the assistance of the city police, asking them to arrest her husband. In turn, the police repeatedly refused to take any action. Finally, the escalating violence culminated in Charles stabbing Tracey and then, in the presence of the police, dropping the knife, kicking Tracey once in the head, running indoors and seizing their child, dropping the child on his wounded mother and then kicking Tracey in the head once again. Despite several police officers witnessing such brutal events, Charles was arrested only after Tracey was placed on a stretcher and Charles, remaining on the scene and continuing to threaten Tracey, approached Tracey once again as she lay on the stretcher.

Id. at 147-48. Approximately one in four women responded by striking back. Id. For further discussion of male reaction to female violence and society’s denial of such violence, see Steinmetz, supra note 6, at 503-04; see also supra notes 127-32 and accompanying text and infra notes 284-89 and accompanying text.

184. See infra Part IV.B.2(a).


186. After her car windshield was smashed by her husband in the presence of a police officer (who had stood by watching Charles scream threats at Tracey prior to smashing the window), Charles had been arrested, convicted of breaching the peace, and placed on probation. Charles’s ongoing menacing contact constituted both a violation of his probation and perhaps a violation of the restraining order obtained by Tracey, thereby providing sufficient basis to re-arrest Charles. Id. at 1524-26.

187. Id.
In her § 1983 action, Tracey successfully argued in response to the city's motion for summary judgment that the police department's inaction violated her Fourteenth Amendment guarantee of equal protection.\footnote{188}

As noted earlier, the \textit{Thurman} decision is routinely cited as a key catalyst in the national adoption of mandatory arrest policies.\footnote{189} For more practical individuals who believe that the purpose of police department policy changes are not to increase public safety but rather to avoid department liability, \textit{Thurman} is cited as the only motivator.\footnote{190} With this degree of significance, \textit{Thurman}'s gender implications must be candidly acknowledged.

In evaluating the Equal Protection claim, \textit{Thurman} described the analysis as a comparison of police treatment of “a woman abused or assaulted by a spouse or boyfriend” with police treatment of “all other persons whose personal safety is threatened, including women not involved in domestic relationships.”\footnote{191} Given this gendered nuance, \textit{Thurman} was afforded intermediate scrutiny.\footnote{192} \textit{Thurman} then held in favor of the plaintiff as the government failed to meet its burden of citing “an important governmental interest” which would jus-
tify “its disparate treatment of women.” Recognizing that wife beating was historically a male prerogative, the court denounced such a practice, finding that it must “join other ‘archaic and overbroad’ premises” which had been traditionally used to condone other acts of gender discrimination.

While not disputing the importance of a non-discriminatory police response to victims of violence, the Equal Protection framework utilized to advance a fair treatment policy must be critically examined. Why did the court compare the treatment of female victims of intimate violence by male partners versus “all others”? Why did the court not simply compare the treatment of victims of intimate violence versus the victims of stranger violence, regardless of gender? Relegated to a footnote, the court responded to such questions. It would accept plaintiff’s allegation of gender-discrimination as true. In so doing, the court cited to only one study which reviewed the cases of criminal defendants charged with intimate violence and found that men were the defendants in twenty-nine out of every thirty cases. Given the nonrepresentative nature of criminal studies which fail to measure the full scope of victims and the general concern regarding prosecutorial studies, the soundness of the court’s decision must be challenged. Yet what is more troubling is the deeper implications of the court’s findings. By adopting a gender-related standard, the court was able to employ an intermediate scrutiny analysis, thus making the government’s justification for its arrest policy more difficult. However, in finding that the government had “failed to put forward any justification” for its arrest policy, it is likely that the government would have failed even if Thurman had adopted a genderless standard, and thereby, compared the police treatment of intimate versus stranger violence at the rational basis level of review. Arguably then, if the court’s interest was ensuring

193. Id. at 1527-28.
194. Id. at 1528 (quoting Craig, 429 U.S. at 198-99). The court also included other recognized examples as gender discrimination. See Stanton v. Stanton, 421 U.S. 7 (1975) (denying the traditional belief that a woman’s place is in the home); Weinberger v. Wiesenfeld, 420 U.S. 636, 643 (1975) (rejecting the misconception that female wages do not contribute to family support); Frontiero v. Richardson, 411 U.S. 677, 689 (1973) (dismissing the belief that husbands of female servicewomen are not dependent on their spouses as compared to wives of male servicemen); Crawford v. Cushman, 531 F.2d 1114 (2d Cir. 1976) (preventing the denial of a pregnant woman’s ability to serve in the military).
196. Id. (citing Stephen Leeds, Family Offense Cases in the Family Court System: A Statistical Description ii (1978)).
197. For a critical discussion of the criminal studies on domestic violence and their use in disputing the family violence studies, see supra Part I.A.1.
198. Thurman, 595 F. Supp. at 1528. As acknowledged by Thurman, at this lowest level of scrutiny, classifications must be “rationally related to a legitimate governmental purpose.” Id. at 1526 (citing San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 55, reh’g denied, 411 U.S. 959 (1973)).
police action in domestic violence cases, it could have achieved this goal without relying on a gender-based analysis. Instead, by so readily relying upon the assertion that only women are victims of domestic violence, the court revealed its own “‘archaic and overbroad’ premises.” While the court was quick to reject stereotypes negatively affecting women, it was not ready to do the same for men. By failing to leave room for the possibility of female violence and that police may also fail to arrest in such cases, Thurman set a standard of liability only in the instance of the police’s failure to arrest male perpetrators.

Following Thurman, the intermediate scrutiny standard for claims made by female domestic violence victims against the police has broadened. Police policies or customs which are facially neutral may be subject to intermediate scrutiny if such policies are applied in a discriminatory fashion which adversely affects battered women. Consequently, in many jurisdictions, proving a police policy or custom providing less protection to victims of domestic violence than nondomestic violence attacks is apparently not sufficient to win an equal protection claim. The plaintiff further has to demonstrate “that discrimination against women was a motivating factor,” and that the plaintiff was indeed injured by such practice.

199. Id. at 1528 (quoting Craig, 429 U.S. at 198-99).


201. Id. at 1031-33 (emphasis added) (vacating the denial of defendant’s motion for summary judgment but remanding action for further proceedings in light of that standard). Interestingly, Hynson’s three part test was based upon Watson v. City of Kansas City, 857 F.2d 690 (10th Cir. 1988). Watson set out two possible equal protection grounds for battered women raising § 1983 claims. Id. at 696. While the first considers domestic violence victims without regard to gender, the second considers gender. Id. Applying these standards, Watson reversed a summary judgment in favor of the defendant city regarding plaintiff’s claim of discriminatory treatment against victims of domestic violence as compared to victims of nondomestic attacks. The court reasoned that a jury could infer discriminatory motive, but found insufficient evidence to reverse summary judgment in favor of the city as to the plaintiff’s claim of class-based discrimination based on sex. Id. at 694-97.

Other cases have adopted a gender-based, intermediate scrutiny analysis. See generally Shipp v. McMahon, 234 F.3d 907 (5th Cir. 2001) (adopting Watson standard to review gender-based equal protection claims against law enforcement in domestic violence cases); Balistreri v. Pacifica Police Dept, 901 F.2d 696, 701 (9th Cir. 1990) (reversing the dismissal of plaintiff’s case regarding police’s failure to arrest male perpetrators of domestic violence); McKee v. City of Rockwall, 877 F.2d 409, 416 (5th Cir. 1989) (adopting gender-based discrimination standard but dismissing plaintiff’s claim as a “complete failure” of proof to support allegation of discrimination against victims of domestic violence and finding it therefore unnecessary to reach allegation of police following policy of intentional discrimination against women).

For further discussion of the equal protection and due process standards relied upon in the review of § 1983 claims by battered women, see generally Hathaway, supra note 188, at 667-77 (for a discussion of Thurman, 595 F. Supp. 1521); Laura S. Harper, Note, Battered Women Suing Police for Failure to Intervene: Viable Legal Avenues After Deshaney v.
As a practical matter, such a standard communicates that police liability can only result in cases of nonintervention on behalf of female victims. By tailoring the standard exclusively to battered women, the courts have further reinforced the belief that only women can be the victims of domestic violence. The effect of Thurman and its progeny has been to train police to respond aggressively in the case of male aggressors while further discounting the complaints of female violence. Thurman, then, should not be acknowledged as responsible for the national adoption of mandatory arrest policies in domestic violence cases. Thurman should be credited for the development of mandatory arrest policies only in the case of male perpetrators and female victims.

2. Prosecution

(a) The Development of Prosecution Policies and Theories

Recognizing that a “coordinated response” is the only means by which a definitive effort to end domestic violence can truly be made, prosecution is regarded as the critical next step following arrest.202 Echoing the justifications for pro-arrest policies, aggressive prosecution is favored as part of an overarching effort to criminalize domestic violence and make it a public crime.203 The strength of this ambition is well illustrated through the recent enactment of the Violence

Winnebago County Department of Social Services, 75 CORNELL L. REV. 1393 (1990) (analyzing the equal protection and due process claims of battered women under § 1983 after the Supreme Court held that there was no general constitutional right to police protection in DeShaney v. Winnebago County Department of Social Services, 489 U.S. 189, 191 (1989)); Kalyani Robbins, Note, No-Drop Prosecution of Domestic Violence: Just Good Policy, or Equal Protection Mandate?, 52 STAN. L. REV. 205, 223-30 (1999).

For the recognition that heightened scrutiny is “not necessary” and that a rational basis test can be successfully utilized when evaluating police failure-to-protect claims made by victims of domestic violence as compared to victims of other crimes of violence, see id. at 227-28 (citing to the successful use of a rational basis review by victims of domestic violence in Bartalone v. County of Berrien, 643 F. Supp. 574 (W.D. Mich. 1986)); Lauren L. McFarlane, Note, Domestic Violence Victims v. Municipalities: Who Pays When the Police Will Not Respond?, 41 CASE W. RES. L. REV. 929, 950-51 (1991) (citing City of Cleburne v. Cleburne Living Ctr., Inc., 473 U.S. 432, 448 (1985) (supporting the argument that “negative attitudes” about victims of domestic violence should be sufficient to find police failure to protect victims of domestic violence based upon a nondeferential rational basis standard)).

202. For the promotion of a “coordinated response” to domestic violence between various legal and social agencies, see supra note 159 and accompanying text. For a comparison of arrest and prosecution efforts, see, for example, BUZAWA & BUZAWA, supra note 142, at 54 (recognizing that problems associated with addressing domestic violence in the arrest context are repeated and become even more severe in the judicial context); Naomi R. Cahn & Lisa G. Lerman, Prosecuting Woman Abuse, in WOMAN BATTERING: POLICY RESPONSES, supra note 116, at 95, 96 (lamenting that the increase in preferred arrest policies is not matched by increased prosecution); Ford & Regoli, supra note 155, at 129 (advocating the maintenance of advances in prosecution on par with increased arrest efforts).

203. For a discussion of the support of pro-arrest policies, see supra notes 160-62 and accompanying text.
Against Women Act (VAWA). The Act aggressively attempts to: mandate the arrest and prosecution of domestic violence offenders; devote greater resources to domestic violence training for law enforcement; make certain domestic violence offenses federal crimes; and provide various civil remedies for victims. However, as with arrest, support for strong prosecution policies is neither universal nor longstanding. VAWA provisions, such as those providing certain federal civil remedies for intrastate acts of domestic violence, have been struck down as unconstitutional. Such resistance is also visible at the state level, where state prosecution of domestic violence offenders, although increasing, is increasing at a very gradual rate. Like arrest, this slow development of aggressive prosecution policies illustrates conflicting feminist goals. And again, as with arrest, any consensus which can be found in support of prosecution reveals only an interest in prosecuting male offenders.

A long list of reasons can be given for the traditional reluctance to prosecute domestic violence. Included among the most objective reasons is that criminal dockets are already overwhelmed by non-domestic cases and cannot absorb the flood of cases which would be introduced by criminalizing all forms of violence between intimates. Decisions then must be made as to which domestic violence cases, if any, should be prosecuted. This may result in prosecutors...


207. In 1996, a survey of prosecutorial developments in medium and large jurisdictions determined there was a “growing commitment” to “vigorous prosecution.” Donald J. Rebovich, Prosecution Response to Domestic Violence: Results of a Survey of Large Jurisdictions, in DO ARRESTS AND RESTRAINING ORDERS WORK?, supra note 96, at 176, 188. For further discussion of prosecutorial developments, see infra notes 216-19 and accompanying text.

208. For discussion of the conflicts in arrest theory, see supra notes 164-69 and accompanying text.

209. In this respect, while the provisions of VAWA are gender-neutral, its title similarly reflects a singular interest at the federal level in addressing only violence against women. For a discussion of VAWA, see supra notes 204-05 and accompanying text.

210. For lengthy discussions of the variety of such reasons and their connection, see, for example, BUZAWA & BUZAWA, supra note 142, at 56-62; FAGAN, supra note 154, at 27-29; Cahn & Lerman, supra note 202, at 96; Ford & Regoli, supra note 155, at 132-43; Hanna, supra note 106, at 1860-63; Hanna, supra note 147, at 1551-54; Hirschel & Hutchinson, supra note 147; Rebovich, supra note 207, at 176-77.

211. BUZAWA & BUZAWA, supra note 142, at 56-58; FAGAN, supra note 154, at 27-28; Hanna, supra note 147, at 1553-54.
strategically deciding only to pursue cases based on legitimate considerations of whether or not they will be successful given the extent of injury and available evidence.\textsuperscript{212} Yet prosecutors, judges, and other court personnel are also recognized to harbor the same social biases that prevent treating domestic violence as anything other than a minor, private matter in which the victim is as much to blame as the abuser.\textsuperscript{213} Such attitudes prevent vigorous prosecution. A victim's own reluctance to prosecute may also be a product of such biases.\textsuperscript{214} Victims may be further dissuaded by fear of the batterer's retaliation, ongoing love of the batterer, economic considerations, and other complicated feelings including what may be best for any children involved.\textsuperscript{215}

Despite these challenges, the strong push to criminalize domestic violence has resulted in a variety of policies and programs intended to increase prosecution.\textsuperscript{216} Among the most prominent are such innovations as no-drop policies limiting prosecutorial discretion,\textsuperscript{217} victim support services within both police and prosecutors' offices,\textsuperscript{218} and specialized domestic violence prosecution units and courts.\textsuperscript{219} Yet even with such developments, prosecution is still recognized as a "relatively rare event."\textsuperscript{220}

\begin{footnotesize}
\begin{enumerate}
\itemstrategicall\end{enumerate}
\end{footnotesize}
(b) The Gendered Promotion of Prosecution

Given that such formal administrative and policy changes have not yielded terrific change, domestic violence prosecution debates now center around what is commonly recognized as the greatest stumbling block to prosecution—victim reluctance.221 Some academics advocate even more aggressive use of no-drop policies which would take the controversial step of forcing victim participation when necessary.222 Grounded in theory and practice, such a position is argued as best suited for ensuring victim safety and effective prosecution as well as sending the critical message that domestic violence is not tolerated.223 Other advocates staunchly defend a victim’s right to choose. Yet in so doing, they also claim reliance upon theoretical and empirical studies to argue that prosecution only effectively deters domestic violence when a victim retains the right to drop the charges.224

221. See, e.g., BUZAWA & BUZAWA, supra note 142, at 58-60 (citing domestic violence victim attrition rates at 60% to 80% while recognizing that prosecutors often further discourage victims to prosecute); Ford & Regoli, supra note 155, at 141-43 (blaming the victim’s desire to drop charges on outside forces); Hanna, supra note 147, at 1551 n.191 (citing victim reluctance to go forward as responsible for lack of prosecution of domestic violence crimes).

222. Taking this step only at the most extreme instance, Professor Hanna also argues strongly in favor of improved efforts to gather and use various forms of nontestimonial evidence that can be obtained from such physical evidence as police investigations, 911 calls, and medical records. Hanna, supra note 106, at 1898-1919.

223. In addition to her own personal experience as a prosecutor of domestic violence cases, Professor Hanna points to a study in San Diego which credited the reduction in domestic violence related homicides to the adoption of a no-drop policy. Id. at 1853, 1864 n.64, 1867 (arguing in favor of “hard” no-drop policies and relying on Mark Hansen, New Strategy in Battering Cases: About a Third of Jurisdictions Prosecute without Victim’s Testimony, 81 A.B.A. J., Aug. 1995, at 14); see also Robbins, supra note 201, at 216-17 n.77 (relying on same San Diego homicide study). In a later work, Professor Hanna further defends her position in favor of aggressive prosecution. In summary, she argues that giving decision-making power to the prosecutor rather than to the victim is the best means to prevent intimidation and reinforce the notion of domestic violence as a public crime, not a private matter. Hanna, supra note 147, at 1506 n.2.

224. The only randomized field experiment to evaluate the deterrence of domestic violence as a result of various prosecution policies was conducted in reliance upon 678 cases of wife battery in Marion County (primarily Indianapolis), Indiana. With the assistance of the prosecutor’s office, the cases were randomly assigned by the prosecutor to four “prosecution tracks” (no prosecution, pretrial diversion, prosecution with rehabilitation, or prosecution with more severe sanctions). These cases included both instances in which the victim was allowed to drop the charges and those in which she was prevented from doing so. The study concluded that regardless of the prosecution track initiated by the prosecutor’s office, rates of recidivism only decreased significantly in the cases in which the victim had the choice to proceed and chose to proceed. The study also cautioned that victims who had chosen to proceed but who ultimately dropped the charges were at increased risk of violence. Ford & Regoli, supra note 155, at 150-53; David A. Ford & Mary Jean Regoli, The Preventive Impacts of Policies for Prosecuting Wife Batterers, in DOMESTIC VIOLENCE: THE CRIMINAL JUSTICE RESPONSE, supra note 142, at 181, 192-205.

For recognition of the Indianapolis experiment and support of providing domestic violence victims with varying degrees of control in criminal proceedings, see, for example,
Such a review of the current developments in domestic violence prosecution indicates that the core controversy remains very much tied to a core tension in feminist theory—a desire to make domestic violence a public crime on the one hand and efforts to preserve female autonomy on the other. What is missing from the debate is any conversation regarding the prosecution of female violence. With the existing discourse rooted in feminist theory, it is virtually impossible for such a subject to be introduced, much less entertained. Regardless of their conflicting conclusions, the few recognized empirical studies which do exist on the issue of effectiveness of prosecution are limited to the study of male batterers. In all discussions of prosecution, it is either assumed, implied, or outright stated that the sole objective is to help battered women. Such comments are made even by scholars who criticize the minimal research and stagnant debates on prosecution and call for more global and radical efforts in order to end domestic violence. Again, the efforts to end domestic violence

225. In her discussion of prosecution efforts, Professor Hanna acknowledges this tension through critical examination of the public/private dichotomy. She further considers the tensions engendered by feminist theory on themes of particularity/generality and agency/victimization. Hanna, supra note 106, at 1869-85.

226. While there are generally recognized to be just two principal studies on the effectiveness of prosecution, only the Indianapolis experiment attempts to determine if prosecution reduces the recidivism rates of violent men. On the Indianapolis experiment, see Ford & Regoli, supra note 155, and Ford & Regoli, supra note 224. The other principal prosecution study examines what variables influence prosecutorial discretion in the decision to prosecute abusive men. Janell Schmidt & Ellen Hochstedler Steury, Prosecutorial Discretion in Filing Charges in Domestic Violence Cases, 27 CRIMINOLOGY 487 (1989). A more recent study similarly devoted itself only to how satisfied female victims of domestic violence are with the legal system’s handling of their cases and what factors impact their satisfaction levels. Fleury, supra note 116; see also supra note 224 and accompanying text (discussing studies on the prosecution of domestic violence). As noted earlier, the six principal studies on the effectiveness of arrest were limited to the arrest of male batterers. See supra notes 151-56 and accompanying text (discussing the Minneapolis Experiment and the five replication studies).

227. The titles of various works on prosecution alone are revealing on this point. See, e.g., Cahn & Lerman, supra note 202; Fleury, supra note 116; Ford & Regoli, supra note 155; Ford & Regoli, supra note 224.

228. See, e.g., FAGAN, supra note 154, at 48 (critiquing the limited existing means and proposing a variety of new means to study the criminalization of domestic violence, but concluding that the objective is to help battered women); Hanna, supra note 106, at 1909 (taking the controversial position of restricting victim control in domestic violence prosecution in order to advance efforts to make domestic violence a public crime, but limiting her discussion to the criminalization of male violence).
against women are commendable and much needed. However, if we are truly committed to ending domestic violence, the violence by women needs to be both acknowledged and addressed. Reacting to such violence must be part of the “coordinated response.” Unfortunately, however, the momentum behind the feminist hold on domestic violence is not contained. Beyond restricting the development of arrest and prosecution policies, feminist theory has debilitated the treatment of both domestic violence batterers and victims.

3. Punishment

(a) Treatment

In any criminal context, the potential influence of the judiciary over domestic violence cannot be overstated. As leading domestic violence commentators well understand:

The effect of judicial attitudes cascades throughout the criminal justice system. The judiciary retains the potential of leading the criminal justice system by example or direction. After all, they are the ultimate authority having the power to ratify or condemn the actions of the police and prosecutors, as well as defining the parameters and seriousness of a particular crime.229

Notwithstanding the strength of the judicial position, the judiciary exercises little practical influence over domestic violence today. The limited possibility of arrest, compounded by the unlikelihood of prosecution, ensures that the powerful judicial link in the “coordinated response” to domestic violence will rarely be reached.230 However, in the event that police do arrest a batterer and prosecution by the state attorney is pursued, what action is taken by the judicial branch? And further, what, if any, distinctions are made between the response to male and female batterers who reach this stage?

Punishment may serve a variety of purposes—among the most prominent being rehabilitation, deterrence (either general or specific), incapacitation, retribution, and condemnation.231 Yet in the domestic violence context, rehabilitation or treatment is almost universally accepted as the sole objective.232 Interestingly, the wide-

229. Buzawa & Buzawa, supra note 142, at 65.
230. For a discussion of the limited use of arrest in the domestic violence context, see supra notes 151-54 and accompanying text. On the unlikelihood of prosecution, see supra notes 210-15 and accompanying text.
231. For a general discussion of the theories of punishment underlying crimes of domestic violence, see Hanna, supra note 147, at 1538-48.
232. Violence in Families, supra note 224, at 178; Ford & Regoli, supra note 155, at 158; Gregory & Erez, supra note 116, at 206; Hanna, supra note 147, at 1508, 1522; Ellen Pence & Melanie Shepard, Integrating Feminist Theory and Practice: The Challenge of the Battered Women’s Movement, in Feminist Perspectives on Wife Abuse, supra note 55, at 282, 284.
spread state and federal “faith in treatment” traces back to the same 1984 U.S. Attorney General’s report which also supported aggressive arrest policies.\textsuperscript{233} Recently, this inconsistency between promoting treatment while endorsing more aggressive arrest and prosecution policies has come under some attack. Suggesting incapacitation, not rehabilitation, to be the appropriate goal of punishment, at least one domestic violence activist and scholar advocates the more aggressive use of incarceration.\textsuperscript{234}

Despite the widespread use of treatment as a criminal response to domestic violence, the critics of treatment are not just the small minority who favor alternative punishments. In addition to those in theoretical opposition to treatment, there is widespread recognition that treatment has, at best, yet to be proven an effective response to domestic violence, and at worst might encourage greater violence.\textsuperscript{235} So why then is treatment the predictable outcome for most domestic violence assault cases which actually reach the courts?

As with prosecution, there is the practical reality that the criminal justice system is ill-equipped to underwrite more expensive and

\begin{itemize}
\item \textsuperscript{233} Per such report, incarceration was to be reserved for only the most serious injury-producing cases of domestic violence. Hanna, supra note 147, at 1526 (discussing the 1984 ATTORNEY GENERAL’S TASK FORCE ON FAMILY VIOLENCE: 1984 FINAL REPORT (1984)). For earlier discussions of the Attorney General’s support of aggressive arrest, see supra notes 149, 155 and accompanying text.
\item \textsuperscript{234} In advocating greater use of incarceration, Professor Hanna also notes the ineffectiveness of treatment. Her support of incarceration is in keeping with her promotion of even more aggressive arrest and prosecution policies. Hanna, supra note 147.
\item \textsuperscript{235} For reports on a variety of treatment studies which fail to conclusively prove that treatment reduces domestic violence, see, for example, Edward W. Gondolf & Alison Snow Jones, The Program Effect of Batterer Programs in Three Cities, 16 VIOLENCE & VICTIMS 693 (2001). The author studied the recidivism rates of domestic violence offenders enrolled in Denver, Houston, and Dallas treatment programs and suggested only tentatively that there may be a “moderate program effect.” Id. at 700; see also FAGAN, supra note 154, at 19 (finding “very little non experimental or quasi-experimental evidence” to properly evaluate batterer programs); JACOBSON & GOTTMAN, supra note 116, at 231 (recognizing research on treatment programs yet to evidence treatment is effective); PENCE & PAYMAR, supra note 101, at 174-77 (surveying shelter workers in seventy-six battered women’s shelters to gain their perspectives on batterer’s programs and reporting results ranging from the belief that such programs helped victims, to the assertion that victims were put at risk of greater violence, to the feeling that there was no change in the violence, or that batterers used such programs to gain leverage in child custody and other matters); VIOLENCE IN FAMILIES, supra note 224, at 180 (suggesting that rehabilitation programs must be studied more carefully before any conclusions regarding their effectiveness can be reached); Gregory & Erez, supra note 116 (conducting interviews with thirty-three women whose batterers were enrolled in treatment programs and concluding that while victims believe that treatment works, such feelings are often mixed with apprehension, fear and guilt); L. Kevin Hamberger & James E. Hastings, Court-Mandated Treatment of Men Who Assault Their Partner: Issues, Controversies, and Outcomes, in LEGAL RESPONSES TO WIFE ASSAULT, supra note 116, at 188, 207-20 (summarizing twenty-eight of the major treatment studies and concluding that the mixed makeup, methodology, and results of such studies yields “[not much] further knowledge about the effectiveness of treatment”); Hanna, supra note 147, at 1514-38 (reviewing a variety of empirical evidence showing treatment is not effective); Pence & Shepard, supra note 232, at 291-92.
\end{itemize}
time-consuming forms of punishment such as incarceration.\textsuperscript{236} Judicial biases against regarding domestic violence as a “real crime” also factor into the explanation.\textsuperscript{237} Such attitudes are exacerbated by treatment advocates who support treatment not because it is effective, but because it allows domestic violence activists to retain control over the issue of domestic violence, generally, and the fate of domestic violence batterers, more specifically.\textsuperscript{238} At the same time, legitimate concern for victim safety and awareness of the ongoing relationships between batterers and victims also exists. Such attitudes may give way to sincere beliefs that trying to change the behavior of batterers may be the best solution, even if it has yet to be proven possible.\textsuperscript{239}

With this array of justifications and concerns, treatment programs run a certain gamut. Courts may mandate treatment as a pretrial diversion tactic, they may defer prosecution with the promise of a reduced sentence upon completion of treatment, or they may order treatment as part of a sentence upon conviction.\textsuperscript{240} Basic differences in such matters as length and size of treatment programs and whether such programs are group or individually oriented also exist.\textsuperscript{241} Notwithstanding the procedural and structural variety, there is

\textsuperscript{236} In supporting greater use of incarceration in the domestic violence context, even Professor Hanna notes the practical difficulties in that cases of assault and battery between intimates may already be treated more seriously than similar cases between strangers. Hanna, supra note 147, at 1524-26. For further discussion of court reluctance to impose harsher sentences, see FAGAN, supra note 154, at 30-33; PENCE & PAYMAR, supra note 101, at xiii; Hanna, supra note 147, at 1556.

\textsuperscript{237} BUZAWA & BUZAWA, supra note 142, at 65; see also Hanna, supra note 147, at 1556; Robbins, supra note 201, at 211-13. For further discussion of the traditional social minimization of domestic violence, see supra Part III.A.1.

\textsuperscript{238} Hanna, supra note 147, at 1550.

\textsuperscript{239} JACOBSON & GOTTMAN, supra note 116, at 231; Hamberger & Hastings, supra note 235, at 189-90. As Cheryl Hanna concludes:

\begin{quote}
Court-mandated treatment programs allow everyone to save face. The prosecutor checks-off ‘conviction’ on his stat sheet; the defense attorney feels like she did some good for her client; the victim has a sense of hope, however false, that the criminal justice system will help her partner change his ways; the offender avoids jail; the judge is not accused of taking these cases too lightly; the treatment program gets yet another client to support its existence; and we all go home happy . . . until the next time.
\end{quote}

Hanna, supra note 147, at 1556.

\textsuperscript{240} VIOLENCE IN FAMILIES, supra note 224, at 178; Hamberger & Hastings, supra note 235, at 190-91; Hanna, supra note 147, at 1522.

\textsuperscript{241} While treatment programs generally favor group counseling as opposed to a one-on-one counseling approach, there tends to be a great deal of variation in other format matters. For a sampling of various program structures, see DUTTON, supra note 96 (explaining a treatment program run by the author). See also PENCE & PAYMAR, supra note 101, at 17-27 (explaining the structural design of the Duluth Model); Gondolf & Snow Jones, supra note 235 (reviewing the programs of Dallas, Denver, and Houston); Hamberger & Hastings, supra note 235, at 203-05 (reviewing treatment format issues), 208-10 (summarizing in chart form the variables of twenty-eight programs); Hanna, supra note 147, at 1528-90 (reviewing the development of treatment programs).
currently a great deal of consistency in the philosophical approach to
treatment. Succinctly characterized, today’s treatment programs are
“feminist-inspired.”

(b) The Feminizing of Treatment

Profeminist treatment has not always been the mainstay. Rather,
its predominance is a natural development in the feminist metamor-
phosis of domestic violence. When domestic violence theory adhered
to a belief that both the batterer and victim had some accountability,
treatment programs relied on models which counseled the couple. As
a result, these programs focused on the shared desire to control and
the common repressed anger, anxiety, or other negative feelings of
the parties which ultimately resulted in violent behavior. Yet as
victim-blaming came under attack and as the cycle of violence and its
emphasis on the male use of physical violence gained popularity,
couple therapy was replaced by anger-management programs which
not only allowed solely male attendance, but also demanded that
men take full accountability for their violent behavior. However,
like the original cycle of violence, this behavioral approach came un-
der criticism for teaching anger-management skills without any ef-
fort to address the underlying, but believed omnipresent, gender-
based motivation for violent male behavior toward women. As the
feminist theory of connection between violence and patriarchy took

242. Hanna, supra note 147, at 1530. For other recognition of the predominance of the
feminist model, see, for example, Jacobson & Gottman, supra note 116, at 231; Physical
Violence, supra note 14, at 11-14; Wexler, supra note 116, at 13; Hamberger & Hast-
ings, supra note 235, at 202-03; Pence & Shepard, supra note 232, at 294; Daniel G. Saun-
ders, Feminist, Cognitive, and Behavioral Group Interventions For Men Who Batter: An
Overview of Rationale and Methods, in Domestic Violence 2000: An Integrated Skills

243. David Adams describes programs of this nature as following an “insight model”
(explaining violent behavior as the symptom of some other underlying conflict); “ventila-
tion model” (addressing the parties’ repressed feelings and teaching “fair fighting” tactics);
and “interaction model” (describing violence as a culmination of efforts by both parties to
gain power and control). David Adams, Treatment Models of Men Who Batter: A Profem-
inist Analysis, in Feminist Perspectives on Wife Abuse, supra note 55, at 176, 178-88;
see also Hamberger & Hastings, supra note 235, at 196-98 (reviewing the insight, ventila-
tion, and interaction models).

For a discussion of early domestic violence theory and its tendency to blame both couple
members, see supra notes 104-08 and accompanying text.

244. Such a model has been described as the “cognitive-behavioral” or “psychoeduca-
tional” model. Adams, supra note 243, at 188-91; see also Hamberger & Hastings, supra
note 235, at 200-01. For a similar recognition of the move away from couples therapy, see
Hanna, supra note 147, at 1527-28. For further discussion of the cycle of violence, see su-
pra notes 103-08 and accompanying text. For further discussion of victim-blaming, see su-
pra note 96 and accompanying text.

245. For a review of such criticisms of the cycle of violence, see supra note 109 and ac-
companying text.
shape, so too did the profeminist model of treatment.\textsuperscript{246} And as the “power and control” wheel of the Domestic Abuse Intervention Project (DAIP) of Duluth, Minnesota became the most popular invention in defining domestic violence, the profeminist Duluth Model became the most widely imitated treatment model.\textsuperscript{247}

\textit{(c) The Duluth Model}

While the Duluth Project began with an anger-management treatment program, it openly rejected this approach in 1984 and adopted a profeminist agenda with a focus on violence as a mechanism of power and control.\textsuperscript{248} Almost twenty years after such a drastic change, the Duluth Model continues to identify itself with several key feminist goals. To have batterers understand that their use of violence is a means of control, “women’s reality and women’s experiences” are a constant part of the curriculum.\textsuperscript{249} Male participants examine their use of violence in its “cultural and social contexts,” with the program striving to reveal how violence is socially learned by men in order to ensure their control over women.\textsuperscript{250} Leaving no room for the suggestion that men may also be the “victims of sexism,” the program’s inevitably all male audience is forced to take full accountability.\textsuperscript{251} No attempt to “psychologize” the problem is tolerated, for such efforts detract from the focus on violence as a tool of power and control and risk providing men with a means of denying responsibility.\textsuperscript{252} Indeed, the patriarchal agenda is so controlling that group facilitators are discouraged from exploring participants’ personal prob-

\textsuperscript{246} For recognition of the current prevalence of the profeminist model, see JACOBSON & GOTTMAN, supra note 116, at 231; Adams, supra note 243, at 190; Hamberger & Hastings, supra note 235, at 200-01; Hanna, supra note 147, at 1528-31; Pence & Shepard, supra note 232, at 282-85; Saunders, supra note 242, at 22.

\textsuperscript{247} For recognition of the Duluth program as the profeminist model most widely imitated, see JACOBSON & GOTTMAN, supra note 116, at 231; WEXLER, supra note 116, at 13; Hanna, supra note 147, at 1530; Pence & Shepard, supra note 232, at 294.

For discussion of the “power and control” wheel and its development within the Duluth project, see PENCE & PAYMAR, supra note 101, at 2-3. See also supra notes 111-15 and accompanying text.

\textsuperscript{248} PENCE & PAYMAR, supra note 101, at 29. It is interesting to note that the project’s change in 1984 coincided with other key developments that year, namely the Attorney General’s Task Force on Domestic Violence Report, the Minneapolis arrest experiment and the decision in \textit{Thurman v. City of Torrington}, 595 F. Supp. 1521 (D. Conn. 1984). For a discussion of these events, see supra note 150 and accompanying text.

\textsuperscript{249} PENCE & PAYMAR, supra note 101, at 30.

\textsuperscript{250} Id. at 5, 29; Pence & Shepard, supra note 232, at 289-90 (“Participants are taught that it is not their anger that leads them to be violent, but rather their belief that they have a right to control and dominate women.”).

\textsuperscript{251} Pence & Shepard, supra note 232, at 295; see also PENCE & PAYMAR, supra note 101, at 29. While some female batterers have also been noted to attend, the program’s approach toward such participants is much less demanding. See infra note 260 and accompanying text (discussing female batterers and the Duluth program).

\textsuperscript{252} WEXLER, supra note 116, at 14.
lems, such as substance abuse or the particulars of their relationships. Confronted with the categorical characterization of every batterer’s violence as a weapon of male domination, the program’s ultimate aspiration is to “undo sexism,” eliminating not just the violence, but all the tactics men use to oppress women. Having unlearned such behavior, it is replaced with more egalitarian attitudes and skills. So executed, a batterer’s treatment program modeled on the Duluth curriculum perfectly follows the flow of domestic violence theory—from explaining male behavior through power and control, to the final goal of equality.

The unyielding support for treatment, in contrast to the equally strong support for aggressive arrest and prosecution policies, evidences a basic feminist tension between the desire to publicly condemn and criminalize domestic violence and the desire to protect a woman’s autonomy by suggesting her life with the abuser may be restored to the status quo—minus the violence and subordination. By one interpretation, the stereotypical female “ethic of care,” the female inherent nature to act in a relational, caring manner as opposed to a more male, isolationist manner is on display through this feminist call for rehabilitation. Yet at a deeper level, the support of treatment, and more specifically feminist treatment, unequivocally reflects a more political goal. Mandating men to treatment gives feminists their most desired audience—men. Once captive to the feminist program, men can be indoctrinated. First convinced that all their violent and nonviolent actions toward women are motivated by a desire to subjugate women, the men may then be educated in the manner which women feel they should be treated.

Certainly, ending gender motivated violence and ensuring women fair treatment are laudable objectives. Yet, ultimately, the control of

254. Id. at 7. For an extensive discussion of the Duluth Model’s curricular breakdown between explaining the patriarchal behavior of violent men and replacing such behavior with “noncontrolling and nonviolent ways of relating to women,” see id. at 30, 29-65.
255. Id. at 7.
256. See Hanna, supra note 147, at 1548-50. Professor Hanna similarly suggests that tension between feminist goals explains the ambivalence feminists manifest in supporting aggressive prosecution while simultaneously resisting “no-drop” policies which prohibit battered women from choosing to have the charges against their batterers dismissed. Hanna, supra note 106, at 1853, 1865-66; see also supra Part IV.B.2 (discussing prosecution goals and conflicts).
257. Hanna, supra note 147, at 1550. The female “ethic of care” originates with psychologist Carol Gilligan. See Carol Gilligan, In A Different Voice (2d. ed. 1993). Since its coinage, the term and its association with women has been subjected to a great deal of criticism. For a sampling of positive reliance on Gilligan’s work by cultural feminists, see, for example, McClain, supra note 124, at 1182-83; Sherry, supra note 124, at 580, 585, 587; Robin West, Jurisprudence and Gender, in Feminist Jurisprudence, supra note 126, at 493, 500-01. For negative treatment of Gilligan in radical feminist thought, see, for example, Mackinnon, supra note 100, at 38-39.
gender is not eradicated through the profeminist agenda; rather, it shifts. For as Cheryl Hanna sharply explains, “[t]reatment programs turn the tables of control from misogynist men to profeminist women and men whose agenda it is to restructure gender relations.”

Beyond the gender empowerment issues, the Duluth Model’s exclusively feminist focus fails to even allow for the possibility that a man’s violence might be caused by other conditions. Recent research suggests that abusive men can be described through a variety of profiles and that more positive, less confrontational models of treatment may be more effective. The Duluth Model’s unyielding reliance on treatment also ignores other legitimate goals of punishment and the methods used to achieve them. Yet, the most gaping hole in today’s response to batterers is the failure to consider the reality of female batterers. Under feminist control, today’s treatment denies the possibility that women can be violent. Accordingly, it fails to provide any means of treating or otherwise responding to female intimate violence. Even the few female offenders who do manage to end up in treatment, despite the layers of social and legal bias, are ultimately not required to assume the accountability demanded of male batterers. Authors of The Duluth Model maintain that such women must have used violence in self-defense.

As with arrest and prosecution policies, the sincere commitment to ending domestic violence is not evident in today’s judicial system. The goal has been lost in the pursuit of the feminist agenda. Likewise, services to victims of domestic violence have also fallen under feminist control.

4. Victim Services

(a) Shelters

By far, shelters are the predominant service for victims of domes-
tic violence. While the shelter movement began in England as early as 1971, the United States soon followed suit. By 1974, the first shelter for women was opened in St. Paul, Minnesota. By 1986, over 700 domestic violence shelters were counted in the United States. Since then, the number has virtually doubled.

As shelters have grown, so have their role. Originally recognized as a “crisis service,” shelters now provide a range of services. In addition to refuge, other immediate assistance is provided to both residents and nonresidents through the availability of emergency hotlines, food, child care, transportation, support groups, and referrals to additional services. Through direct counseling services, as well as access to the network of legal and community service providers, shelters are considered critical to victims of domestic violence who seek permanent lifestyle changes by: finding alternative living arrangements, pursuing child custody and divorce, filing criminal charges, receiving job training or employment leads, and working to separate from their abusers.

Despite the growth in and reliance on shelters, there is little empirical evidence of their effectiveness. While shelter studies have used such measures as the recurrence of violence, the long-term separation of victim and abuser, and rates of victim satisfaction, the results are all equivocal. Legitimate concerns are raised as to whether any of such yardsticks accurately reflect the success of shelters. Victim evaluation studies are overly subjective and have tended to review attitudes toward various types of counseling, not shelter services per se. Because the threat of separation assault attaches to any effort to leave a violent relationship, separation is routinely

---

261. Shelters have recently been reported to account for 1200 of the 1800 programs in the United States for victims of domestic violence who are either spouses or intimate partners. VIOLENCE IN FAMILIES, supra note 224, at 110.


263. Known as “Women’s House,” the shelter was opened by the Women’s Advocates. WALLACE, supra note 219, at 177; Pence & Shepard, supra note 232, at 282-83.

264. Berk et al., supra note 262, at 482.

265. VIOLENCE IN FAMILIES, supra note 224, at 110 (noting the existence of 1200 domestic violence shelters in the United States by the late 1990s).

266. GONDOLF & FISHER, supra note 130, at 77.

267. For an overview of the intervention services offered by shelters, see GONDOLF & FISHER, supra note 130, at 78; Dutton-Douglas & Dionne, supra note 262, at 114-21.

268. Dutton-Douglas & Dionne, supra note 262.

269. GONDOLF & FISHER, supra note 130, at 81; Berk et al., supra note 262, at 481; Dutton-Douglas & Dionne, supra note 262, at 113.

270. Dutton-Douglas & Dionne, supra note 262, at 121.

271. Self-reported evaluations have generally investigated victims’ attitudes toward services provided by women’s groups as opposed to religious or more general social organizations. Id. at 124.
recognized to be an inadequate measure of shelter success as shelter stays may lead to more, rather than less violence.\textsuperscript{272} Violence recurrence studies may be the best means of evaluating shelter success. However, they too are flawed. Surveying only domestic violence victims who have sought shelter services, such studies lack a control group and instead rely on nonrandom samples that may not accurately reflect the diverse pool of domestic violence victims.\textsuperscript{273} Yet, after accounting for such methodological concerns and recognizing the risk of increased violence upon separation, violence recurrence studies generally associate a victim’s use of shelter services with a reduction in violence.\textsuperscript{274} But instead of finding a causal link between shelters and violence reduction, such studies conclude that it is not the victim’s decision to utilize shelter services but rather the victim’s assertion of control which stops violence. When a victim’s use of shelter services is not perceived by the batterer as part of a “serious” decision to leave the relationship, the action is simply another display of disobedience that must be met with more violence for punitive effect.\textsuperscript{275} Such findings and criticisms of shelters certainly do not explain the support shelters enjoy. Yet, rather than detracting from the endorsement of shelters, such empirical findings have been relied upon to advocate for shelters as an integral part of a coordinated effort to supply a wider range of victim services.\textsuperscript{276} Why are shelters so strongly supported?

\textsuperscript{272} Id. at 121-23 (finding separation studies to be “[a]t best . . . only an indirect measure of intervention effectiveness”); GONDOLF & FISHER, supra note 130, at 80, 82. For a thorough discussion of separation assault, see Mahoney, supra note 108. See also supra note 138 and accompanying text.

\textsuperscript{273} Berk et al., supra note 262, at 484.

\textsuperscript{274} For a collection of violence recurrence studies, see id. (studying the recurrence of violence for victims seeking shelter services in Santa Barbara, CA); Dutton-Douglas & Dionne, supra note 262, at 124 (reviewing various studies); Maryse Rinfret-Raynor & Solange Cantin, Feminist Therapy for Battered Women: An Assessment, in OUT OF THE DARKNESS: CONTEMPORARY PERSPECTIVES ON FAMILY VIOLENCE, supra note 90, at 219 (studying violence recurrence and other factors for battered women participating in various therapies).

\textsuperscript{275} Berk et al., supra note 262, at 484; see also GONDOLF & FISHER, supra note 130, at 82; Dutton-Douglas & Dionne, supra note 262, at 124.

\textsuperscript{276} One study stated:

The overall implication in these findings is that shelters cannot do it alone. More than shelter refuge is needed to help women maintain some semblance of safety. Battered women need a wide variety of services and resources to leave their batterers. There also needs to be vigilant coordination with batterer programs to assure that the batterer’s counseling does not mistakenly lure women back to an unsafe relationship. In sum, the community help sources need to be more systematically coordinated in order to effectively assist wife abuse.

GONDOLF & FISHER, supra note 130, at 88.
(b) The Feminist Politics of Shelters

As with other aspects of the domestic violence system, because shelters began in the “formative years” of the battered women’s movement, they too are built upon the feminist framework. Overriding the shelter objective of helping victims of domestic violence to safety is the belief that women live in a world of “male imposed isolation” and that a shelter offers a means of escape from this world. This feminist politicization of shelters is both openly acknowledged and celebrated. Shelters provide women’s groups with:

- a place where batterers and their coconspirators in the system cannot control the discussion; they cannot interpret the facts; they cannot silence women’s minds nor keep women from speaking . . . .
- [T]he groups are designed to create a safe space for women to find their personal power to join with other women to take back control of their lives.

Or more simply, “[t]he role of the shelter and the advocate cannot be underestimated.”

Given such strong feminist influence, it is no surprise that the wide variety of residential and nonresidential services offered by shelters is extended almost exclusively to female victims of domestic violence. Only a handful of men’s shelters have ever existed. The belief that men could not possibly need the services of domestic violence shelters is so widespread that it has been implicitly assumed in the judicial review of gender-based challenges made by battered women denied the opportunity to build a shelter.

While it is acknowledged that the few shelters that have opened for battered men have been closed for lack of support and patronage, this history does not lend itself to arguments against services for

---

277. See Pence & Shepard, supra note 232, at 282-83.
278. Id. at 291.
279. See Pence & Paymar, supra note 101, at 11; see also Pence & Shepard, supra note 232, at 282. Yet for current support for the depoliticization of clinics and greater attention to gender-neutral services and family violence generally, see infra notes 296-301 and accompanying text.
281. Id.
282. See Cook, supra note 44, at 54, 156, 168 (noting that only one or two shelters for battered men have existed).
283. See, e.g., Doe v. City of Butler, 892 F.2d 315, 328 (3d Cir. 1989). In this case, the shelter application had been denied as a result of a local zoning ordinance which prevented transitional dwelling in residential areas which would house more than six people (including children). In rejecting the women’s disparate impact claim made pursuant to the Fair Housing Act, the court found the battered women’s denial of shelter space to be no greater than the denial which would come to a similarly situated group of men who might need a shelter for other purposes, such as alcoholism. In so doing, the court implicitly assumed both that only women could be abused and that only men could be alcoholics. Id. at 328.
male victims of domestic violence. The lack of support is not a result of the lack of need of services for battered men. Rather, it can more fairly be attributed to the widespread denial of female violence and the stigmatization felt by any man claiming to be a victim of such. It may be argued that shelters for homeless men are ample enough to absorb any population of battered men. However, in some areas, particularly large metropolitan areas, there is not sufficient shelter space to assist all men seeking shelter. In the case of battered men accompanied by their children, the lack of adequate physical space becomes more critical. There is terrific difficulty in finding suitable shelter for homeless families, particularly those headed by men.

Even if it is conceded that battered women may have a greater need for shelter space than battered men, such concession does not mandate that both the services and the space provided by a battered women’s shelter cannot be utilized to accommodate battered men. Existing space is often already partitioned in such a way to give families separate living quarters. Future space can be built to better accommodate battered men. Yet, perhaps most importantly, as is recognized in the support of domestic violence shelters, shelters provide more than a place of physical safety. Domestic violence shelters offer “hope, support, and counseling specifically targeted to the victims of domestic violence.” Such an offer should be as readily made to battered men as it is to battered women.

V. THE CHALLENGE

From the theoretical definition of domestic violence, to the practical treatment of arrest, prosecution, punishment and victim services, feminism controls domestic violence. Such an approach is a loss for both sexes. Battered men are prevented from seeking relief. Violent women are not restrained. To the extent legitimate, non-patriarchal explanations of domestic violence are suppressed, battered women also remain at risk. But as feminists have strongly held in their work on domestic violence, “[i]t is the constant interaction of theory and practice that keeps a movement dynamic, growing, changing, and most importantly, moving.” To remain truly committed to keeping

---

284. See Cook, supra note 44, at 156.
285. See id. at 155. For further discussion of societal attitudes toward female violence, see supra note 90 and accompanying text.
286. See Cook, supra note 44, at 155.
287. See id. at 155-57.
288. See id. at 156.
289. See id. at 156-57.
290. Id. at 156.
the domestic violence movement moving, domestic violence theory must now be challenged. Revitalizing the domestic violence movement calls upon the strength of both theoreticians and practitioners and upon both women and men. It also requires suggestions for change. Such efforts can be initiated by first recognizing the value of many current discussions and projects and then building upon them.

A. A Challenge to Theory

Just as we have successfully challenged the “essential” battered woman, the image of the “essential” battering male is ripe for challenge.292 Contrary to feminist assumptions, not all male batterers are motivated by gender domination. Recent profile studies of male batterers dispel the “homogeneous” batterer myth.293 In its place are findings which suggest that batterers can be divided into subtypes based upon considerations of the extent of their marital violence, general violence (which is directed toward both strangers and intimates) and any psychopathological or personality disorders.294 From such typologies, research may still support the theory that a batterer’s violence may be motivated by patriarchal interests. However, this characterization cannot be made in all cases. To the extent violent behavior is indiscriminate, or a product of either a physical or mental disorder, no patriarchal charge can be made.295 Pursuing such sophisticated, complex theories which explore the variety of reasons a man might commit domestic violence, will yield more effective tools for addressing and preventing domestic violence.296

292. See Hanna, supra note 147, at 1561-62. The prototypical battered woman suffered from learned helplessness and failed to account for the diversity of needs and interests brought on by differences in such factors as race, religion, class, age, culture, or immigration status. There have been various contributions deessentializing the battered woman. See, e.g., Crenshaw, supra note 108 (exploring the various intersectional contexts of battered women); Hart, supra note 96, at 99 (warning against perpetuating a homogeneous battered woman); Kelly, supra note 108, at 311-14 (reviewing the cultural and immigration complications faced by battered women); Meier, supra note 106, at 1305-07 (discussing the fallacies of learned helplessness theory); Nilda Rimonte, A Question of Culture: Cultural Approval of Violence Against Women in the Pacific-Asian Community and the Cultural Defense, 43 STAN. L. REV. 1311 (1991) (examining cultural attitudes toward domestic violence held by individuals from Asian and Pacific Rim countries in the United States); Seymore, supra note 116, at 1065 n.214 (noting the diversity amongst battered women).

293. FAGAN, supra note 154, at 36.

294. Id. (citing Amy Holtzworth-Munroe & Gregory L. Stuart, Typologies of Male Batterers: Three Subtypes and Differences Among Them, 116 PSYCHOL. BULL. 476 (1994); see also Hanna, supra note 147, at 1563-72 (reviewing the research on the reasons why men batter).

295. See FAGAN, supra note 154, at 36.

296. See Hanna, supra note 147, at 1561; Daniel G. Saunders, Husbands Who Assault: Multiple Profiles Requiring Multiple Responses, in LEGAL RESPONSES TO WIFE ASSAULT, supra note 116, at 9-30. For a discussion of practical proposals regarding domestic violence prosecution and treatment built on developing theory, see infra Part V.B.
Yet, beyond more effectively addressing the violence of men, such work must be recognized for its contribution to recognizing the violence of women. Once the patriarchal element is removed and replaced by gender neutral explanations for violence, it logically follows that violence may be committed by women as well as men. This theoretical move is consistent with other theoretical efforts that have already implicitly accepted that domestic violence is not only committed by men and that it is not always motivated by gender. Domestic violence takes form in elder abuse, child abuse, and sibling abuse. Homosexual abuse, including lesbian battering, is also finally being acknowledged. In combination with the magnitude of

297. For a selection from the work on elder abuse, see, for example, Lynda Aitken & Gabriele Griffen, Gender Issues in Elder Abuse (1996); Gerry Bennett & Paul Kingston, Elder Abuse: Concepts, Theories and Interventions (1993); Michael Brogden & Preeti Nijhar, Crime, Abuse and the Elderly (2000); Patricia J. Bronell, Family Crimes Against the Elderly: Elder Abuse and the Criminal Justice System (1998); Frances Merchant Carp, Elder Abuse in the Family: An Interdisciplinary Model for Research (2000); Yvonne Joan Craig, Elder Abuse and Mediation: Exploratory Studies in America, Britain and Europe (1997); Richard J. Gelles & Claire Pedrick Cornell, Intimate Violence in Families 100-04 (2d ed. 1990); Intimate Violence in Families, supra note 14, at 113-19; The Mistreatment of Elderly People (Peter Decalmer & Frank Glendenning eds., 1997); Understanding Elder Abuse in Minority Populations (Toshio Tatara ed., 1999); Wallace, supra note 219, at 237-57.


299. For various accounts of sibling violence, see, for example, Gelles & Cornell, supra note 297, at 85-90; Intimate Violence, supra note 14, at 59-60; Intimate Violence in Families, supra note 14, at 97-103; Straus & Gelles, supra note 14, at 449-52; Wallace, supra note 219, at 108-20.

studies reporting on female intimate violence, such research powerfully challenges existing feminist assumptions. Women are capable of intimate violence. Men can be victims.

B. A Challenge to Practice

It is indisputable that feminists have engaged in hard and sincere work on the domestic violence front. The feminist “discovery” of domestic violence and subsequent efforts have made it an issue of terrific public importance. The value of these feminist efforts should never be minimized. However, while not detracting from such progress, it must also be recognized that the limits of feminist theory prevent a full understanding of domestic violence. In addition to its advances, feminist theory is responsible for the limits of existing domestic violence theory and policies. By expanding the theoretical foundation to allow for the possibility of domestic violence in every intimate human relationship, new perspectives on how to effectively address domestic violence will inevitably develop. In new explorations of such issues as punishment and victim services, such work has already begun.

In the punishment context, the appreciation of the multidimensional profiles of batterers has led to the recommendation of an array of treatment programs and other responses, which go well beyond the Duluth Model. While a great deal of such work remains limited to the context of male batterers, it rejects the premise that all violence is a function of patriarchy. Rather than helping to build “respectful relationships,” such feminist “shame-based” programs instead “[establish] a power hierarchy in the treatment setting that subtly reinforces power tactics—and that alienates the very population we want to reach.” Advocates of more “client-centered” solutions reject “confrontational approaches” which focus exclusively on gender and power issues. Their proposals include counseling and substance abuse programs geared to address different mental and physical conditions. Characterizing some batterers as “[b]orderline and sociopathic abusers” who are untreatable, the use of incarcera-
tion is also recommended.\textsuperscript{308} Similarly, advocates in the shelter movement are also adopting new perspectives. Rather than prioritizing the feminist agenda, greater attention is being paid to the delivery of more practical services such as those relating to housing and crisis intervention.\textsuperscript{309}

**CONCLUSION**

Such work marks the beginning of a positive, new direction for the domestic violence movement. Of course, much more needs to be done. For academicians and scholars, greater exploration of the physical, mental, and social conditions that give rise to domestic violence must be undertaken. Such work will lay the groundwork for activists who can develop more effective policies and programs for police, court officials, social workers, and others engaged in the fight against domestic violence. Legitimate concerns about how best to use limited funding can also be raised without having to rely upon more underhanded tactics meant to suppress any recognition of female violence. Through such a hard but sincere struggle, feminists and their work will not be ignored. Instead, patriarchy and gender will be recognized as one of many reasons why domestic violence exists. In so doing, feminism will maintain its legitimacy. And, with a great deal of hope, effort, and time, a real end to domestic violence will be found.

\textsuperscript{308} Hanna, supra note 147, at 1577.

\textsuperscript{309} Id. at 1531 n.117 (relying on Elizabeth M. Schneider, *The Violence of Privacy*, 23 Conn. L. Rev. 973, 993 (1991)). Not all recognition of such change is supported, see, for example, Norma Jean Profitt, *Women Survivors, Psychological Trauma, and the Politics of Resistance* 24 (2000) (noting such changes in the Canadian shelter system and criticizing them for failing to recognize the role of female victims). For further discussion of the politicization of shelters, see supra Part IV.B.4.(b).
SHOULD KLANSMEN BE LAWYERS? RACISM AS AN ETHICAL BARRIER TO THE LEGAL PROFESSION

CARLA D. PRATT*

I. INTRODUCTION..................................................................................................... 858
II. DEFINING MORAL CHARACTER AND FITNESS...................................................... 864
   A. The First Amendment as a Shield, Not a Sword ........................................ 867
   B. Lawyers’ First Amendment Freedoms Are Already Restricted When Necessary to Promote the Interests of Justice ................................. 871
   C.Balancing First Amendment Rights With Other Constitutional Interests ... 874
   D. Serving Two Masters .................................................................................... 877
III. LAWYERS SHOULD NOT EXEMPT THEMSELVES FROM ANTI-DISCRIMINATION LAWS .................................................................................................................... 884
IV. ADMISSION OF WHITE SUPREMACISTS WOULD BE PREJUDICIAL TO THE PUBLIC, THE LEGAL PROFESSION AND OUR SYSTEM OF JUSTICE ....................... 886
   A. Racism Obstructs the Administration of Justice ......................................... 886
   B. The Parade of Horribles ............................................................................... 890
   C. Admitting Racists to the Bar Undermines Equal Protection For All .......... 894
V. WHY RACISM? ...................................................................................................... 897
VI. THE TOUGH CASES............................................................................................... 901
   A. The Reformed Racist .................................................................................... 901
   B. The Open Racist vs. The Closet Racist ........................................................ 901
   C. The Mythological Black Supremacist .......................................................... 905
   D. The Unconscious Racist................................................................................ 908
VII. CONCLUSION ........................................................................................................ 908

No man can serve two masters: for either he will hate the one, and love the other; or else he will hold to one, and despise the other.1

* Assistant Professor of Law, The Dickinson School of Law of the Pennsylvania State University. B.A., East Texas State University, 1989; J.D., Howard University School of Law, 1992. I am grateful to my colleagues at Dickinson School of Law for their insightful and thought-provoking comments after I presented the idea for this paper to the faculty. I would like to convey special thanks to Dean Peter Glenn for the research grant that supported this work and to Peter Alexander, Larry Backer, David Richman, Victor Romero, and Laurel Terry for their thoughtful comments and suggestions. I also wish to express appreciation for the opportunity to present this paper at an Ethics Conference at Shippensburg University and at the Mid-Atlantic People of Color Scholarship Conference at Georgetown University Law Center. The dialogue at each conference was thought-provoking and contributed to shaping the thoughts presented in this paper. I would also like to extend my gratitude to Colleen Grygier, Jane Smedley, and Regina Ross for their able research assistance. Finally, I thank my husband, Michael, who supported me and encouraged me to write this Article despite the inherent vulnerability that is associated with being the “underdog.” This Article is dedicated to our children, Payton and Christopher, who I hope will grow up with a system of justice free from the infestation of racism.

1. Matthew 6:24 (King James). I thank Derrick Bell for his essay, which focused my attention on this verse. See Derrick A. Bell, Jr., Serving Two Masters: Integration Ideals and Client Interests in School Desegregation Litigation, 85 YALE L.J. 470 (1976).
I. INTRODUCTION

Within the legal community, it is often said that there are certain core values of the legal profession: loyalty, confidentiality and competence.\(^2\) Often overlooked, but recognized by some, is the core value of justice.\(^3\) From our history, which includes the drafting of our Constitution, are derived two core values that serve as the cornerstones of our system of justice, which necessarily includes the legal profession since it has the monopolistic privilege of administering justice.\(^4\) Those core values are fairness and the notion of equal justice for all.\(^5\) Equal justice is a core value of the legal profession because history demonstrates that without a commitment to this value, our legal system will commit atrocities against the powerless and the oppressed. Without a commitment to equal justice, lawyers are empowered to commit the racist mistakes of past administration of justice to our


\(^3\) See Robert L. Nelson, The Futures of American Lawyers: A Demographic Profile of a Changing Profession in a Changing Society, 44 CASE W. RES. L. REV. 345, 384 (1994) (arguing that the core value around which the legal profession should be organized is justice); see also Judith L. Maute, Pro Bono Publico in Oklahoma: Time For Change, 53 OKLA. L. REV. 527, 549 n.123 (2000) (citing Conference Report: ABA Mid-Year Meeting, 69 U.S.L.W. 2042 (U.S. July 18, 2000), in proposing that equal access to justice and pro bono services are core values of the legal profession).

\(^4\) It could be argued that a lawyer administers justice any time the lawyer acts in a capacity which constitutes the practice of law because the lawyer’s role arguably affects the outcome for the client and therefore affects the “justice” that the client receives. I would probably define the term “administering justice” more narrowly to include functions wherein the lawyer is empowered with the ability to make decisions that could negatively impact those persons seeking justice. Lawyers administer justice in many ways, including but not limited to selecting jurors in \textit{voir dire}, performing the function of civil or criminal prosecutor on behalf of the government, and/or serving as judge, arbitrator or mediator.

\(^5\) I use the term “equal justice” not merely as a reference to equal access to our system of justice, but as a reference to the underlying premise that all persons subject to the jurisdiction of our system of justice are entitled to fairness and equal treatment in the administration of the law regardless of, for example, race or socioeconomic background. The preamble to the MODEL RULES OF PROFESSIONAL CONDUCT, as revised by the Ethics 2000 Commission and adopted by the ABA House of Delegates at its February 2002 meeting, supports this contention by providing that “all lawyers should devote professional time and resources and use civic influence to ensure equal access to our system of justice for all those who, because of economic or social barriers, cannot afford or secure adequate legal counsel.” MODEL RULES OF PROF'L CONDUCT pmbl. (2001).

\(^6\) The ABA has intimated that equal justice is a core value of the profession by adopting a resolution in response to the multi-disciplinary practice (MDP) controversy. The resolution acknowledges the obvious core values of client loyalty, independent professional judgment, confidentiality, and competence, while also acknowledging as a core value of the profession “the lawyer’s duty . . . [as] an officer of the legal system, and a public citizen having special responsibility for the quality of justice.” ABA House of Delegates, Revised Resolution 10F (July 2000), available at http://www.abanet.org/cpr/mdprec0010f.html (last visited Mar. 15, 2003) (on file with author) (emphasis added).
minority population. A commitment to equal justice is important to the professional role of a lawyer because without it lawyers will be perceived as biased and unfair administrators of justice who will threaten the orderly administration of justice by causing the public to question the integrity of the entire judicial system. Justice is not a commodity for only the majoritarian group; it is for all. Hence in order to be a lawyer, one must be committed to justice for all regardless of race. The thesis of this Article is that in order to possess the requisite moral character to be an officer of the court and an administrator of justice, also known as a lawyer or attorney, an individual must subscribe to the core value of equal justice that serves as a cornerstone of our entire system of justice. Members of any group advocating racial caste systems, including but not limited to white supremacists, do not subscribe to these core values because they do not believe that all people are entitled to fairness and equal justice without regard to their race. White supremacists believe that justice is a commodity that only the whitest of white people should enjoy. Being a legal realist, I too believe that the world in which we live and practice “law” is a nomos contextualized by the narratives that locate law and give it meaning. The genre of narrative that locates

---

7. I use this term, as many scholars have, to mean the racial majority, which in this country happens to be “white people.” The definition of “whiteness” has evolved with our culture. For an understanding of who historically has been considered white and who today is considered white, see JUAN F. PEREA ET AL., RACE AND RACES: CASES AND RESOURCES FOR A DIVERSE AMERICA 429-54 (2000).

8. I use the term “white supremacist” to refer to an array of groups and individuals who include, but are not necessarily limited to, Nazis, Skinheads, Klansmen, and the like. The white supremacist is an individual who views non-whites and Jewish people as the enemy and who believes that whites are inherently superior to non-whites. The white supremacist rhetoric and propaganda describe a disdain for non-whites that is nothing short of hatred and suggest an extreme social, economic, and often times violent agenda for the genocide of such groups. For readings on white supremacy, see JOHN GEORGE & L AIRD WILCOX, AMERICAN EXTREMISTS (1996) and J AMES RIDGEWAY, BLOOD IN THE FACE: THE KU KLUX KLAN, A RYAN NATIONS, NAZI SKINHEADS, AND THE RISE OF A NEW WHITE CULTURE (2d ed. 1995).


our civil rights and equal protection jurisprudence in this country is history, and our history regretfully incorporates the narrative of racism. Although racism is an ancient concept that long pre-dates the birth of our United States Constitution, American racism is a unique narrative. Our long history of racism needs not to be recapitulated here in order to be accepted and understood as an integral part of our contemporary view of law and culture. Suffice it to say that our nation embraced the notion of racial inferiority for generations, and upon attempting to correct its evil mistake, organizations such as the infamous Ku Klux Klan (hereinafter “KKK”) were formed in an effort to maintain the status quo of white privilege and to resist the ever-evolving political, social and jurisprudential trend toward racial equality.

In addition to the KKK, various other groups have perpetuated
the cultural myth of white supremacy. Christian Identity, Aryan Nation, and Skinheads are but a few of the better-known organized groups that profess white supremacy and disdain for all non-white and Jewish people. Perhaps less well-known is The World Church of the Creator (“World Church”), which is a largely internet-based organization that promotes white supremacy.

At the time this Article was written, the World Church was headed by Matthew Hale. Hale is a graduate of Southern Illinois University School of Law who passed the bar examination and applied for admission to the Illinois bar. Despite his successful performance in both law school and on the bar exam, Hale was denied admission to the Illinois bar because of a finding by bar authorities


15. Members of the World Church subscribe to 16 “religious” commandments, and among them are the following:

IV. The guiding principle of all of your actions shall be: What is best for the White Race?

. . .

VI. Your first loyalty belongs to the White Race.

. . .

VII. Show preferential treatment in business dealings to members of your own race. Phase out all dealings with Jews as soon as possible. Do not employ niggers or other coloreds.

THE WORLD CHURCH OF THE CREATOR MEMBERSHIP MANUAL, available at http://www.creator.org/manual (last visited Mar. 16, 2003). “[W]hat is good for the White Race is the highest virtue, and what is bad for the White Race is the ultimate sin.” Id. Although Hale’s public speeches as head of the organization outwardly condemn violence against non-white people, he does threaten that violence will erupt if his organization and its beliefs are not heeded. For example, after being denied membership to the Illinois bar and prior to his appeal to the Illinois Supreme Court, Hale stated, “If the courthouse is closed to Non Approved Religions, America can only be headed for violence.” See Press Release, Illinois State Bar Forbids Freedom of Speech and Religion! (July 5, 1999) (on file with author) (quoting Hale). Indeed, two days after Hale was denied admission to the bar, Benjamin Nathaniel Smith, who had been a member of Hale’s World Church, went on a weekend shooting spree of non-whites in Illinois and Indiana, killing two people before killing himself. See Mark Skertic & Abdon M. Pallasch, Spree Ends, Alleged Gunman Kills Self Near Salem, Ill., CHI. SUN-TIMES, July 5, 1999, at 1. After the killings, Hale said that it was possible that the shootings were the result of the character and fitness committee’s denial of his application for a license to practice law. See Court rejects Hale’s appeal for law license, St. J.-REG. (Springfield, Ill.), June 27, 2000, at 9. In an interview documented on videotape by the Southern Poverty Law Center, Matthew Hale expressed no regret or sympathy for this loss of non-white life. For more information on the World Church’s racist propaganda, see Richard L. Sloane, Note, Barbarian at the Gates: Revisiting the Case of Matthew F. Hale to Reaffirm that Character and Fitness Evaluations Appropriately Preclude Racist from the Practice of Law, 15 GEO. J. LEGAL ETHICS 397, 418-20 (2002).

that his racist activities both prior to and during his leadership of the World Church demonstrated a lack of moral character, rendering Hale unfit to practice law.\textsuperscript{17} In the judgment of this author, \textit{In Re Hale}\textsuperscript{18} was correctly decided, and it stands for the proposition that the bar may legitimately exclude a person from the legal profession if that person participates in racist activities because such activities obstruct the administration of justice and undermine the core values of the legal profession. Nonetheless, the \textit{Hale} decision has been widely criticized as abridging First Amendment rights.\textsuperscript{19} It is because

\begin{itemize}
\item On December 16, 1998, the Illinois Bar Inquiry Panel refused to certify that Hale had the necessary moral character and fitness to practice law. Hale appealed the Inquiry Panel’s decision to the Illinois Committee on Character and Fitness (“Committee”), which reviewed Hale’s application and, after a hearing, concluded that Hale failed to meet the requirement of “good moral character and general fitness to practice law” as that requirement has been interpreted in Illinois. See \textit{Geoffrey C. Hazard Et Al., The Law And Ethics Of Lawyering} 875-84 (3d ed. 1999) (reporting the Committee’s opinion). Hale appealed the Committee’s decision to the Supreme Court of the State of Illinois, which dismissed the appeal. \textit{In re Hale}, 723 N.E.2d 206 (Ill. 1999). He subsequently filed a petition for certiorari to the United States Supreme Court, which was denied without opinion. Hale v. Comm. on Character & Fitness of the Ill. Bar, 530 U.S. 1261 (2000). The decision to deny Hale admission to the bar could have been based on the fact that Hale failed to disclose on his application several arrests and disciplinary actions taken against him. See Bob Van Voris, \textit{Muddying the Waters: Illinois Racist’s Free Speech Case is Complicated by His Arrest Record}, NAT’L L.J., Feb. 21, 2000, at A1. Failure to disclose information in a bar application can and has served as the basis for denying admission. See, e.g., \textit{In re Florida Bd. of Bar Exm'n's ex rel. John Doe}, 770 So. 2d 670 (Fla. 2000) (denying admission to Florida bar for an applicant’s failure to disclose pending criminal charges); \textit{In re Cunningham}, 502 N.W.2d 53 (Minn. 1993) (denying admission to the Minnesota bar for an applicant’s failure to disclose an arrest and paternity proceedings). Moreover, the committee could have cited poor judgment as demonstrated by a letter Hale sent to a woman expressing the opinion that the “nigger race” is “inferior” and inquiring, “Is it going to take your rape at the hands of a nigger beast or your murder before you become aware of the problem?” See Brief of Amicus Curiae Illinois State Bar Association at 8-9, \textit{In re Hale}, 723 N.E.2d 206 (Ill. 1999), cert. denied, 530 U.S. 1261 (2000) (on file with author). Despite the existence of sufficient “benign” reasons to deny Hale admission, the Illinois Inquiry Panel chose to focus on the more disturbing aspect of Hale’s application—the fact that he was an admitted white supremacist. For a detailed reporting of the facts and circumstances surrounding Hale’s denial of admission to the bar, see Emelie E. East, \textit{Note, The Case of Matthew F. Hale: Implications for First Amendment Rights, Social Mores and the Direction of Bar Examiners in an Era of Intolerance of Hatred}, \textbf{13} GEO. J. LEGAL ETHICS 741 (2000). Unlike the cited piece, the instant Article focuses on the ethics of allowing a white supremacist to practice law and offers a legal rationale to support the ethical conclusion.
\item See W. Bradley Wendel, \textit{Free Speech for Lawyers}, \textit{28} HASTINGS CONST. L.Q. 305, 320 (2001) (discussing the application of free speech principles to the legal profession and arguing that anti-discrimination values should not be preferred to First Amendment values). Wendel acknowledges the odious nature of Hale’s belief in white supremacy, but argues that these beliefs are constitutionally protected. Id.; see also W. William Hodes, \textit{Accepting and Rejecting Clients—The Moral Autonomy of the Second-to-the-Last Lawyer in Town}, \textit{48} U. KAN. L. REV. 977, 988-89 (2000) (acknowledging that he has criticized the Hale decision as implicating the First Amendment); Nadine Strossen, \textit{Incitement to Hatred: Should There Be a Limit?}, \textit{25} S. ILL. U. L.J. 243 (2003) (arguing that the First Amendment precludes bar authorities from discriminating against persons who express racist views).
\end{itemize}
of this criticism that I feel compelled to write this Article.20

This Article explores the issue of whether Klansmen or other white supremacists should be lawyers. It contends that white supremacists should not be lawyers and uses the example of Matthew Hale to illustrate its points. To make its case, this Article looks first in Part II to the definition of moral character in an effort to determine whether white supremacists such as Matthew Hale possess the requisite moral character to be lawyers. Parts II.A and B examine cases in which candidates for admission to the bar have been denied admission on the basis of moral character in an effort to locate the rationale for the moral character requirement. It argues that this rationale is furthered by the exclusion of white supremacists such as Hale from the bar. Part II also looks to relevant case law to determine whether there is precedent for holding an applicant morally unfit to practice law because of conduct that is arguably protected by the First Amendment. In doing so, Part II.C argues for a balanced application of the First Amendment by balancing the state’s interest in providing a system of justice free from racism against the individual’s free speech interests. It also argues for harmonization of the competing interests of First Amendment liberty with Fourteenth Amendment protections in the bar admission context.

Next, Part II.D looks at the “professional role obligations”21 of lawyers and asks whether a white supremacist like Hale can meet the obligations of the profession while practicing his racist “religion” or beliefs. It explores several additional questions. May an applicant’s personal moral code in any way contradict the profession’s code of ethics? May an applicant’s political and/or religious beliefs and/or agenda be contrary to existing law? Part III seeks to distinguish the rest of us from the white supremacists in an effort to explain why many candidates, despite their beliefs and biases, meet the moral character requirement while others do not. Part IV proffers a policy rationale for the Hale decision and contends that the profession would be harmed by the admission of white supremacists. To make this case, Part V argues that the admission of a person to the bar who does not subscribe to the core value of equality would illegiti-

20. Since his denial of admission in Illinois, Hale has been on a forum-shopping spree and has sought admission in Montana and Ohio, both of which denied his application. See Anderson v. Hale, 2001 U.S. Dist. LEXIS 3774 (N.D. Ill. Mar. 29, 2001) (mentioning Hale’s denial of admission in Montana in late March 2001). Although no jurisdiction had admitted Hale at the time this Article was nearing completion, the potential remains for Hale or other white supremacists to be admitted to the bar either because the white supremacist does not have a record containing sufficient grounds for denial that are unrelated to beliefs or because the jurisdiction in question views the First Amendment as an obstacle to precluding admission of an open racist.

mize a fundamental principle of our system of justice: the task of administering equal justice to all.

Part VI acknowledges that there are cases in which the application of the proposed rationale of Parts II and III would be more difficult than it was in the Hale case. It discusses those difficult cases and seeks to distinguish or harmonize the more difficult cases with the Hale decision. Part VI acknowledges that the term “racist” is defined according to degree. Ultimately, Part VI seeks to deconstruct discourses of racism and racist identity in a critical manner to clarify the complex relationship between the two and to identify how racist lawyers or bar applicants should be identified and treated by bar authorities to best further the principles of the profession.

This Article attempts to provide a basis for removing from racism the veil of protection it sometimes has received from the First Amendment in the bar admission context.

II. DEFINING MORAL CHARACTER AND FITNESS

White supremacists lack the requisite moral character and fitness to be lawyers. All United States jurisdictions require that applicants for admission to the bar possess and demonstrate good moral character.22 The issue of whether a candidate for admission to the bar is of good moral character is an important one given the degree of discretion that lawyers inevitably exercise in even the most routine of law practices.23 Good moral character is required of lawyers in order to protect the orderly administration of justice.24 “Since our ‘fortunes, reputations, domestic peace . . . nay, our liberty and life itself’ rest in the hands of legal advocates, ‘[t]heir character must be not only without a stain, but without suspicion.’”25 The moral character requirement is intended to ensure that the applicant, if admitted to the bar, would not obstruct the administration of justice or otherwise act unscrupulously in his or her capacity as an officer of the court.26

Good moral character—what does that mean?27 The inherent prob-


27. For a brief history of the “good moral character” requirement and an argument in favor of its retention, see Sloane, supra note 15, an article that focuses solely on the Hale
lem with assessing moral character in any instance is that there is no firm definition of the term. The concept is fluid and quite subjective.28 Conduct that may demonstrate to one individual that the applicant lacks moral character will have little or no significance in another person’s analysis of whether the applicant has the requisite moral character to be a lawyer.29 Reasonable minds can disagree about what constitutes “moral character.”30 Any person who knowingly and intentionally practices racism lacks moral character because racism31 is immoral.32 Assuming reasonable minds can agree on what type of conduct is immoral, the inquiry does not stop there. Adultery is immoral behavior, but if it were used as a litmus test for disqualifying bar applicants and practicing lawyers, there would be a significant decrease in the number of lawyers. Accordingly, not all immoral behavior serves as a disqualifier for purposes of determining bar admission. So what type of immoral conduct is acceptable in lawyers and what type is not?

Courts have offered some guidance in defining and applying the amorphous criterion of moral character.33 In order for behavior that society characterizes as immoral to become relevant to bar authorities, there must be a nexus between the alleged immoral behavior and the practice of law.34 In other words, if the immoral behavior will not reflect adversely on the person’s fitness to practice law, then it is case. This article uses the Hale case to construct a framework for discussion of the larger questions underlying the Hale case in an effort to aid in the analysis of future cases similar to Hale.

28. Rhode, supra note 25, at 529.
29. Id. at 530-31.
30. Id. at 507.
31. Defining racism is beyond the scope of this Article. However, I use the term to encompass racist beliefs (the belief that a person’s race makes that person inferior to other people) that necessarily determine one’s practices and behavior and thereby render the latter action racist. For a discussion of the definition of racism, see PEREA ET AL., supra note 7, at 6-49.
32. See Dr. Martin Luther King Jr., Letter From Birmingham Jail (Apr. 16, 1963), available at http://www.mlkonline.com/jail.html (last visited Mar. 16, 2003) (arguing that racism in the form of segregation is morally wrong); see also HAZARD ET AL., supra note 17, at 878 n.4 (citing In re A.L., 702 N.E.2d 1021, 1024-25 (Ill. App. Ct. 1998), in which “the Appellate Court cited bigotry as evidence of depravity which has been defined as the opposite of good moral character,” and In re Abdullah, 423 N.E.2d 915, 917 (Ill. 1981), which held that “depravity is ‘an inherent deficiency of moral sense and rectitude’”); Charles R. Lawrence III, The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism, 39 STAN. L. REV. 317, 322-23 (1987) (discussing how society has come to accept the idea that racism is immoral).
33. It is also important to note that the applicant bears the burden of proving that she is of good moral character, so if the applicant fails to meet this burden of proof, denial of admission to the bar is appropriate. Konigsberg v. State Bar of Cal., 366 U.S. 36, 40-41 (1961).
34. In re Eimers, 358 So. 2d 7, 10 (Fla. 1978) (holding that while homosexuality "affronts public conventions," there is no substantial nexus between homosexual acts and the ability of the applicant to live up to the professional responsibility and conduct required of a lawyer).
not the type of immoral conduct that disqualifies a candidate from bar admission. For example, adultery, although still considered immoral behavior in our society, does not reflect adversely on a candidate’s fitness to practice law—one can be an adulterer and a competent and effective lawyer at the same time. Unlike adultery, the practice of racism is a type of immoral behavior that is integrally connected to the practice of law. A lawyer who commits adultery in no way harms his clients, the profession, or his ability to provide competent and loyal representation. On the other hand, a lawyer who practices racism is likely to harm non-white clients either by refusing them representation or providing them with less than loyal and competent representation through reduced zeal caused by his personal conflict of interest arising from racism. A lawyer who practices racism is also likely to have a detrimental impact on the public image of the profession. Members of the public, particularly minority members, will likely view our system of justice as less fair if it is administered by racially biased lawyers and judges.

In *Konigsberg v. State Bar of California*, the United States Supreme Court defined the parameters of the moral character inquiry as seeking to determine whether “a reasonable [person] could fairly find that there were substantial doubts about [an applicant’s] ‘honesty, fairness and respect for the rights of others and for the laws of the state and nation.’” Admittedly, the moral character requirement is a subjective standard for determining fitness to practice law. Most bar admission cases that seek to determine moral character fall within one or more of the following categories: political/religious belief and conduct; “misconduct in the bar admission process; past illegal conduct; financial malfeasance; and emotional or mental instability.” Both past illegal conduct and misconduct in the bar admission process are non-controversial factors that routinely have been used as a basis for determining that a candidate does not possess the requisite moral character for admission to the bar. Factors that indicate that a candidate is dishonest are generally agreed upon as legitimate grounds for denial of membership to the bar. The more controversial factors that bar authorities have used to reach the conclusion that an applicant does not possess the requisite moral character include

35. *In re Haukebo*, 352 N.W.2d 752, 754-55 (Minn. 1984).
36. This statement assumes, of course, that the lawyer’s accomplice/partner in the adulterous act is not his/her client.
39. *Id.* at 264
40. See McChrystal, *supra* note 22, at 73.
membership in controversial organizations and repetitive use of speech and/or the media to personally attack others. These factors are controversial because they involve the bar candidate's exercise of First Amendment rights such as the right of association or the right to practice one's religion without governmental intrusion. Thus, these factors have provoked disagreement about whether conduct that is arguably protected by the First Amendment, such as advocating and practicing private racism and discrimination, can be considered by bar authorities in their assessment of an applicant's moral character. This question must be answered in the affirmative if the profession is to preserve its integrity and its core values, and thereby not to undermine the fair administration of justice.

A. The First Amendment as a Shield, Not a Sword

Those who argue that First Amendment rights trump all others, frequently referred to as "free speech absolutists," argue that the First Amendment guarantees to white supremacists such as Hale the right to say whatever they choose and to associate with whomever they choose without governmental consequence or punishment. Those who adopt this myopic absolutist view are willing to suffer any social evil in order to prevent any restriction of speech. Such a view stretches the First Amendment beyond its reasonable limits by allowing the First Amendment to subvert other important constitutional guarantees. Even acknowledging the individualistic rights-based stance of American jurisprudence, our Supreme Court has recognized that the interest in free expression must yield at times to other more important considerations. The First Amendment should be interpreted as a shield that protects us all, not a sword that is used to protect some while disemboweling others. Restrictions on the

41. Konigsberg v. State Bar of Cal., 366 U.S. 36 (1961) (upholding the denial of admission of an applicant who argued that the California Bar Commission violated the applicant's First Amendment rights by forcing him to reveal whether he was affiliated with the Communist Party).
42. In re Converse, 602 N.W.2d 500 (Neb. 1999) (denying admission to a bar candidate due to several instances of using letters, the newspaper, and artwork on t-shirts to personally attack his opponents in disputes).
43. At least one court has agreed with this proposition in holding that a bar Commission may consider conduct arguably protected by the First Amendment when investigating whether an applicant possesses the requisite moral character and fitness to practice law.
45. For an enlightening and prolific argument that the First Amendment is not absolute and should not be used as a sword to permit harm to defenseless people, see id. See also RICHARD DELGADO & JEAN STEFANCIC, MUST WE DEFEND NAZIS? (1997).
46. I recognize that the First Amendment has been a tool used by minorities for the advancement of their interests. Protesters during the Civil Rights movement were to some extent protected by the First Amendment. See DERRICK BELL, RACE, RACISM AND AMERICAN LAW 653-734 (Chapter 11, entitled Parameters of Racial Protest) (4th ed. 2000).
content of speech are imposed when the social value of the speech is significantly outweighed by the harm it imposes. Such restrictions include, but are not limited to, those on child pornography, obscenity, defamation, fighting words, sexually harassing speech, and false or misleading advertising.

In the case of allowing white supremacists into the bar, the value of racist speech is outweighed by the need to protect minority citizens from the high risk of harm, just as the right of free association was outweighed by the interest in racially integrated public schools. The First Amendment must be balanced against other competing interests guaranteed by the Constitution. If First Amendment principles were absolute and could not be balanced against the often competing interest of promoting equal protection under the law, we could not have made strides in achieving integrated public schools in this country. The arguments that free speech absolutists make in oppo-

48. Osborne v. Ohio, 495 U.S. 103 (1990) (upholding an anti-child pornography statute that was challenged on First Amendment grounds); see also Victor C. Romero, Restricting Hate Speech Against “Private Figures”: Lessons in Power-Based Censorship from Defamation Law, 33 COLUM. HUM. RTS. L. REV. 1, 13 (2001).
49. See Roth v. United States, 354 U.S. 476 (1957) (holding that obscenity is not protected by the First Amendment); see also Miller v. California, 413 U.S. 15 (1973) (affirming the Roth decision that obscenity is not protected by the First Amendment).
51. Chaplinsky, 315 U.S. at 571-72 (upholding a conviction on the ground that fighting words are not protected under the First Amendment). But see Romero, supra note 48, at 13 (noting that at least one scholar has argued that Chaplinsky should be overruled by citing Note, The Demise of the Chaplinsky Fighting Words Doctrine: An Argument for Its Interment, 106 HARV. L. REV. 1129 (1993)).
52. Title VII regulations forbid sexually harassing speech, even when unaccompanied by conduct, to the extent that such speech creates a hostile work environment for the employee. 29 C.F.R. § 1604.11(a) (2001). Racist speech in the workplace is also banned to the extent that such speech causes a hostile work environment. See Daniels v. Essex Group Inc., 937 F.2d 1264 (7th Cir. 1991) (finding that an employee has a cause of action under Title VII when persistent racist speech creates a hostile work environment for the employee); see also Andrews v. City of Philadelphia, 895 F.2d 1469, 1482 (3d Cir. 1990) (listing the elements of a claim for a racially hostile work environment).
53. See Friedman v. Rogers, 440 U.S. 1 (1979) (upholding a provision of the Texas Optometry Act that prohibited the practice of optometry under a trade name and holding that the provision did not violate the First Amendment because it furthered the state’s interest in protecting the public from deceptive and misleading use of optometric trade names).
54. Admittedly, I do not see any value in racist speech, but I use this framework for those who do.
55. Indeed, Brown v. Board of Education, 347 U.S. 483 (1954), demonstrates that any First Amendment right of white parents to choose to have their white children associate exclusively with other white children must necessarily yield at the point that exercise of this First Amendment right tramples the Fourteenth Amendment rights of children of color. I do not contend that the mandates of Brown have ever been fully achieved in this country, but to the extent that there are no longer positive rules of law that preclude children of color from attending predominately “white” public schools, I will acknowledge that public schools are integrated at least in theory. For a recent study indicating that segrega-
sition to the Hale decision are reminiscent of arguments made by opponents of public school integration in the 1950s. During the 1950s, those who practiced racial bigotry used First Amendment absolutism to argue in favor of continued segregation by claiming that they and their children had the First Amendment right to associate or not associate with persons of their choosing. Accordingly, they argued that forced integration administered through court orders, busing programs, and police escorts violated their First Amendment rights to free association. Proponents of free speech absolutism argue that there should be no government regulation of speech, and to the extent that regulation is imposed, it should be content-neutral, meaning that the government should not regulate the content of speech or impose regulations based on the content of the speech. This view ignores the harm that a white supremacist is likely to inflict upon the legal profession and the administration of justice, not to mention our non-white citizens who are subject to our justice system. Like the majoritarian group, racial minorities have a right to a fair justice system that does not make judgments based on race.


56. Segregationists found support in the writings of constitutional scholar Herbert Wechsler. See Herbert Wechsler, Toward Neutral Principles of Constitutional Law, 73 Harv. L. Rev. 1, 34 (1959) (arguing that integration forces an association upon those who find it unpleasant or repugnant). For an extant articulation of the argument that people should have an absolute right to not associate, see Robert W. McGee, The Right to Not Associate: The Case for an Absolute Freedom of Negative Association, 23 UWLA L. Rev. 123 (1992). Of course, outside the realm of racial discrimination, the United States Supreme Court has recognized that the First Amendment right to free association may trump a state’s desire to eliminate sexual orientation discrimination. See, e.g., Boy Scouts of Am. v. Dale, 530 U.S. 640 (2000).

57. Wechsler, supra note 56, at 34.


59. For a discussion of the “costs” of absolute tolerance of speech and how society’s tolerance of extremist speech may become an excuse for excessive intolerance in other areas, see Lee C. Bollinger, The Tolerant Society 58-75 (1986).

60. Even our highest court has acknowledged that the “particular character of the Klan’s activities, involving acts of unlawful intimidation and violence” warrants the allowance of some intrusion into the First Amendment veil if the government can establish a compelling justification for so doing. NAACP v. Alabama, 357 U.S. 449, 465 (1958).
mission of racists to the bar. As Derrick Bell and Charles Lawrence have pointed out, it is not uncommon for the white majority to sacrifice the constitutional rights of non-whites in order to preserve the interests of the majority. In this case, the most obvious beneficiaries of a judicial system free of white supremacist administrators are people of color and perhaps Jews. However, excluding white supremacists from the legal profession will benefit all of us, not just people of color. The benefit to whites in yielding First Amendment liberty in this context will be a system of justice that is less contaminated with racism. Why should whites desire a justice system less contaminated by racism? What benefits to whites would be derived by such a system? The system would be more efficient and effective because race would become less of a factor in administering justice. For example, if Mark Furman had not been a racist police officer, O.J. Simpson would probably be in prison now. By allowing a racist to enforce our laws, society paid a price—we allowed racism to create reasonable doubt in the minds of the jurors. A racist police officer has a motive to frame a black man, whereas a fair-minded police officer has no more motive to frame a black man than any other person. Thus, a system of justice not administered by racists makes it less likely that actions of administrators will be questioned as being racially motivated and more likely that the actions of our lawyers will

61. Charles R. Lawrence III, If He Hollers Let Him Go: Regulating Racist Speech on Campus, 1990 DUKE L.J. 431, 478 (arguing that some civil libertarians minimalize the harm to victims of racist speech by characterizing racial harassment as isolated incidents in an otherwise racism-free community).

62. See id. at 475 (citing BELL, supra note 46, at 30).

63. In this Article, I sometimes use the terms “black” and “African-American” interchangeably to refer to people of African descent. I elect to use the term “black” rather than African-American in most instances because “black” is broader in the sense that it encompasses people of African heritage who may not be American, but who nonetheless experience racism in America.

64. Examples of challenges to decisions rendered by our system of justice on the basis of perceived racism abound. See RANDALL KENNEDY, NIGGER: THE STRANGE CAREER OF A TROUBLESOME WORD 61-63 (2002) (discussing cases in which decisions of judges have been overturned due to the judges’ use of the “N” word and citing United States v. Brown, 539 F.2d 467, 468 (5th Cir. 1976)). Kennedy also discusses the removal of an elected district attorney for his use of the “N” word. Id. at 65-72. Kennedy opines that the district attorney’s removal was justified because statements he made rendered him unfit to fulfill his public responsibility. Such a responsibility entails a commitment to the idea that all people, regardless of race, should be treated equally and with respect before the bar of justice. By calling [the defendant] Ray Jacobs a nigger, Jerry Spivey [the district attorney] cast a pall over public confidence in his commitment to accord all people due respect regardless of race. Id. at 70-71 (citing In re Spivey, 480 S.E.2d 693 (N.C. 1997)). For a current example of how racism in the judicial process creates inefficiency and causes doubt about the validity of judicial action, see Sara Rimer, In Dallas, Dismissal of Black Jurors Leads to Appeal by Death Row Inmate, N.Y. TIMES, Feb. 13, 2002, at A24.
be perceived as rational, fair-minded decisions. And from that we all benefit.

B. **Lawyers’ First Amendment Freedoms Are Already Restricted When Necessary to Promote the Interests of Justice**

The denial of bar admission to Matthew Hale is neither the first nor the only instance in which bar authorities have imposed restrictions on the First Amendment freedoms of lawyers. Lawyers already surrender some First Amendment rights in exchange for the privilege of administering justice. There are several ethical rules that impose limitations on the First Amendment freedoms of lawyers. The rules pertaining to advertising and client solicitation restrict lawyers’ First Amendment freedoms.65 These rules restrict commercial speech, which is subject to the less stringent tests of time, place, and manner.

There are other instances where non-commercial speech of lawyers is restricted in an effort to safeguard the administration of justice.66 Rule 3.5 of the *Model Rules of Professional Conduct*, as drafted by the American Bar Association and adopted by many jurisdictions, provides that a “lawyer shall not: seek to influence a judge, juror, prospective juror or other official by means prohibited by law; communicate ex parte with such a person except as permitted by law; or engage in conduct intended to disrupt a tribunal.”67 The rule expressly restricts a lawyer’s First Amendment right to free speech by restricting communications that lawyers may have with jurors and judges. The rule also prohibits speech that is intended to disrupt the decorum of the courtroom.68 The rule has a dual purpose, which is to prevent prejudice to the adjudicative process and to protect and promote the impartiality—as well as the appearance of impartiality—and fairness of the judicial system. Despite constitutional challenges to the rule, it has been upheld as not violative of a lawyer’s right to free speech.69

65. See *Model Rules of Prof’l Conduct* Rs. 7.1, 7.3 (2001).

66. See id. R. 4.2 (prohibiting a lawyer from communicating with a person whom the lawyer knows is represented by another lawyer); see also id. R. 3.6 (prohibiting a lawyer from making an “extrajudicial statement” if the lawyer should know that the statement would have a “substantial likelihood of materially prejudicing an adjudicative proceeding”); id. R. 3.8(g) (supplementing Rule 3.6 by precluding a prosecutor from making an extrajudicial statement that would have a “substantial likelihood of heightening public condemnation of the accused”).

67. Id. R. 3.5.

68. See State v. Swisher, 676 S.W.2d 576 (Tenn. Crim. App. 1984) (providing an example of how a lawyer may be sanctioned by the court for speaking in court when said speech disrupts the decorum of the court).

Likewise, political speech of lawyers is restricted by ethical rules in the context of judicial election campaigns. Lawyers who are candidates for judicial office are precluded from expressing a political position.\textsuperscript{70} The obvious purpose of this rule is to promote the public's confidence in the rule of law and our judicial system. By prohibiting political statements of judicial candidates, the state ensures that the public views judicial candidates as fair and impartial decision-makers rather than as political decision-makers. Public confidence is essential to our system of justice because our legal system depends upon support from the public to maintain its authority.\textsuperscript{71} The exclusion of white supremacists from the bar serves a purpose similar to the purpose of Model Rule 3.5 and the rule governing political speech of judicial candidates. That purpose is to prevent racial prejudice from affecting the judicial process, thereby promoting the public's confidence in the rule of law derived from the judicial process. A system of justice that allows a devout racist to administer justice to a citizenry that includes people who are the object of the racist's hatred and discriminatory practices promotes the appearance of unfairness in our justice system and will cause a loss of public confidence in the system.

The decision in \textit{Hale} is supported also by the decision in \textit{Bob Jones University v. United States},\textsuperscript{72} which held that the government has a “fundamental, overriding interest in eradicating racial discrimination in education” that “substantially outweighs whatever burden [that the] denial of tax benefits” placed on the University's exercise of its religious beliefs.\textsuperscript{73} Likewise, a state has an overriding interest in eradicating racial discrimination in the administration of justice that substantially outweighs whatever burden the denial of a law license may place on a white supremacist's freedom of expression or exercise of religion. The \textit{Hale} decision is not censorship. Like the restriction on speech in \textit{Bob Jones}, excluding Hale from the bar is merely an incidental restriction on speech. In other words, the state's objective in excluding a racist from the bar is to eradicate racial discrimination in the administration of justice, not to censor speech. In-


\textsuperscript{72} 461 U.S. 574 (1983) (upholding an IRS ruling denying section 501(c)(3) tax exempt status to the private Bob Jones University on the ground that the university practiced racial discrimination that was contrary to the public policy promoted by the tax exemption for charitable organizations). The university prohibited interracial dating and interracial marriage and denied admission or expelled any student who was engaged in such an interracial relationship. \textit{Id.} at 580-81.

\textsuperscript{73} \textit{Id.} at 604.
cidental limitation on First Amendment freedom is permissible and justified when the government regulation furthers a substantial governmental interest and that governmental interest is unrelated to the suppression of free expression.74

Not only is the restriction on speech in the Hale case incidental, it is insignificant when compared to the interest that the state is seeking to promote—eradicating racism in the administration of justice. The incidental restriction on racist speech in this context is a permissible restriction on speech because it leaves open alternative avenues for the white supremacist to communicate his message of racial superiority and hatred75 toward minorities. Matthew Hale is free to continue to practice his “religion,” hold his beliefs, and say whatever he pleases. Moreover, the state’s interest in eradicating racism in the administration of justice is compelling and outweighs the individual racist’s interest in practicing law. Free speech absolutists, who argue in favor of admitting racists to the bar, either do not consider important values expressed elsewhere in the Constitution or elevate First Amendment rights over other constitutional rights.76 First Amendment rights should not undermine the administration of justice nor trample Fourteenth Amendment rights. In First Amendment jurisprudence, like much American jurisprudence, courts make value judgments and determine the confines of the law based on those preconceived values. Unfortunately, protecting the rights of the colored minority at the expense of the privileged majority has never been a popular cause. Nonetheless, if we value equality and truly want to strive to be a color-blind nation with color-blind justice, we must elevate the right of the minority to be free from racial oppression over the right of the majority to inflict and/or ignore such racial oppression under the auspices of expression. The refusal of our courts to restrict hate speech is an example of how our system of justice makes such value judgments.77 Hate speech, like pornography, has no societal worth, yet it is valued and protected over the rights of those who

75. Matthew Hale, through his use of the WHITE MAN’S BIBLE, admits that the Credo of the World Church is to hate “Jews, blacks and other colored people,” but argues that loving the “White Race” makes hate for the enemies of the white race inevitable. According to the WHITE MAN’S BIBLE, those who profess love for their enemies are liars and hypocrites. See BEN KLASSEN, WHITE MAN’S BIBLE: CREATIVE CREDO NO. 71, available at http://www.creator.org/holybooks/wmb/credo71.html (last visited March 26, 2002) (on file with author). I define the term “hatred” as the absence of love. For those who need a definition of love, see 1 Corinthians 13:4 (New International).
76. See Lawrence, supra note 61, at 436-37 (noting that traditional civil libertarians do not take into account important values expressed elsewhere in the Constitution).
77. See, e.g., R.A.V. v. City of St. Paul, 505 U.S. 377 (1992) (striking down a bias-motivated crime ordinance as unconstitutional). Banning white supremacists from the bar is easier to defend than banning hate speech due to the unique nature of the legal profession and the limited and incidental nature of the restriction on speech.
are its target to be free from such harm. 78

To allow a white supremacist to become a lawyer would be to privilege the interests of racists over the interests of Jews and people of color. The white supremacist masquerading under the guise of the strict libertarian asks our system of justice to privilege his interest in promoting racism above the interests of people of color and the interests of our justice system, and he cites the First Amendment as the authority for granting such privilege. Meanwhile, Jews and people of color demand from our government a system of justice that is not infested with racist administrators who may intentionally and covertly sabotage their rights. The resolution of the First Amendment conflict, therefore, necessarily lies in determining ab initio whose interest is most valued by the state, or, assuming the interests are equally valued, whether such competing interests can be reconciled.

Communitarianism is perhaps the most rational mechanism for reconciling the competing interests of civil liberties and civil rights. 79 Communitarians believe that, in a rights-based society, rights have limits as well as concomitant responsibilities. 80 Underlying the theory of communitarianism is the idea that, as members of a community, we are morally obligated to undertake certain responsibilities, and that these responsibilities may exist without an imminent payback in the form of rights. 81 In other words, communitarianists recognize that individual rights cannot be absolute, but must be tempered by that which is in the best interest of the community, 82 which in this case is the legal profession and all those who are governed by our system of justice. The court in Hale applied a communitarian-type rationale by choosing first to recognize, and then second to value, the civil rights of those who come in contact with the legal profession, as well as by choosing to value the sanctity of the legal profession and its goal of administering justice.

C. Balancing First Amendment Rights With Other Constitutional Interests

The First Amendment co-exists with other rights in our Constitution—specifically, the Fourteenth Amendment, which obviously was

78. See Matsuda et al., supra note 44, at 79.
79. Linda E. Fisher, A Communitarian Compromise On Speech Codes: Restraining the Hostile Environment Concept, 44 Cath. U. L. Rev. 97, 122-25 (1994) (discussing the merit of campus speech codes and arguing that in the academic community, justice is attained when the institution balances the rights and needs of all of the members who comprise the academic community).
81. Id. at 653.
82. Id. at 654.
added to our Constitution after the First Amendment. Prior to the adoption of the Fourteenth Amendment, the United States Constitution was interpreted to permit essentially all types of racist conduct by the states. States were free to adopt laws that furthered the interests of whites while sacrificing the rights and interests of non-whites. The Fourteenth Amendment necessarily restricted the states’ rights to elevate the First Amendment rights of whites over the equality rights of non-whites. A basic maxim of statutory construction provides that to the extent that a prior provision of a positive rule of law conflicts with a subsequent provision, the prior provision must yield to the subsequent provision. 83 This canon reasonably presumes that the drafters of the rule of law were aware of the prior provision and therefore must have intended the subsequent provision to limit the prior one to the extent that the two provisions cannot be reconciled. 84 This elementary canon of statutory construction is no less applicable to our Constitution than it is to statutes enacted by Congress. 85 States should not permit First Amendment principles to impede the state’s effort to eradicate racial discrimination in the administration of justice. 86 Moreover, the First Amendment should not be used as a sword to cut away Fourteenth Amendment rights held by the group of people for whom the Fourteenth Amendment was deemed necessary. If a state permits self-proclaimed racists to administer justice within that state, the state is elevating the interests of free speech over the state’s interest in providing a system of justice free from racism. The state may also be denying equal protection of the law to the group of people who would be the obvious targets of lawyers’ racism. 87

Synthesized to their core, the cases applying the moral character requirement can be understood to hold that a candidate for admission to the bar is required to prove by clear and convincing evidence that he is honest, fair, and respectful of both the laws of this nation and the rights of others. Surely reasonable minds would agree that an applicant who advocates white supremacy through such notions as racial disenfranchisement, racial segregation, the denial of employment based on race, and even the deportation of all non-white United States citizens does not demonstrate respect for the rights of

84. Id.
85. See Bd. of Trs. of Univ. of Ala. v. Garrett, 531 U.S. 356, 364 (2001) (restating the holding of Fitzpatrick v. Bitzer, 427 U.S. 445, 456 (1976), that the Eleventh Amendment to the Constitution is necessarily limited by the enforcement provisions of the Fourteenth Amendment).
87. This argument is premised upon a group justice view of the Fourteenth Amendment that our current Supreme Court likely would refuse to adopt.
non-white people, nor a willingness to treat such people fairly. Accordingly, the unwillingness to deal fairly with non-whites and the lack of respect for the rights of such persons that render white supremacists such as Matthew Hale deficient in the area of moral character as that term has been applied to lawyers means he is thereby unfit to be a lawyer. At the heart of the moral character inquiry is the desire by bar authorities to ascertain whether the candidate is someone who can be trusted with the power of lawyers and who will treat people fairly so that “justice” is done. Bar authorities simply cannot trust that a racist lawyer will fairly administer justice to our citizens of color.

Supporters of white supremacist lawyers will undoubtedly point out that a lawyer need not subscribe to or believe in all the laws of our country. A Catholic applicant is not excluded from bar membership because she opposes abortion or the death penalty, both of which are legal in our country. Indeed, a lawyer need not support abortion rights laws or the imposition of the death penalty in order to be admitted to the bar. However, neither abortion nor the death penalty are core values of the legal profession. Unlike abortion or the death penalty, equal justice is a core value of the legal profession. Acceptance of and adherence to this core value are appropriate criteria for entry into the legal profession. All persons who desire to become an officer of the court must subscribe to the core values of the profession. An applicant who is not committed to equal justice for all is necessarily lacking the requisite moral character to become a lawyer because that applicant cannot demonstrate that he is fair and respectful of the rights of other people regardless of their race. If the applicant cannot demonstrate a commitment to equal justice for all, that applicant is not qualified to be an administrator of justice in a system of justice that is founded upon and committed to such a core value. A lawyer who is anti-abortion or anti-death penalty is qualified to administer justice because her political position or belief does not demonstrate a lack of commitment to the core values of the profession, which are competence, loyalty, confidentiality, and equal justice for all. A lawyer must respect the rights of our citizenry. Our Fourteenth Amendment grants all of our citizens the right to equal protection under the law. This right has been interpreted to mean that a person is entitled to fair and non-discriminatory administra-

88. Lawyers have more power than laypeople. Lawyers have the power to initiate civil lawsuits or criminal charges and to compel the appearance of witnesses and the production of documents. Lawyers even have the power to gain access to certain venues not open to the public.
tion of the law.\textsuperscript{89} A white supremacist does not respect the rights of other people unless those people are purely white. This lack of respect for the basic human rights of all people, a concept so fundamental to most organized governments of the world,\textsuperscript{90} makes the white supremacist lack the moral character required for all lawyers.

\textbf{D. Serving Two Masters}

The wisdom of Matthew 6:24 is infinite. “No man can serve two masters: for either he will hate the one, and love the other; or else he will hold to one, and despise the other.”\textsuperscript{91} Hale’s religious and/or personal beliefs require him to use his profession and its instrumentalities, principally the courts, to engage in discriminatory treatment of non-whites, yet the profession that he sought to enter is charged with administering equal justice and prohibits discriminatory treatment based on race. In essence, Hale seeks to serve two masters—his racist “religion” and the legal profession.

Lawyers are not merely hired guns who advocate their client’s position. Lawyers are also officers of the court and thereby public servants. Even a lawyer who has a “private practice” is a public servant—the public has entrusted him with the power of being a lawyer. It is the public that has awarded him the title “officer of the court” and the monopolistic privilege of administering justice. As such, a lawyer has certain responsibilities to the public. Many of those responsibilities are codified in the rules of professional conduct adopted by the jurisdiction in which the lawyer practices. The rules of professional conduct arguably establish the ethical floor beneath which no lawyer should fall. The rules outline the duties of a lawyer, which range from loyalty to the client to the duty of candor to the tribunal. The rules of professional responsibility proscribe certain conduct as unprofessional and therefore subject to discipline. A lawyer cannot pick and choose which of the rules he or she will follow. Accepting the rules and agreeing to adhere to them is an all-or-nothing proposition. To the extent that the lawyer’s personal morality conflicts with rules

\textsuperscript{89} See Batson v. Kentucky, 476 U.S. 79 (1986) (holding that a state that allows a lawyer to prevent a black person, on account of his or her race, from serving as a juror in the administration of justice violates the Equal Protection Clause).


\textsuperscript{91} Matthew 6:24 (King James).
of professional conduct, the latter wins.92

If admitted to the bar, Matthew Hale, like all lawyers, would have certain “binding professional role obligations”93 that would directly conflict with his personal and/or religious obligations as leader of the World Church. Lawyers should strive to promote equality in the legal system.94 This aspirational goal of lawyers would be ignored, if not expressly violated, by the white supremacist lawyer. More importantly, the white supremacist lawyer is likely to violate the non-aspirational rules of law governing lawyers. If an applicant’s religious canons mandate that the individual engage in conduct that violates the law of lawyering, that individual is presented with a dilemma and must choose which master he or she will serve, for it will not be possible for such an individual to comply with his or her religious canons and be a lawyer if those religious canons command conduct that will violate a rule of professional responsibility.

Thus, if an applicant’s religious beliefs mandate conduct that is contrary to the ethical obligations of an attorney, it is both fair and necessary to exclude such a person from the profession for the purpose of preserving the integrity of the legal system as a whole. Matthew Hale was faced with this very dilemma—he wanted to serve two masters: the World Church and the legal profession. Hale’s sworn pledge to the profession to uphold the laws of this country, including the rules of professional conduct, would have been in direct opposition to the canons of his church because many jurisdictions now prohibit lawyers from participating in discriminatory conduct in the management or operation of a law practice.95 Hale practiced a re-

93. See Wilkins, supra note 21, at 1580.
94. Phoebe A. Haddon, Education for a Public Calling in the 21st Century, 69 WASH. L. REV. 573 (1994) (asserting that lawyers have a professional obligation to promote equality in the legal system); see also MODEL RULES OF PROF’L CONDUCT pmbl. (as revised by the Ethics 2000 Commission and adopted by the ABA House of Delegates at its February 2002 meeting). The preamble provides that “all lawyers should devote professional time and resources and use civic influence to ensure equal access to our system of justice for all those who because of economic or social barriers cannot afford or secure [adequate] legal counsel.” See also id. R. 6.1.
95. CAL. R. PROF’L CONDUCT R. 2-400 (prohibiting discriminatory conduct in a law practice and providing in part that a member of the bar in the management or operation of a law practice “shall not unlawfully discriminate or knowingly permit unlawful discrimination on the basis of race, national origin, sex, sexual orientation, religion, age or disability in: hiring, promoting, discharging or otherwise determining the conditions of employment of any person; or accepting or terminating representation of any client”); see also CAL. BUS. & PROF. CODE § 125.6 (West 2003) (permitting the state to discipline any licensed attorney who refuses to perform legal services because of the potential client’s “race, color, sex, religion, ancestry, disability, marital status, or national origin”). Other jurisdictions have also adopted anti-discrimination rules. See, e.g., FLA. R. PROF’L CONDUCT R. 4-8.4(d); IDAHO R. PROF’L CONDUCT R. 4.4(a); ILL. SUP. CT. R. 8.4(a)(5); MICH. R. PROF’L CONDUCT R.
ligion that mandates discriminatory conduct against non-whites. Hale’s “religion” forbids him from hiring non-whites and mandates that his actions promote white supremacy, not racial equality. If the bar ignored his commitment and intended adherence to the racist precepts of the World Church, it would render the oath that we take as lawyers meaningless. Lawyers affirm to uphold the Constitution and promise to not maliciously hinder the cause of any person. If facts in the record demonstrate that an applicant lacks sincerity in professing a willingness to abide by the attorney oath, such facts may serve as the basis for denial of admission to the profession. If the candidate for admission is a devout racist who has pledged to do everything in his power to promote the white race and to subjugate all others, bar authorities are justified in doubting that applicant’s declarations of willingness to abide by the attorney’s oath.

Ignoring the fact that a bar applicant engages in racist practices is tantamount to ignoring the fact that an elementary school teacher applicant admittedly enjoys virtual child pornography. Does the fact that the teacher has no criminal record for engaging in real child

6.5; MINN. STAT. R. PROF'L CONDUCT R. 8.4; N.J. R. PROF'L CONDUCT 8.4(g); N.M. R. PROF'L CONDUCT R. 16-300; N.Y. CODE PROF'L RESPONSIBILITY DISCIPLINARY R. 1-102(A)(6); OHIO REV. CODE ANN., CODE PROF'L RESPONSIBILITY R. 1-102; R.I. SUP. CT. R. PROF'L CONDUCT R. 8.4(d); TEX. GOV'T CODE ANN. R. 5.08; WASH. R. PROF'L CONDUCT R. 8.4(g).

96. The World Church canons preclude a member/believer from hiring African-Americans. BEN KLASSEN, 2 NATURE'S ETERNAL RELIGION: THE SALVATION (1992) (see Chapter 4, The 16 Commandments) [hereinafter THE SALVATION]. In its writings, the World Church utilizes the pejorative term “niggers,” which, given its historical usage and meaning as well as the context in which it is used, has been interpreted by this author to be an intended reference to African-Americans.

97. Lawyers in many jurisdictions take an oath similar to the following:

I do solemnly swear (or affirm) that I will support, obey and defend the Constitution of the United States and the Constitution of this Commonwealth and that I will discharge the duties of my office with fidelity, as well as to the court as to the client, that I will use no falsehood, nor delay the cause of any person for lucre or malice.

42 PA. CONS. STAT. § 2522 (2000). Any person refusing to take the oath or affirmation shall forfeit his office. Id.

98. See In re Roots, 762 A.2d 1161 (R.I. 2000) (denying an applicant admission to the Rhode Island bar based in part on the fact that the applicant’s racist writings demonstrated insincerity in professing a willingness to take and abide by the attorney’s oath).

99. Absolutists argue that virtual child pornography, which is computer-generated images of children engaging in sexual acts, is protected by the First Amendment and is not subject to regulation because the pornography does not use real children and therefore does not afford the state the same harms analysis as pornography that utilizes real children. Child advocates counter this argument by arguing that children are harmed by even virtual pornography because permitting such material promotes a culture that often leads to pedophilic conduct by those who indulge in child pornography. The U.S. Supreme Court recently sided with the more absolutist view by holding that Congress cannot ban virtual child pornography. See Ashcroft v. Free Speech Coalition, 535 U.S. 234 (2002) (an opinion from which Justices O’Connor, Rehnquist, and Scalia dissented). For an enlightening discussion of how pornography is the practice of sex discrimination that should not be protected by the First Amendment, see Catharine A. MacKinnon, Pornography, Civil Rights, and Speech, 20 HARV. C.R.-C.L. L. REV. 1 (1985).
pornography or pedophilic conduct mean that we can entrust our children to this person? Do we cite to the First Amendment to protect the right of the teacher to work in our schools? Not if we value our children enough to protect them from the risk of harm of a potential pedophile. If there are facts in the record that demonstrate that an applicant lacks sincerity in professing a willingness to abide by the attorney oath, such facts may serve as the basis for denial of admission.\textsuperscript{100} Because Hale’s religious beliefs mandate racist discriminatory practices that would violate the rules of professional responsibility, his beliefs cannot be reconciled with the rules of our profession or the oath required by attorneys. Accordingly, this impossibility of reconciling the mandates of the two sets of canons supports the decision to deny Hale and other white supremacists admittance to the bar.

Even jurisdictions that do not have a rule expressly precluding discriminatory conduct by a lawyer may still impose discipline for such conduct under subsection (d) of the general misconduct rule of the \textit{Model Rules of Professional Conduct}\textsuperscript{101} or subsection (5) of Disciplinary Rule 1-102(A) of the \textit{Model Code of Professional Responsibility}.\textsuperscript{102} Those subsections prohibit a lawyer from engaging in conduct that is prejudicial to the administration of justice. Discriminatory conduct by a lawyer is prejudicial to the administration of justice because it causes the public to lose confidence in the system and it renders the system unfair.\textsuperscript{103} Hence, such conduct would violate the rules of professional responsibility even in the absence of an explicit anti-discrimination rule.\textsuperscript{104}

Devilish advocacy could perhaps force one to argue that a bar applicant such as Hale could legitimately hold these beliefs and still respect the rights of non-whites while practicing law. Hale supporters may ask, Why do we assume that Hale will violate the rules of professional conduct in his role as a professional? Can’t Hale set aside his personal hatred and the World Church mission for purposes of

\textsuperscript{100} \textit{In re Roots}, 762 A.2d 1161.

\textsuperscript{101} \textit{Model Rules of Prof’l Conduct} R. 8.4(d) (2001).

\textsuperscript{102} \textit{Model Code of Prof’l Responsibility} DR 1-102(A)(5) (1981).

\textsuperscript{103} \textit{In re Goodfarb}, 880 P.2d 620, 621-22 (Ariz. 1994) (finding the judge’s use of racial epithets to be prejudicial to the administration of justice because it caused many citizens to lose faith in the judge’s judgment).

\textsuperscript{104} \textit{See In re Williams}, 414 N.W.2d 394, 398-99 (Minn. 1987), \textit{appeal dismissed sub nom.} Williams v. Lawyers Prof’l Responsibility Bd., 485 U.S. 950 (1988) (publicly reprimanding a lawyer for his use of a racial slur against a Jewish adversary at a deposition and holding that such conduct violated rule 8.4(d) of the \textit{Minnesota Rules of Professional Conduct}; \textit{see also In re Stevens}, 645 P.2d 99 (Cal. 1982) (censuring a judge for repeated use of racial epithets that the court found to be prejudicial to the administration of justice); People v. Sharpe, 781 P.2d 659, 660 (Colo. 1989) (holding that a prosecutor who used racially hostile language to refer to a Hispanic defendant violated DR1-102(A)(6) of the \textit{Colorado Rules of Professional Conduct}; Florida Bar v. Uhrig, 666 So. 2d 887 (Fla. 1996) (holding that an attorney violated rule 4-8.4(d) of the \textit{Rules Regulating the Florida Bar} by mailing a racially insulting letter to an opposing party who was a minority).
practicing law and administering justice? This argument makes several assumptions that may lack veracity. Perhaps the most interesting assumption lurking behind this question is the assumption that Hale could and would set aside his racist beliefs and practices while practicing law. By making this assumption, those who support allowing white supremacists to be lawyers demonstrate that they would rather risk the civil rights of all non-white people before possibly restricting the liberties of a few white people. They also assume that the applicant who has adopted racist religious beliefs has done so only for purposes of his private life and has decided that his religious beliefs do not apply to his professional life, or that the applicant can serve both masters—the profession and the religion. Those who are serious about their religious beliefs try to apply the commandments of their faith to all aspects of their lives. Religious beliefs may be at the core of who the person is, and for white supremacists, racism is at the core of their being. Thus, we cannot assume that a white supremacist applicant will ignore the mandates of his racist religious canons while practicing law.

Assuming a white supremacist did promise to set aside his racist agenda while practicing law—and we could somehow be assured that this promise was not illusory and would be kept—there would still be problems with admitting such a person to the bar. As lawyers, we are charged with protecting the public and the image of the profession. If a white supremacist is admitted to the bar and thereby given the title of “officer of the court,” the image of the profession would be tarnished. No longer would officers of the court be perceived by the public as fair-minded, respectable ministers of justice. Public confidence in the profession would be diminished.

Even if a candidate promised to disavow racist canons while practicing law, such a promise might be no more than chimerical, and the risk of harm is not worth taking. Bar authorities are justified when excluding an applicant who poses a significant risk of engaging in hidden racial discrimination in the practice of law. Bar authorities, like the rest of us, know that a racist agenda can be carried out in a

105. In the Hale decision from the Illinois Supreme Court, Justice Heiple seemed to be concerned with this possibility in his dissent. See In re Hale, 723 N.E.2d 206, 206-07 (Ill. 1999).

106. Indeed the World Church commandments mandate that Hale apply his racist beliefs to all aspects of his life, including his professional life. See THE SALVATION, supra note 96.

107. Indeed, Hale appeared unwilling to disavow the World Church canons even within the confines of law practice. Ironically, if Hale were willing to violate the canons of the World Church while practicing law, such a willingness to violate his own religious canons when doing so would promote his self-interest in obtaining a law license would raise serious concerns about his willingness to violate ethical canons of the legal profession when doing so would promote his racist agenda.

108. See Part IV regarding harm to the profession and accompanying footnotes.
covert manner in which pretextual rationales for racist conduct present themselves not as pretextual at all, but rather as plausible and rational reasons for conduct ultimately deemed race-neutral. Because racism can be perpetrated without detection, it poses an especially insidious threat to the cause of equal justice. Thus, the challenge for bar authorities is to determine who cannot reasonably be trusted to follow the dictates of equal justice if admitted to the bar. To determine whether a candidate for admission to the bar is likely to violate an ethical canon if admitted to the bar, bar authorities should look to the applicant’s past conduct. If the applicant’s conduct demonstrates firmly held beliefs that are incompatible with practicing law, exclusion from the bar is justified. In Hale’s case, the finding was easily made. Faced with the evidence of Hale’s active and overt racism, a conscientious bar examiner could not fail to conclude that Hale’s odious personal convictions were at war with the interests and obligations of the legal profession. Thus, the inquiry is not simply one of kind, but of degree.

Hale’s supporters overlook the gate-keeping function of the bar authorities. The primary responsibility of bar authorities is not to punish lawyers who engage in wrongdoing, but rather to safeguard the administration of justice and to protect the public from the misconduct of lawyers.109 The process of screening for moral fitness is a prophylactic measure to prevent injury to the public and the legal profession. “To achieve ‘[t]he greatest protection for the public, the courts, and potential clients,’ bar spokesmen have advocated ‘eliminating the diseased dogs before they inflict their first bite.’”110 Whether a white supremacist or an applicant who has embezzled funds from a former employer, the bar authorities cannot determine with certainty that the applicant will in fact violate a canon of professional ethics if admitted to the bar. Nonetheless, bar authorities use such information as an indicator of the applicant’s potential for harming the public and/or the profession in an effort to prevent future harm to our system of justice, which includes clients, the courts, and the image of the profession. The legal profession as a self-governing body has therefore determined that pre-admission conduct can be used to indicate whether an applicant would violate disciplinary rules at some point in the future if admitted to practice.111 If bar authorities abandoned these prophylactic measures, the profession could admit all applicants and theoretically disbar them later when and if they violate an ethical canon.

110. Rhode, supra note 25, at 509.
111. See generally McChrystal, supra note 22.
This approach is problematic for several reasons. First, once admitted, a lawyer lacking in moral character may commit many violations of the ethical rules before someone finally reports him to the disciplinary authorities. “The vast majority of attorney misconduct remains undetected, unreported, or unprosecuted, and bar disciplinary authorities have proved highly reluctant to withdraw individuals’ means of livelihood. Given the difficulties of ex post policing, entry restrictions appear to be a logical means of maximizing public protection.”112 Due to the stealthy nature of racism, a bar candidate who practices racism and is admitted to the profession is less likely to have such misconduct detected than a lawyer who steals client funds. Applicants are frequently denied admission to the bar if their past conduct indicates that they cannot be entrusted with client funds.113 Theft of client funds is a tangible, traceable, identifiable form of misconduct that is more easily identified and more easily proven than misconduct that is committed in the form of racism. Conduct that may appear racist to one person may appear benign to another. Moreover, even if conduct appears racist at first blush, it can often be explained away by some non-racist, well-reasoned rationale for the conduct. Hence, it is because misconduct in the form of racism or racial discrimination is too hard to detect and even harder to prove that the prophylactic measure of denying admission is required.

Because much misconduct goes unreported or undetected, many members of the public may be harmed before bar authorities are informed of the lawyer’s misconduct. By screening candidates for admission, we seek to avoid harm to the public and to preserve public confidence in the profession. Accordingly, bar authorities should screen out hardened racists to ensure confidence in the legal profession in the minds of people who question whether the legal profession is committed to treating whites and non-whites equally. Taking this paternalistic approach of trying to weed out the bad eggs helps assure the public that our system of justice is fair and not corrupt, thereby improving the image of the profession and the public’s faith in our system of justice.

112. See Rhode, supra note 25, at 509.
113. See, e.g., In re Mustafa, 631 A.2d 45 (D.C. 1993) (denying admission to a candidate for the bar of the District of Columbia because the candidate, while a student in law school, embezzled funds from his law school’s moot court program).
III. LAWYERS SHOULD NOT EXEMPT THEMSELVES FROM ANTI-DISCRIMINATION LAWS

Hale informed the Illinois Committee attempting to assess his character that he would probably never represent a black client.114 One should not be surprised that a white supremacist who feels no remorse or sense of loss for the death of several non-white people at the hands of one of his followers would not want to represent a black client.115 One may be surprised to learn that historically lawyers have had the autonomy to choose which clients to represent, and that this autonomy permitted the lawyer to deny representation on the basis of the client’s race.116 Those who argue against a prohibition on such race discrimination by lawyers argue that an anti-discrimination rule against lawyers would be illusory because it would force lawyers to lie about the reason for declining the representation. Indeed, Harvard Law Professor Martha Minow acknowledges that in the face of an anti-discrimination rule, lawyers are likely to offer pretextual reasons to cover up their practice of racial discrimination.117 One could argue that there is no problem with Matthew Hale refusing to represent black clients because there are other lawyers who will provide representation to such clients, and therefore the clients are not harmed by the denial of representation unless Matthew Hale is the last lawyer in town.118 Moreover, it could also be argued that blacks

115. See Hate.com: Extremists on the Internet (HBO television broadcast, Oct. 23, 2000) (on file with author) (featuring Hale’s statement that he feels no remorse when black people are murdered).
116. See Jennifer Tetenenbaum Miller, Free Exercise v. Legal Ethics: Can a Religious Lawyer Discriminate in Choosing Clients?, 13 GEO. J. LEGAL ETHICS 161, 165-66 (1999) (noting that traditionally lawyers have held an unencumbered right to decline representation, citing CHARLES W. WOLFRAM, MODERN LEGAL ETHICS § 10.2.2, at 573 (1986)). See also CANONS OF PROF’L ETHICS Canon 31 (1908), providing that:

[n]o lawyer is obliged to act either as adviser or advocate for every person who may wish to become his client. He has the right to decline employment. Every lawyer upon his own responsibility must decide what employment he will accept as counsel, what causes he will bring into Court for plaintiffs, what cases he will contest in Court for defendants.

117. Martha Minow, A Duty to Represent? Critical Reflection on Stropnicky v. Nathanson Foreword: Of Legal Ethics, Taxis, and Doing the Right Thing, 20 W. NEW ENG. L. REV. 5, 6 (1998). Professor Minow argues that a rule prohibiting a lawyer from discriminating in the selection of clients will force lawyers to lie and proffer pretextual reasons for denying the potential client representation. Id. While Professor Minow’s assumption is probably correct to some degree, the tendency of lawyers to lie in this situation should not serve as a roadblock to protecting the rights of the clientele of the legal profession. If Professor Minow’s reasoning were adopted, we would need to abolish almost all anti-discrimination laws, including the rule proffered in Batson, which is frequently circumvented through pretextual rationales which Professor Minow accurately labels as lies.

118. See Hodes, supra note 19, at 990 (outlining the “last lawyer in town” argument and arguing that lawyers have moral autonomy to choose which clients to represent).
should be grateful that a white supremacist lawyer would refuse to represent them given the harm that could be imposed upon the black client by the racist lawyer.119

While it is true that lawyers have historically enjoyed the privilege of determining which clients to represent, the public policy promoted by such lawyer autonomy is the promotion of zealous advocacy, arguably because a lawyer will not be as zealous in the representation of a client if the lawyer either disbelieves in or disapproves of the client or the client’s cause. Lawyer autonomy in client selection still exists, but it is not absolute. Limiting lawyer autonomy in client selection is not a new or novel concept. Lawyer autonomy in client selection is restricted already by the conflict rules which prohibit a lawyer from representing a client when such representation would create a conflict of interest between the potential client and an existing client or the lawyer’s own personal interests.120 Hence, the privilege of lawyer autonomy in client selection must be reconciled with both society’s and the profession’s commitment to equal justice for all, which translates into a commitment to eradicate racial discrimination in the administration of justice, which necessarily includes the selection of clients. Lawyers should consider only legitimate factors in declining representation such as the merits of the case, the client’s ability to pay the fee charged, and even whether the client’s cause is morally repugnant to the lawyer. By way of example, it is perfectly reasonable for a lawyer to decline representation because the lawyer believes that she will be unable to prove the potential client’s case. A lawyer who does not believe that the client’s case has merit would arguably not be a zealous advocate for the client. Likewise, lawyers are not expected to provide representation without adequate compensation unless the lawyer is acting to meet her pro bono publico obligations.121 Accordingly, it is reasonable to decline the representation of a client who is unable to pay for the services of the lawyer. Again, the argument is that a lawyer who is forced to represent a client without compensation might be less zealous in the representation.122

Finally, the ethics rules permit discrimination in client selection where the lawyer finds the client’s cause morally repugnant. Again, the rationale behind this rule is that the lawyer arguably would not be able to represent such a client with zeal. Arguably a lawyer who

120. See Model Rules of Prof’l Conduct Rs. 1.7, 1.8, 1.9 (2001).
121. See id. R. 6.1.
harbors racial hatred for a client would also be less likely to provide competent zealous representation for such client. Nonetheless, client disdain based solely on the immutable characteristic of the race of the client should not be a permissible basis for the denial of representation. Rather than argue that racist lawyers will not provide zealous advocacy, the legal profession should demand that lawyers be capable of providing zealous representation to all without regard to race. Those who cannot do so should not be in the profession. Thus, if we preclude the admission of racists to the bar in the first instance, the profession can uphold its commitment to equal justice for all and ensure zealous representation. In other words, if the profession only admits persons who are committed to the notion of equal justice for all, then the profession need not be troubled by a rule prohibiting discriminatory conduct in the client selection process or in any other aspect of the administration of justice. Hence, the exclusion of white supremacists is necessary to protect the integrity of the profession and to ensure compliance with ethical rules prohibiting discrimination by lawyers in the practice of law.\textsuperscript{123} Most importantly, though, the exclusion of white supremacists should be mandated in order to ensure the fair and orderly administration of justice.

IV. ADMISSION OF WHITE SUPREMACISTS WOULD BE PREJUDICIAL TO THE PUBLIC, THE LEGAL PROFESSION AND OUR SYSTEM OF JUSTICE

A. Racism Obstructs the Administration of Justice

Politically or religiously inspired conduct that is prejudicial to the administration of justice can serve as the basis for denying admission

\textsuperscript{123} Some may find this argument reminiscent of the arguments made in favor of excluding women from the bar in the case of \textit{Bradwell v. Illinois}, in which the state argued that women lacked the requisite moral character to be lawyers because: (1) they were frail; (2) the legislature did not contemplate their admission; and (3) their clients would not be able to sue them to enforce their contracts because the law at that time would not enforce a contract against a married woman. 83 U.S. 130, 139 (1872) (citing the \textit{Slaughter-House Cases}, 83 U.S. (16 Wall.) 36 (1872), which found that the Fourteenth Amendment applied only to State discrimination against “negroes”). The Court held that excluding women from the bar did not violate the privileges and immunities clause because the power to control: the granting of license to practice law in the courts of a State is one of those powers which are not transferred for its protection to the Federal government, and its exercise is in no manner governed or controlled by citizenship of the United States in the party seeking such license.

\textit{Id.} at 139 (citing \textit{Slaughter-House Cases}, 83 U.S. (16 Wall.) 36). Obviously, \textit{Bradwell} is no longer the law since the Fourteenth Amendment was subsequently interpreted to extend protection to women. The distinction between white supremacists and female candidates for bar admission should be obvious as well. Unlike women, white supremacists are not a protected class under the Fourteenth Amendment. See, e.g., 42 U.S.C. § 2000e-2 (2000) (prohibiting several types of discrimination on the basis of "race, color, religion, sex, or national origin"). Perhaps more importantly, white supremacists pose a more significant risk to the public and the administration of justice than the alleged risk posed by women as articulated in \textit{Bradwell}. 
to the bar. Denying admission to an individual whose stated “religious” mission requires the individual in part to transgress the limitations of the law by refusing to represent or hire blacks, for example, is not only reasonable, it is necessary to preserve the integrity of our entire system of justice. “Maintaining public respect for the laws and the courts is essential to the effective administration of justice.” Our profession plays a crucial role in our legal and political order. Thus, any conduct by a lawyer that decreases public confidence in our legal system should be condemned.

How could a lawyer who regards blacks as inherently inferior and undeserving of equal treatment under the law adequately represent a black person? Such preconceived notions about the worth of the client would “likely impair [the lawyer’s] ability to represent the client” and would seriously damage and erode the lawyer-client relationship. This argument assumes, of course, that a white supremacist lawyer would even accept a black client. The Model Rules specifically allow a lawyer to refuse to accept a court appointment in circumstances where “the client or the cause is so repugnant to the lawyer as to be likely to impair the client-lawyer relationship or the lawyer’s ability to represent the client.” Moreover, a lawyer has the right to control his or her own labor, meaning that a lawyer does not have to accept as a client, every person who seeks his services. Thus, a racist lawyer, which I will define in this instance as one whose beliefs mandate discriminatory behavior against non-whites, necessarily will be more likely to discriminate against a black in determining whether to accept that individual’s case. Since he has the intellectual acumen to become a lawyer, he probably will be astute enough to offer a pretextual reason for declining to accept a black client. For example, Hale could deny a black person representation with the explanation that he thinks the case lacked merit or that his current caseload or other commitments would not permit the representation. These rationales are race-neutral on their face and reasonable. In fact, Hale’s real reason for denying the representation may be the race of the potential client. Rather than allow this type of

124. See McChrystal, supra note 22, at 74.
125. See THE SALVATION, supra note 96.
127. See Wilkins, supra note 21, at 1512.
128. See Kiovsky, supra note 126, at 201.
130. Matthew Hale admitted that he would probably never represent a black client.
subtle and often unprovable discriminatory behavior, bar examiners
are correct to not empower white supremacists with a license to prac-
tice law.

While some jurisdictions still do not have ethical rules precluding
lawyers from refusing to represent an individual based on the race of
the would-be client, this author submits that it is morally and ethi-
cally improper for a lawyer to refuse to accept a matter if that refusal
is based even partially upon the race of the potential client. Indeed,
even in the absence of an ethical rule precluding discrimination in
the acceptance of clients, such conduct may be illegal under federal,
state, and/or local public accommodation laws, which guarantees to blacks the same right to make
equipments as to whites. Thus a lawyer who refuses to
enter into a lawyer-client relationship, which is a contractual rela-
tionship, solely on the basis of the potential client’s race would be
violating Section 1981. This “status-based discrimination” under-
mines the legitimate social purposes underlying the rules regard-
ing client selection. The law governing lawyers has historically per-
mits a lawyer to choose which clients and causes the lawyer will
represent. The reason that the law has granted a lawyer this type of
autonomy is so that the lawyer’s independent professional judgment
will not be compromised by the lawyer’s personal beliefs and/or mor-
als that may conflict with the client or the client’s legal position.
The classic example of such a conflict is the lawyer who is asked to
represent a person charged with a sex offense against a child. Many

133. See Wilkins, supra note 21, at 1575-76 (citing Stropnick v. Nathanson, 19 Mass.
     1999 WL 33453078 (Mass. Comm'n Against Discrimination July 26, 1999) (finding that a
     lawyer’s office is a public accommodation under the terms of the relevant Massachusetts
     anti-discrimination statute). For a discussion of whether lawyers should be allowed to de-
     cline representation based on the race of the potential client, see Issue One of Volume 20 of
     the Western New England Law Review (1998), which has several essays discussing lawyer
     autonomy in client selection.
134. 42 U.S.C.A. § 1981(a), (c) (2000) provides in relevant part:
     All persons within the jurisdiction of the United States shall have the same
     right in every State and Territory to make and enforce contracts, to sue, be par-
     ties, give evidence, and to the full and equal benefit of all laws and proceedings
     for the security of persons and property as is enjoyed by white citizens, and
     shall be subject to like punishment, pains, penalties, taxes, licenses, and exac-
     tions of every kind, and to no other . . . . The rights protected by this section are
     protected against impairment by nongovernmental discrimination and impair-
     ment under color of State law.

135. Id.

     United States Code prohibits private schools from excluding qualified children solely be-
     cause they are black).

137. See Wilkins, supra note 21.

138. See Hodes, supra note 19, at 990.
lawyers would refuse this representation because their own beliefs and morals are so offended by the alleged offense that they would be unable to zealously represent such a client.

Lawyers as administrators of justice stand as an instrumentality of government. The judicial process itself is viewed as government. Lawyers have a government-granted monopoly on the administration of justice in this country which gives them power to affect the lives of others. Racially discriminatory or racially charged behavior by lawyers obstructs the fair administration of justice.\(^\text{139}\) Granting a license to practice law to an individual who practices racial hatred and discrimination effectively gives that person more ammunition to add to his arsenal to support his personal race war because a lawyer has more powers and privileges than he had as a lay person.\(^\text{140}\) Thus, whether a law office constitutes a public accommodation or not, the profession’s commitment to “equal access under law” and “justice for all” is undermined if an individual lawyer is permitted to refuse to represent individuals on the basis of race or other “considerations that have nothing to do with either their moral worth as human beings or the legitimate interests of attorneys,” such as the merits of the case.\(^\text{141}\)

There is precedent for excluding white supremacists from a profession or occupation. The military has found that active participation in hate groups such as the KKK is incompatible with military service, which requires trust, cohesiveness, and discipline among service members.\(^\text{142}\) While the military is easily distinguished from the legal profession due to its status as a separate community,\(^\text{143}\) the

\(^{139}\) See Brenda Jones Quick, Regulating a Lawyer’s Discriminatory Conduct: Constitutional Limitations, 21 OHIO N.U. L. REV. 897, 911 (1995) (citing In re Vincenti, 554 A.2d 470 (N.J. 1989), in which the court held that a lawyer who uttered racial slurs against opposing counsel had undermined the administration of justice). Quick also cites Gonzalez v. Commission on Judicial Performance, 657 P.2d 372 (Cal. 1983), which held a judge guilty of misconduct because he had made racially discriminatory remarks to a juror and a defendant. Quick, supra, at 911.

\(^{140}\) See L. Ray Patterson, The Fundamentals of Professionalism, 45 S.C. L. REV. 707, 716, 719-20 (1994) (noting the power that lawyers have to control the affairs of others and arguing that this power is derived both from the client and the state).

\(^{141}\) See Wilkins, supra note 21, at 1581.

\(^{142}\) Military Forbids Active Role of Soldiers in Hate Groups, N.Y. TIMES, Sept. 12, 1986, at A25. The military also has limited other activities protected by the First Amendment in an effort to ensure that the military is ready, willing and able to perform its vital role in protecting national security. For example, the military found desecration of the United States flag by a military service member to constitute the military offense of dereliction of duty despite the First Amendment because the military has a compelling government interest in ensuring a respect for duty and discipline among its members. See United States v. Wilson, 33 M.J. 797 (A.C.M.R. 1991).

\(^{143}\) See Douglas Daniels, Freedom of Hate and Service in the United States Coast Guard: Rights vs. Duty, 9 U.S. A.F. ACAD. J. LEG. STUD. 147 (1998/1999) (describing the distinction between civil and military law based on the view of the military as a separate community).
legal profession, not unlike the military, has a compelling interest in ensuring that its lawyers adhere to the values of the community, one of which is equal justice. Public employers other than the military have attempted with mixed success to deny employment based on membership in a hate group. Thus, police forces that have sought to exclude members of hate groups from their ranks have only been successful when the hate groups in question had been involved in the commission of violent crimes.

B. The Parade of Horribles

A license to practice law is broad and unrestricted. Once admitted to the bar, an individual can pursue many areas of practice, including public service. Indeed, if admitted to the bar, Hale or any other white supremacist could seek a position as a criminal prosecutor. Criminal prosecutors have broad discretion, ranging from the powers to determine who to prosecute and what charges to bring against the accused to the powers to offer a plea bargain or downgrade charges. Already there are questions of whether white prosecutors who claim to be objective are subconsciously abusing this discretionary power in implementing the criminal justice process.

Because prosecutors have such broad discretion and power over the lives of people who become a part of our criminal justice system, a white supremacist cast in such a role would have the power to implement his racist agenda by race-based exercises of discretion in the charging function, plea bargaining process, and in sentencing recommendation. For example, a white supremacist prosecutor could choose to prosecute all blacks charged with a crime to the fullest de-

144. Id. Daniels acknowledges that police and firefighters restrict active membership in hate groups such as the KKK in an effort to avoid biased treatment of those who depend on the crucial services provided by these professions. Id. at 157 (citing McMullen v. Carson, 568 F. Supp. 937 (M.D. Fla. 1983), in which dismissal of police officer based on officer’s active membership in the KKK was upheld). Daniels defines biased treatment by police officers as including unlawful apprehension and/or assault on the basis of race and points to the Rodney King tragedy as evidence of the harm that racist police officers are able to inflict. Id. Daniels also points to an incident wherein a firefighter was charged with intentionally allowing two black children to die in a fire to demonstrate the nature and severity of harm that racists in the ranks of the fire department may cause. Id. at 157.


146. The disproportionate number of black men who are prosecuted, convicted, and sent to jail in this country, along with racist guidelines and practices uncovered in some prosecutors’ offices and contextualized by our racist history, have caused many lawyers and scholars to opine that our criminal justice system is racially biased. See Anthony Alfieri, Prosecuting Race, 48 DUKE L.J. 1157 (1999); Angela Davis, Prosecution and Race: The Power and Privilege of Discretion, 67 FORDHAM L. REV. 13 (1998); Sheri Johnson, Unconscious Racism and Criminal Law, 73 CORNELL L. REV. 1016 (1988); Richard McAdams, Race and Selective Prosecution: Discovering the Pitfalls of Armstrong, 73 CHI.-KENT L. REV. 605 (1998).
gree possible under the law despite the circumstances of the case and could look for mitigating factors to support a decision to charge white defendants with lesser charges or not at all. The white supremacist prosecutor could also refuse to plea bargain with black defendants. On the surface, this refusal could be supported under the law through an easily proffered non-race-related rationale, but underneath that determination lies the racist belief that blacks are evil and undeserving of leniency, a second chance, or even a fair process. Such abuse would be difficult if not impossible to prove because the law gives the prosecutor this untamed discretion. Not only is the prosecutor’s subjective motive virtually impossible to prove, the standard of proof that a minority seeking to challenge an official’s implementation of discretion must meet is “exceptionally clear proof.” In this circumstance, the black criminal defendant would have no power or legal basis upon which to reject the system’s use of the white supremacist as a minister of justice.

It is reasonably foreseeable that Hale or any other white supremacist would be more inclined than the average lawyer to deny black jurors the privilege of serving on a jury, especially in cases involving black criminal defendants charged with victimizing whites. After all, the precepts of Hale’s religion mandate that he perform in a manner consistent with what he perceives as best for the white race. Admittedly, there exists positive law that prohibits a lawyer’s use of peremptory challenges to exclude jurors based on race. Nonetheless, because discrimination is an intent offense that occurs within the confines of the perpetrator’s mind, it is relatively easy to camouflage the offense as permissible action. Lawyers can frequently offer a pretextual reason for excluding a juror that, on its face, is a rational, race-neutral basis for excluding the juror.

Once admitted to the bar, Hale might even be elevated to the

---

147. McCleskey v. Kemp, 481 U.S. 279, 297 (1987) (rejecting the Baldus study’s evidence of the disparate impact that the death penalty has upon blacks).


149. See Tracy M.Y. Choy, Branding Neutral Explanations Pretextual Under Batson v. Kentucky: An Examination of the Role of the Trial Judge in Jury Selection, 48 Hastings L.J. 577 (1997) (discussing the problem of identifying pretextual reasons for discrimination in jury selection). Examples of explanations that have been accepted by the courts as non-pretextual reasons for striking minority jurors include too much eye contact, too little eye contact, living in public housing, poor attitude, and poor posture and demeanor. See, e.g., United States v. Tindle, 860 F.2d 125, 129 (4th Cir. 1988), cert. denied, 490 U.S. 1114 (1989); United States v. Mathews, 803 F.2d 325, 331-32 (7th Cir. 1986), rev’d on other grounds, 485 U.S. 58 (1988); United States v. Forbes, 816 F.2d 1006, 1010-11 (5th Cir. 1987); United States v. Vaccaro, 816 F.2d 443, 457 (9th Cir. 1987); United States v. Caridgide, 808 F.2d 1064, 1070 (5th Cir. 1987); see also Purkett v. Elem, 514 U.S. 765 (1995); Bill, supra note 46, at 552 n.20; Douglas L. Colbert, Challenging the Challenge: Thirteenth Amendment as a Prohibition Against the Racial Use of Peremptory Challenges, 76 Cornell L. Rev. 97-98 (1990).
If Hale were a judge, affiliation with the World Church would certainly be problematic because of the restraints on affiliation placed on judges by the Model Code of Judicial Conduct, which has been adopted in many jurisdictions. The Code of Judicial Conduct expressly provides that a “judge shall not hold membership in any organization that practices invidious discrimination on the basis of race, sex, religion or national origin.” Arguably this rule restricts a judge’s freedom of expression and/or association under the First Amendment in the same manner as Hale’s denial of admission to the bar. The lawyer or judge must choose her master; she cannot serve both. The commentary to this canon explains the rationale behind the rule. Judges in our system of justice are supposed to be impartial arbiters, and one of the reasons that we give deference to the decisions of judges is because we perceive them as being fair and impartial.

If a judge is also a Klansman or affiliated with any white supremacist group, our perception of that judge’s impartiality is impaired because we quite logically think that he is unable to be impartial when presiding over matters presented by non-white lawyers in-

---

150. The argument that Hale’s racist views would act as a barrier to an appointment or election to the bench is not persuasive in light of the political success of lawyers for whom a history of racist politics has not been a hindrance to judicial appointment. The case of Charles Pickering is instructive. Judge Pickering, who was recently nominated for appointment to the Court of Appeals for the Fifth Circuit from his current district court judgeship, wrote a law review article advocating an amendment to Mississippi’s anti-miscegenation law to strengthen the law. See Charles W. Pickering, Criminal Law, Miscegenation, Incest, 30 Miss. L.J. 326 (1959). Many opposed Judge Pickering’s appointment due to his allegedly racist past. See Our Opinions: Extremist judge is unfit to sit on appeals court, ATLANTA J.-CONST., Feb. 7, 2002, at A17 (noting that Pickering has offered no apologies for a law review article he wrote in 1959 suggesting ways to strengthen Mississippi’s anti-miscegenation law and mentioning his support for a segregationist organization in the 1970s). But see David Firestone, Blacks at Home Support a Judge Liberals Assail, Feb. 17, 2002, at http://query.nytimes.com/gst/abstract.html?res=F90E10F83A5B0C748000AB0894DA404482 (last visited Mar. 16, 2003) (on file with author) (reporting that some blacks in Laurel, Mississippi, the Judge’s hometown, do not see him as racist). Despite the fact that Judge Pickering’s nomination to the Fifth Circuit was defeated, he still holds a position on the bench, and Pickering has since been re-nominated to the Fifth Circuit by President George Bush. See Ditching over Pickering, N.Y. POST, Jan. 19, 2003, at 26. This author makes no judgment about whether Judge Pickering is a racist, but merely acknowledges that serious questions exist regarding his racial views and actions.

151. MODEL CODE OF JUDICIAL CONDUCT Canon 2(C) (2001).

152. See id. Canon 2, cmt; see also Leslie W. Abramson, Canon 2 of the Code of Judicial Conduct, 79 MARQ. L. REV. 949 (1996) (outlining the law interpreting Canon 2 and acknowledging that the Canon requires judges to be fair and impartial).

153. See MODEL RULES OF PROF'L CONDUCT pmbl. (2001) (as adopted by the ABA House of Delegates at the February 2002 Meeting and stating in relevant part that “legal institutions in a constitutional democracy depend on popular participation and support to maintain their authority”); see also id. R.6; MODEL CODE OF JUDICIAL CONDUCT Canon 1, cmt. (2001) (“Deference to the judgments and rulings of courts depends upon public confidence in the integrity and independence of judges.”).
volving non-white litigants or involving issues of race. Since all white supremacist groups require members to be white, such groups without question practice discrimination. One might ask how a judge’s passive membership in a white supremacist group affects the judicial system if that judge swears that he is ready, willing, and able to put his personal beliefs aside for the purpose of fulfilling his duties as judge. The judge’s passive membership in a white supremacist organization expresses to the public that the judge approves of invidious discrimination against non-white persons. Thus, the judge’s affiliation with such an organization manifests bias on behalf of the judge, thereby creating the “appearance of impropriety” and diminishing “public confidence in the integrity and impartiality of the judiciary.”

The Model Judicial Code also mandates that judges perform their judicial duties without bias or prejudice. The commentary to the canon explains that a “judge who manifests bias on any basis in a proceeding impairs the fairness of the proceeding and brings the judiciary into disrepute.” Hence, a commitment to non-bias on behalf of those administering justice is essential to preserving the integrity of our judicial process.

Although lawyers, as officers of the court, do not have to meet the standard of impartiality that judges do it would seem that they too should not appear unfair or tainted by discriminatory beliefs and/or affiliations. Lawyers tainted by such discriminatory views would also give the appearance of impropriety in our system of justice. As officers of the court, lawyers have been entrusted with certain unique powers. Even if we assume that a Klansman can set his white supremacy beliefs aside for purposes of practicing law, do we not still have a perception problem from the perspective of the public, which has delegated the duty of administering justice to lawyers? Awarding a license to practice law to a Klansman would diminish the public’s

154. See KENNEDY, supra note 64, at 61 (citing United States v. Brown, 539 F.2d 467, 468 (5th Cir. 1976), wherein a judge who used a racial epithet was found to have called into question his own ability to be fair and impartial). The question of whether an organization practices invidious discrimination is answered by looking at such factors as “how the organization selects its members,” meaning whether the organization arbitrarily excludes people based on race, religion, sex, or national origin. MODEL CODE OF JUDICIAL CONDUCT Canon 2(C), cmt. 1 (1990).

155. Id. Canon 2(C), cmt. 2.

156. MODEL CODE OF JUDICIAL CONDUCT, Canon 3(B)(5) (2000), requires judges to perform their duties “without bias or prejudice.” In pertinent part, Canon 3(B)(5) states:
A judge shall perform judicial duties without bias or prejudice. A judge shall not, in the performance of judicial duties, by words or conduct manifest bias or prejudice, including but not limited to bias or prejudice based upon race . . . and shall not permit staff, court officials and others subject to the judge’s direction and control to do so.

Id.

157. Id.
confidence in the integrity and dignity of our profession, especially the confidence of those whom white supremacists deem inferior.

Some may wonder why Hale is distinguishable from the rest of us who may hold “unpopular” views and bias yet still have the privilege of sitting at the bar. Hale’s views are not merely “unpopular,” many of them have been excluded from the basis of our social and political organization by the Constitution. While I will concede that we all bring our own personal bias and idiosyncracies to the bar, none of us should be committed to promoting racial discrimination or racial violence, both of which are promoted by white supremacists. A white supremacist, unlike the rest of us, has at the core of his very being a burning passion to inflict harm\(^\text{158}\) upon non-white people. Unlike the white supremacist, we, as members of the bar, are committed to the core value of equality and justice for all, and we strive to implement those notions in our practice despite the fact that we may from time to time fall short of our efforts. It is our desire and our effort to achieve equality and justice for all that distinguishes the rest of us from the Matthew Hales of society.

C. Admitting Racists to the Bar Undermines Equal Protection For All

The supreme law of our land, the Constitution of the United States, ensures justice for all and equal protection under the law for all citizens regardless of their race, creed, or national origin.\(^\text{159}\) Measures designed to ensure equal protection in the enforcement of the law by the police are necessary in light of the genre of narrative that locates and contextualizes our equal protection jurisprudence.\(^\text{160}\) Likewise, measures designed to ensure equal protection in the administration of justice by lawyers is necessary in light of our history. Moreover, such measures are justified by the compelling governmental interests of ensuring safety and equal protection of its citizenry under the law, and are therefore consistent with the guarantees of the First Amendment.\(^\text{161}\)

---

\(^\text{158}\) Harm is not limited to physical violence, but also includes the damaging economic effects of discriminatory practices in employment, housing, and other spheres of human activity.

\(^\text{159}\) U.S. CONST. amend. XIV.

\(^\text{160}\) See Sean Hecker, \textit{Race and Pretextual Traffic Stops: An Expanded Role For Civilian Review Board}, 28 COLUM. HUM. RTS. L. REV. 551 (1997) (chronicling the problem of racial profiling by the police and arguing that the Civilian Review Board could be an effective tool for preventing and remedying racial profiling).

\(^\text{161}\) See Robin D. Barnes, \textit{Blue by Day and White by (K)Night: Regulating the Political Affiliations of Law Enforcement and Military Personnel}, 81 IOWA L. REV. 1079 (1996) (arguing that hate group restrictions for police and military members are necessary and legal under the First Amendment).
Granting a law license to a white supremacist would violate the spirit if not the letter of the Fourteenth Amendment, which mandates equal protection under the law. Once admitted to the bar, a lawyer becomes an officer of the court in the state in which he or she is licensed. In some jurisdictions, lawyer oaths acknowledge this status of lawyers by referring to the lawyer as an officer of the state or as holding an office within the state.162 Moreover, the Proposed Model Rules of Professional Conduct, as proposed by the Ethics 2000 Commission established by the American Bar Association, has acknowledged the unique status of lawyers by defining a lawyer as an “officer of the legal system and a public citizen.”163 It has been held in the context of liquor licensing that the state’s issuance of a license does not constitute “state action” as that term has been defined in Fourteenth Amendment jurisprudence.164 Nonetheless, the state’s issuance of a license to practice law is distinguishable from the issuance of a liquor license. The licensing of a lawyer has significantly graver consequences than licensing an establishment to serve alcohol. The licensing of a lawyer makes that person an officer of the state and arguably a state actor because that person now has the power to administer justice, which is a function of the state. Conversely, selling liquor is not a function of the state.

Even if a lawyer is not in all respects a state actor, a lawyer has the power to transform herself into a state actor. For example, a lawyer who chooses to work for the government, such as a prosecutor, is clearly a state actor for purposes of the Fourteenth Amendment. Even a lawyer in private practice may be a state actor when that lawyer exercises state-granted powers.165 Thus, by issuing a license to practice law, a state is creating a state actor. Accordingly, the act of creating a state actor must itself be state action. Even if a lawyer is not a state actor, the act of licensing a person to practice law may

162. See, e.g., N.J. STAT. ANN. § 2A:13-1 (West 2001) (requiring attorneys to “solemnly promise and swear” that they will “perform the duties of [their] office faithfully”)


164. See Moose Lodge No. 107 v. Irvis, 407 U.S. 163 (1972) (holding that the state’s issuance of a liquor license to a private club did not constitute state action under the Fourteenth Amendment); see also Jackson v. Metro. Edison Co., 419 U.S. 345 (1974) (holding that a utility company’s termination of service to a household did not constitute state action despite the fact that the utility company was engaged in a business affecting public interest, was subject to extensive state regulation, and had a partial monopoly).

165. See Edmonson v. Leesville Concrete, 500 U.S. 614, 623-25 (1991) (holding that a private lawyer’s exercise of a peremptory challenge in a civil trial constitutes state action because peremptory challenges derive from state authority and because the private lawyer made extensive use of those government procedures with “overt, significant participation of the government”).
still be state action because the states are delegating governmental power to lawyers—the power to administer justice. States should not be permitted to delegate this power to someone who is a known racist because to do so would serve to encourage racial discrimination in the administration of justice, which states are arguably prohibited from doing under the Fourteenth Amendment.

Even if licensing a lawyer does not constitute state action and therefore does not violate the Fourteenth Amendment, such action by the state does convey a message to the public—particularly the non-white public—that the legal profession tolerates racial bigotry, the Thirteenth and Fourteenth Amendments notwithstanding. Moreover, admission of a white supremacist would serve as the state’s concession that racism is either a benign component of our legal system that we all must tolerate, or a benign component of personal character not worthy of concern or challenge. Even worse, admission of an open and notorious white supremacist by the state might even communicate the message that racial inferiority is not a false or evil ideology, but rather one that is open to debate. The admission of openly racist applicants to the bar would contribute to a racist culture that promotes the subordination and marginalization of the minority and heightens the risk that minorities will not receive justice and fairness from our judicial process. If the law permits a white supremacist to become a lawyer, then the law necessarily contributes in a significant way to perpetuating a racist culture. Moreover, admission of a white supremacist to the bar vividly displays that the profession is not committed to ridding itself of the patterns and practices that perpetuate the problem of institutional racism.

Once empowered with a law license, a white supremacist has more power to effectuate his racist agenda than he would have as a layperson. Thus, by granting a license to a white supremacist such as Hale, the state would be indirectly supporting Hale’s racist agenda and, through Hale, would be endangering the equal protection rights of all non-whites who are subject to our system of justice—or at least those non-whites who would encounter Hale in his capacity as a lawyer. The failure to protect the minority members of the public from the racist agenda of Hale or any other white supremacist lawyer would elevate the rights and liberty interests of white racists over

---

166. Many free speech absolutists argue that one can never be sure that an opinion is a false opinion. See BOLLINGER, supra note 59, at 260 n.24.
167. Andrew E. Taslitz, Condemning the Racist Personality: Why the Critics of Hate Crimes Legislation Are Wrong, 40 B.C. L. REV. 739, 742 (1999) (arguing that hate crimes contribute to a racist culture and create a heightened risk of physical harm to racial minority groups).
the rights and liberty interests of not only non-whites, but also non-racist whites. Absent a law license, one can dismiss Hale as a marginal individual unlikely to effectuate much harm if he operates within the confines of the law, and he will likely be apprehended and punished if he crosses the legal line. But when the state grants a white supremacist such as Hale a license to practice law, the white supremacist agenda is legitimized if not endorsed by the entire system of justice. The state essentially would be saying that lawyers do not have to believe in or uphold a fundamental principle of our system of justice, which is that all persons are entitled to equal justice.

V. WHY RACISM?

Racism is intimately woven into our traditional social and political ordering. Much governmental and social effort in the twentieth century has been directed toward eliminating racism from institutional organizations of government. Admitting white supremacists to the bar is antithetical to the efforts of the last century.

Some may ask why we should be concerned with racism and how we can bar racism when the courts of the McCarthy era seemed to say that bar authorities cannot exclude candidates for admission based on their beliefs. The loyalty oath cases of the McCarthy era should not be read so broadly. When synthesized, the cases merely hold that the First Amendment prohibits a state from inquiring into the associations of bar applicants without demonstrating that a legitimate state interest is served by the intrusion. In those cases, the applicants refused to reveal their beliefs and bar authorities denied admission based on their refusal to answer questions of associa-

169. Of course, the political belief that was of concern during the 1950s and 1960s was the belief in communism. See generally Konigsberg v. State Bar of Calif., 366 U.S. 36 (1961). For a thorough history of communism and the First Amendment, see Marc Rohr, Communists and the First Amendment: The Shaping of Freedom of Advocacy in the Cold War Era, 28 SAN DIEGO L. REV. 1 (1991).

170. In re Stolar, 401 U.S. 23, 28-29 (1971) (holding that “the First Amendment prohibits Ohio from penalizing an applicant by denying him admission to the Bar solely because of his membership in an organization” or because “he personally . . . ‘espouses illegal aims’”). The Court concluded that the question on the application seeking to determine whether the applicant was a member of an organization that advocates overthrow of the United States government by force violated the First Amendment because “no legitimate state interest . . . is served by a question which sweeps so broadly into areas of belief and association.” Id. at 30. Likewise, in Baird v. State Bar of Arizona, 401 U.S. 1 (1971), the Court held that the state failed to show that asking an applicant to identify membership in an organization was necessary to protect a legitimate state interest. Hence, the finding that the applicant was barred from admission was reversed. In Law Students Research Council v. Wadmond, 401 U.S. 154 (1971), New York had a rule requiring an applicant to believe in and be loyal to the government of the United States. The Court held that a state could deny admission to the bar if an applicant’s advocacy of government overthrow or membership in an organization advocating forceful overthrow was coupled with the “specific intent” to achieve that illegal goal. Id.
During this period in history, states argued that they had legitimate governmental interests in knowing whether an applicant is or was a member of any organization which advocates the overthrow of the United States government by force or whether the applicant personally espoused such illegal aims. The Court rejected this claim and found that whether an applicant is a communist or advocates overthrow of the United States government is irrelevant to his fitness to practice law. The rules of ethics for lawyers do not proscribe communist beliefs or practices. Unlike communism or other political beliefs or ideologies, whether an applicant is a devout racist is relevant to his fitness to practice law for all of the reasons previously discussed in this Article. Unlike the racist, the communist does not seek to subvert the rights of a particular class of citizens identified solely on the irrational basis of an immutable characteristic. Also, unlike communism, racism is, in many contexts, proscribed by the law.

Moreover, unlike anti-communism or other political beliefs, the premise of equality is at the core of the legal profession and the system of justice which the legal profession is entrusted to administer. The legal profession is charged with the duty of administering justice for all people without regard to race or color. Justice for people of color is only achieved in a system that is not entrusted to and administered by racists because racism obstructs the fair administration of justice by bringing the non-relevant factor of race into the process. Hence, a commitment to racial equality should be a prerequisite for admission to the profession. Those who fail to meet the equal justice threshold necessarily lack the qualification to become a lawyer. Finally, unlike communism, racism is a tool that has been used by our government in the administration of justice to systematically deny justice to a group of people, identified only by their race, in an effort to subordinate those people. In light of this history and the pervasiveness of racism in present-day law and society, government should take affirmative steps to eradicate racism from the administration of justice.

As for the First Amendment concern, the First Amendment does not prohibit a state from taking steps to protect its citizens from harm. For those who are wed to First Amendment precedent, the

171. In the case of Matthew Hale, who is an open racist, there was no need to inquire about his association with a racist group or his willingness to discriminate against people based on their race because he openly advocated such practices and is the public leader of a white supremacist group.
174. For an intelligent discourse on how race continues to permeate the law and social interactions between people, see K. ANTHONY APPIAH & AMY GUTMANN, COLOR CONSCIOUS: THE POLITICAL MORALITY OF RACE (1996).
First Amendment does not apply to conduct unless it is expressive conduct such as burning the flag to make a political statement. But expressive conduct can be prohibited when the prohibition is not related to expression. Moreover, expressive conduct can be prohibited even when the state’s regulation of such conduct is related to expression, as long as the prohibition is necessary to protect a compelling governmental interest. The exclusion of a white supremacist from the bar is not based solely on his beliefs. In the Hale case, there was conduct in the record that supported a finding that Hale did not merely hold racist beliefs, but engaged in racist and discriminatory activities. This author posits that a person who harbors racist beliefs does not do so in the abstract. Racism is a unique narrative that is not easily defined due to its multifaceted conformation. It is not merely a belief held in the abstract; it is, by its very nature, both belief and conduct inseparably merged like a scrambled egg.

Racism informs and influences everyday decision-making and interaction with others. It is not merely thought; it is a practice. It is practiced in ways too numerous to record here. Thus, excluding a racist from the bar is not an exclusion based solely on his or her beliefs, but an exclusion based on the reasonably anticipated conduct that the racist will engage in if admitted to the bar. Accordingly, unlike communism, racism is relevant to an applicant’s fitness to practice law. Knowing the true nature of racism as both a practice and a belief, it is not acceptable to confine the white supremacist’s racism to “his beliefs” and protect him under the First Amendment. It is reasonable for bar authorities to conclude that a white supremacist whose “religion” mandates conduct that violates the anti-discrimination laws of our country, including the law of lawyering, would likely violate those laws while practicing law.

The question arises as to whether the white supremacist’s racist conduct is expressive conduct, because if it is, it is entitled to a

175. See, e.g., Texas v. Johnson, 491 U.S. 397 (1989) (holding that although the state has a legitimate interest in encouraging appropriate treatment of the flag, it may not promote its view of the flag by prohibiting expressive conduct relating to it by criminally punishing a person who burned the flag as a means of political protest). The Court pointed out, though, that not all action taken with respect to our flag would constitute expressive conduct. Id.


178. Lawrence, supra note 61, at 443-44 (arguing that the inseparability of the idea and practice of racism was central to the decision in Brown, which held that segregation is unconstitutional). Lawrence further argues that we do not see most racist conduct because much of it is considered unrelated to race. Id.

heightened level of scrutiny under the First Amendment. The racist conduct of a white supremacist is not expressive conduct. The white supremacist who adopts a way of life devoted to violating antidiscrimination laws is not likely doing so to make a political statement, but rather to promote his or her own self-interest in maintaining white supremacy. But even if we accept that the white supremacist is being racist for the purpose of making a political statement, the next issue that must be determined is whether a state regulation banning white supremacists from the bar relates to the suppression of free expression. Banning white supremacists from the legal profession does not relate to the suppression of free expression, but rather it relates to protecting the rights of our citizens, particularly our minority citizens who historically have been denied rights due to their race.

Arguably, a ban on white supremacists is related to the suppression of free expression because to the extent one wants to be a lawyer, one would not be permitted to engage in expressive racist conduct. Nonetheless, even if a ban on white supremacists does suppress free expression, it is not unconstitutional under the First Amendment because the constitutional interest that the state is seeking to promote—equality—outweighs the racists’ interest in being able to practice racism without consequence. Unlike communism or other political ideology, racism has a special “juridical category”\textsuperscript{180} created for it within the text of our Constitution. Hence, unlike anti-communist ideology, the principle of equality or anti-racist ideology is embedded in the very foundation of our law. “The First Amendment does not guarantee that other concepts virtually sacred to our Nation as a whole—such as the principle that discrimination on the basis of race is odious and destructive—will go unquestioned in the marketplace of ideas.”\textsuperscript{181} Nonetheless, the First Amendment does not prohibit the state from taking action that imposes an incidental limitation on speech when such action is aimed at eradicating racial discrimination.\textsuperscript{182}

\textsuperscript{180} Texas v. Johnson, 491 U.S. at 417 (discussing the fact that a Texas law banning flag burning violated the First Amendment when burning the flag was expressive conduct because there is no indication in the Constitution or other law that a separate “juridical category exists for the American flag” so as to permit the state to restrict First Amendment rights with respect to the flag).

\textsuperscript{181} Id. at 418.

\textsuperscript{182} Bob Jones Univ. v. United States, 461 U.S. 574 (1983).
VI. THE TOUGH CASES

A. The Reformed Racist

What do we do with the applicant who is an ex-Klansman? Most of us are continually evolving in our viewpoints and the way in which we conduct ourselves. As such, it is imprudent to apply the “once a racist, always a racist” standard to an applicant for bar admission. Once it is determined that the applicant participated in discriminatory or racist conduct in the past, bar authorities should conduct the same analysis they would for any other past conduct reflecting adversely on a candidate’s moral character and fitness to be a lawyer.

Bar authorities should consider how long ago the racist conduct occurred, and whether the candidate has had a clear record of non-discriminatory conduct for a significant period of time so as to demonstrate reformation or rehabilitation. For example, a bar candidate who was a member of a racist organization less than two years before applying to the bar should not be deemed to have demonstrated reformation because only an insubstantial amount of time has passed since the misconduct. A bar candidate should not be denied admission for conduct that occurred twenty years before applying to the bar when she was an impressionable youth. Similarly, a bar candidate should not be denied admission for past racist conduct when the candidate has conducted herself in a fashion demonstrating fitness to practice law over a long period of time. Finally, bar authorities should also consider whether the candidate has accepted responsibility for the conduct as well as whether the candidate expresses remorse for the conduct. Again, it is important to remember that the applicant bears the burden of proving that she is of good moral character, so if the applicant fails to meet this burden of proof, denial of admission to the bar is appropriate.

B. The Open Racist vs. The Closet Racist

White supremacists are extremists marginalized by American society at large. Their viewpoint and political propaganda are “politi-

183. See In re Prager, 661 N.E.2d 84, 89 (Mass. 1996) (acknowledging that a bar candidate who has committed an act reflecting adversely on her fitness to practice law may be sufficiently rehabilitated to be admitted to the bar after the passage of time during which the candidate’s actions demonstrate good moral character).
184. Id.
185. In re Mustafa, 631 A.2d 45 (D.C. 1993) (denying bar admission to a candidate who had embezzled and repaid funds while in law school since only one year had passed since the misconduct).
186. In re Zbiegien, 433 N.W.2d 871 (Minn. 1988).
cally incorrect” and abhorred by most members of the bar. And while there are those who think that white supremacists should not be excluded from the bar based on their status as white supremacists, the decision of the Justices of the Supreme Court of Illinois in In re Hale suggests that those justices and many lawyers, especially lawyers of color, tend to agree that the exclusion of white supremacists from the bar is beyond dispute.

The more challenging question arguably is what degree of racism by a member of the bar is permissible. Do persons who are not members of extreme hate groups, but who openly express and exhibit racist viewpoints and practices, meet the requisite moral character standard? Such persons I will call “open racists.” The open racist presents many of the same problems presented by the white supremacist. The open racist is someone that bar authorities can easily identify as presenting a danger to the administration of justice. Like the white supremacist, the open racist would harm the image of the profession, particularly from the viewpoint of non-whites who are subject to our system of justice. Moreover, the open racist would present the same “parade of horribles” that were discussed with respect to the white supremacist. The open racist is likely to engage in discriminatory treatment of minorities while practicing law especially in those areas where discrimination is easily masked with a pretextual rationale for the negative treatment. As such, the open racist should also be excluded from the legal profession under the same rationale as the white supremacist.

The “closet racist,” on the other hand, is a more difficult case. The profile of the closet racist is a familiar one. The closet racist is someone who harbors racist views in his heart or the confines of his mind, and may express those views around other known racists, but is politically and socially astute enough to avoid the expression of racist views in “mixed company” or a public forum. Secretly though, the closet racist seeks to effectuate policy that operates to maintain and/or establish the political and social supremacy of whites and reveals in the opportunity to oppress non-whites and/or maintain the status quo of white privilege. The closet racist is not facially identifiable as a villain. The closet racist ostensibly proclaims that racism is wrong and argues that we should take a color-blind approach to jurisprudence. Indeed, she may be an upstanding citizen with no

188. It is important to acknowledge that even most free speech absolutists express an abhorrence of racism. Nonetheless, they are willing to tolerate it because the alternative sacrifices some First Amendment freedom which they apparently value above the right of the historically oppressed to be free from racism. See Wendel, supra note 19.

criminal record and instead a record of public or community service. He may have no conduct that one can easily identify as racist, and he may even have one or more black “friends.” The closet racist arguably presents more danger than the open racist because the closet racist ostensibly demonstrates a commitment to equal justice while secretly looking for pretextual reasons to deny rights, opportunities, and benefits to minorities. As such, this person is likely to be a respected member of the community, whereas the white supremacist generally is not. It is admittedly at this point that the exclusion of racists from the bar becomes a difficult rule to apply.

The closet racist presents evidentiary challenges. How will we determine whether a candidate for bar admission is a closet racist? Bar authorities could ask more probing questions on bar admission applications and in interviews with the candidate’s references. However, the questions could not be so overly broad so as to sweep every person who has any level of bias into the net of bar authorities for a hearing on whether the candidate is racist. Even if bar authorities could ask narrowly tailored questions that stay within the parameters of the First Amendment, it is likely that a politically savvy applicant will answer questions on an application with politically correct responses. If that is the case, we do not want bar authorities conducting investigations into every candidate’s background to determine whether the candidate has ever committed an act that is arguably racist or discriminatory. Such investigations are likely to lead to “witch hunts” in which bar authorities affirmatively seek to determine the presence of a racist taint of each candidate for admission. Even if bar authorities were as lucky as Johnny Cochran in the O.J. trial and located audiotapes or other independent evidence demonstrating the racist nature of the candidate for admission to the bar, the inquiry could not cease there. Bar authorities would also have to determine whether this candidate’s alleged racial bias is so substantial that it would likely preclude the candidate from fairly

190. Questions related to racist practices could be posed to both candidates and their references. For example, bar authorities could ask, (1) Do you participate in any racially discriminatory activities? Explain; (2) Do you practice racial discrimination or differing treatment of persons outside your racial group in any aspect of your life? Explain. These questions are different from the general communist questions asked by bar authorities in the past because racism is rationally related to the fair administration of justice as well as the core values of the legal profession.

191. Moreover, it is likely that the references provided by the candidate will be biased toward the candidate and would express ignorance of the candidate’s racist views or practices.

192. The reader may recall that when O.J. Simpson was on trial for the murder of his ex-wife, he was represented by attorney Johnnie Cochran, who introduced into evidence audio tapes of police officer Mark Fuhrman using the “N” word to refer to blacks. See JOHNNIE COCHRAN & DAVID FISHER, A LAWYER’S LIFE (2002) (discussing the Fuhrman tapes).
and effectively administering justice as a lawyer. After all, arguably we all are racist to some extent because we have been socialized by a racist society.193 At the very least, we all have our own biases. Thus, the problem is one of degree, and the issue for determination is this: To what extent does one become too tainted by racial bias to be a lawyer?

Given that some scholars have defined us all as racist,194 the efforts of bar authorities could render us all unfit to practice law if a clear line is not drawn. Bar authorities could adopt a definition of racism that would find that persons supporting affirmative action are racist against white people and therefore not fit to practice law. Likewise, bar authorities could just as easily determine that people who oppose affirmative action are racist because they are attempting to maintain the status quo of white privilege.195 Clearly, bar authorities should not be targeting proponents or opponents of affirmative action.196 Hence, the line should be drawn high enough to afford lawyers and bar applicants the room to hold and express political thought while protecting against a substantial risk of harm to the legal profession and the public it serves.

As stated previously, it is likely that the closet racist candidate would answer bar authorities’ questions in a politically correct fashion and would assure them that he would not discriminate based on race in the selection of clients or any other aspect of practicing law. Accordingly, a consumer protectionism rationale for excluding the closet racist would not suffice unless bar authorities refused to accept the closet racist’s assurances based on a finding that such assurances lack credibility due to some independent evidence of racism that cast doubt on the candidate’s commitment to equal justice. The best rationale for excluding the closet racist is the professional protectionism rationale wherein we seek not only to protect the image of lawyers, but also to protect the image of our system of justice so that it

193. Lawrence, supra note 32, at 317-24 (arguing that because Americans share a common historical and cultural heritage in which racism has played an integral role, we are all racist to the extent that this cultural belief system has influenced us). But see JODY DAVID ARMOUR, NEGROPHOBIA AND REASONABLE RACISM: THE HIDDEN COSTS OF BEING BLACK IN AMERICA 17, 126-30 (1997) (acknowledging that unconscious racial discrimination, “Negrophobia,” influences the judgment of all Americans, while also discussing the different types of racists, including the aversive racist, and arguing that not all persons who are influenced by cultural stereotypes are racist because some people renounce cultural stereotypes and develop their own personal beliefs about race).

194. Lawrence, supra note 32, at 317-24; see also ARMOUR, supra note 193, at 68-80 (discussing unconscious bias).

195. Wildman & Davis, supra note 189, at 11-20 (acknowledging that whites live their daily lives as beneficiaries of privilege created by systematic racism).

196. For an enlightening discussion of the arguments in support of and against affirmative action, including a discussion of whether affirmative action is racist discrimination against members of the majoritarian group, see Myrl L. Duncan, The Future of Affirmative Action: A Jurisprudential/Legal Critique, 17 HARV. C.R.-C.L. L. REV. 503 (1982).
appears fair, not arbitrary, and color-blind, not racist. This rationale alone, however, is not sufficient to warrant exclusion of the closet racist from the bar.

Accordingly, bar authorities should not affirmatively seek to exclude the closet racist from the bar. Bar authorities should restrict their moral character analyses to determining whether there is evidence in the record demonstrating that the candidate for admission is so racist that the candidate is unlikely to adhere to the anti-discrimination laws of our country and/or the rules of professional conduct so as to create a significant risk of harm to the population served by the profession and of harm to the profession itself. Some may argue that a policy that excludes white supremacists and other openly racist candidates while allowing the secretly racist candidates to practice law does not correct the problem of racism in the administration of justice. I agree. The problem of the closet racist will persist, but there are other mechanisms for dealing with the closet racist that do not involve the rule of law. The law is not the only tool for eradicating discrimination. Nonetheless, law should do its part. A rule excluding open racists from the bar ensures that the severely “diseased dogs” will be excluded from the profession and will not be permitted to use the legal system to effectuate their racist agendas. Likewise, the implementation of a policy against racism demonstrates to applicants, lawyers, and the public that the legal profession demands non-discriminatory behavior from lawyers and is committed to providing equal justice.

C. The Mythological Black Supremacist

When I told students and colleagues from the majoritarian group about the thesis of this Article, universally I was asked, “What about the black supremacist? Would your thesis be the same?” Intentionally displaying a state of ignorance, my reply was consistently, “What do you mean by black supremacist? Give me an example.” The two most common examples proffered were Al Sharpton and Louis Farrakhan. As I anticipated, my students and colleagues were comparing oranges with apples. Neither Al Sharpton nor Louis Farrakhan is

197. Political power (voting), economic power (spending), and education have proven effective tools in combating racism.
198. Rhode, supra note 25, at 509.
199. Admittedly, this policy is similar to the “don’t ask, don’t tell” policy adopted by President William Clinton pertaining to gays in the military. See 10 U.S.C. § 654 (2000).
200. Interestingly, rarely was this question presented by students or colleagues of color. Rather than argue the reason for this, I offer it as something worth pondering and leave it to the psychologists to deconstruct.
like Matthew Hale. Neither Al Sharpton nor Louis Farrakhan promotes or advocates subjugation or marginalization of whites. Rather, these figures point to the harm that whites inflict upon the black community and call upon the black community to unite and utilize what power it has to fight oppression.202 Neither leader reasonably could think that blacks are better than or superior to white people because one half of a millennium of history and its present-day remnants, including politics and social customs, tell them as well as the rest of us otherwise. So even if they do proffer statements proclaiming black supremacy, their statements and actions, like the statements and actions of white supremacists, must not be viewed under an anti-historical lens, but rather a lens that accurately reflects the nomos in which such statements were made. Such statements could only have been made after, and therefore presumably in response to, the social, political, and legal institutions which had already firmly established, through hundreds of years of laws and social custom, that blacks are inferior in the eyes of the law and society.

While I am inclined to agree with the argument that the “black supremacist” is a mythological being that has yet to be seen, I am compelled to acknowledge the underlying question which is at the heart of the majoritarian concern. The majoritarian question regarding the black supremacist assumes that blacks can be racist against whites. Indeed, critical race theorists have attacked this presumption and quite persuasively argued that blacks, at least on the group level, do not have the requisite power to “race” whites.203 While I agree with this theory when applied on the group level, it becomes less persuasive when applied on the individual level. On the individual level, there could be a black applicant to the bar who belongs to an underground organization or movement that seeks to and is preparing to annihilate, deport, and/or enslave white people. There could be a black applicant whose religious tenants mandate that he never hire a white person nor have any business with a white per-

202. See NBC interview with Minister Farrakhan, available at http://www.abbc2.com/islam/english/toread/farnbc.htm (last visited Mar. 16, 2003) (wherein Minister Farrakhan stated his position as being based on the notions of “equity and reciprocity”). Referring to black people, Minister Farrakhan stated, “We cannot allow ourselves to be controlled by any outside group. We must take control of our own destiny. That is what I preach, and that is what I believe, and that is what I’m striving for.” Id. See also Transcript from Minister Louis Farrakhan’s Remarks at the Million Man March, Oct. 17, 1995, available at http://www.cnn.com/US/9510/megamarch/10-16/transcript/index.html (last visited Mar. 16, 2003) (saying among other things that “white supremacy has to die in order for humanity to live” and challenging blacks to be productive members of society by saying, “Black man, you don’t have to bash white people, all we gotta do is go back home and turn our communities into productive places.”)

son.\textsuperscript{204} If this is in fact the black applicant’s agenda, such an applicant could also harm the legal profession. The extent and gravity of the harm is likely to be less than the harm that a white supremacist could impose because the system of justice will be more circumspect of a black lawyer’s actions than a white lawyer’s actions because history has demonstrated that blacks lack the credibility of whites in our system of justice,\textsuperscript{205} and because our system of justice is predominately white, which makes it a place where the white supremacist can more easily operate “under cover” while the black lawyer stands out like the proverbial “fly in buttermilk.”

Nonetheless, the harm that a black supremacist/separatist lawyer could impose on the system is not so \textit{de minimis} as to justify his admission. If the black separatist is more likely than the average lawyer to violate the rights of a white person, as well as the laws of lawyering, he too is a threat to the administration of justice and should be excluded from the bar. It should be noted, however, that the black separatist would have a diminished impact on the profession because the profession is not controlled by his racial group, but rather by the majoritarian group. Thus, it is unlikely that the black separatist will be admitted by the majoritarian-controlled bar, which could likely find some other reason to exclude the candidate other that his racist beliefs. Even if admitted, the black separatist would find few allies in the majoritarian-dominated profession and would likely have a difficult time being appointed to a position such as prosecutor or judge since it is predominately the majoritarian group that controls such appointments. Thus, while the harm that the black separatist could impose is intolerable to this author, it is likely to be significantly less than the potential harm that could be imposed by the white supremacist who, although not accepted by the majoritarian group, is less feared than the black separatist.

\textsuperscript{204} Such blacks are frequently referred to as black separatists, not black supremacists, because they preach separatism—blacks living apart from white society in order to escape racism and white supremacy. The separatist movement also preaches black pride, which is an effort to motivate a disenfranchised and oppressed group of people. The black separatist movement seems quite distinguishable from the white supremacist movement because the underlying premise of the respective movements is different as well as the historical context. The white supremacist movement developed in response to efforts to achieve racial equality, and its agenda seeks a return to the days of segregation, or even racial “cleansing” of the non-white races, on the theory that such non-white people are inferior and undeserving of equality with whites. The black separatist movement, on the other hand, is a response to oppression and has as its agenda the separation of the races as a way of self-preservation. This agenda is based upon the historical and arguably present-day reality that people of color, particularly blacks, have been oppressed by whites, and hence the desire to avoid oppression by whites fuels the idea of separatism. In common terms, the theory behind the separatist movement is that if the dog bites, one is best assured of safety by staying away from it rather than trying to domesticate it.

\textsuperscript{205} For a persuasive article supporting this proposition, see Sheri Lynn Johnson, \textit{The Color of Truth: Race and the Assessment of Credibility}, 1 MICH. J. RACE \\& L. 261 (1996).
D. The Unconscious Racist

The unconscious racist is someone who does not consider herself a racist. She professes disdain for racism and those who practice it. The unconscious racist has been defined as the person who does not reveal racist tendencies at all, except as “persistent unconscious fantasies.” It has been argued that all of us are unconscious racists to the extent that we have been influenced by a racist cultural belief system. We are unconscious racists because we “do not recognize the ways in which our cultural experience has influenced our beliefs about race or the occasions on which those beliefs affect our actions.” We function on a daily basis by allowing race to influence our decision-making process and our value judgments of others without even realizing that we are doing it.

The unconscious racist should not be the target of bar authorities because the unconscious racist has made a conscious commitment to equal justice. The unconscious racist aspires to be a fair-minded, non-discriminatory actor. In other words, the unconscious racist has made a commitment to and accepts the core value of equal justice even though she may from time to time fall short of her aspirations of equality. Bar authorities cannot seek absence of bias; all they can demand from an applicant is a commitment to equal justice. This commitment should be a criterion for determining whether a candidate for admission to the bar possesses the requisite moral character to be an officer of the court and an administrator of justice.

VII. CONCLUSION

Racism has taken the form of black slavery in our country, which was remedied by the Thirteenth Amendment to our Constitution. It has also taken the form of denying blacks and other people of color the right to serve on juries and to vote, which was remedied by the Fourteenth and Fifteenth Amendments to our Constitution. Finally, racism has taken the form of denying citizens of color the right to eat in restaurants and to sleep in hotels, which was remedied through Congress’ public accommodation laws. In other words, history teaches us that, if allowed, racism can be integrally connected with the administration of justice. If racism is allowed to remain a component of the administration of justice through racist administrators,  

207. See Lawrence, supra note 32, at 317-24.
208. Id. at 322.
209. One need only re-read the decisions in Dred Scott v. Sandford, 60 U.S. 393 (1856), and Plessy v. Ferguson, 163 U.S. 537 (1896), to be reminded of this. For a more recent but equally tragic example of how racism can infect the administration of justice, see United States v. Price, 383 U.S. 787 (1966), in which three racist officials, with the aid of fifteen other whites in the state of Mississippi, murdered three civil rights workers.
it will prejudice our administration of justice, just as it has in the past, albeit much more covertly and virtually imperceptibly in this circumstance. If racist applicants are allowed to become lawyers, there will be no group justice for those who are the target of such lawyers' prejudice.

Much discussion has been had about the color-blind theory of justice and the color-blind nature of our Constitution. Our highest Court has espoused the view that our society should strive to achieve and administer color-blind justice. If color-blind justice is the goal that the legal profession seeks to obtain, the admission of racists to the bar would serve to thwart that effort. Entrusting racists with the power to administer justice to non-whites threatens to carry us further away from the goal of a judicial system that is color-blind. So if color-blindness is the goal for our legal system and not just a catchy phrase to be used to promote the interests of whites, then persons who seek to use color as the basis to deny our citizens of color full participation in and protection under our legal system should be excluded from the bar. If we learn nothing from the mistakes of our past, we are destined to repeat them.

210. See Wilkins, supra note 21, at 1514-15 (discussing “bleached out professionalism” and its link to the color-blind theory of justice and noting that the phrase “color blind” as used to refer to the nature of our Constitution was first made famous by Justice Harlan in his dissent in Plessy v. Ferguson, 163 U.S. 537, 557 (1896) (Harlan, J., dissenting), overruled by Brown v. Board of Education, 347 U.S. 483 (1954)); Appiah & Gutmann, supra note 174 (arguing that fairness in a color-conscious society demands color-conscious law and policy because color-conscious policies are instrumental in overcoming historically imposed racial injustice); see also Neil Gotanda, A Critique of “Our Constitution is Color-blind”, 44 STAN. L. REV. 1 (1991) (arguing that the U.S. Supreme Court’s use of color-blind constitutionalism fosters white racial domination). Cf. Wechsler, supra note 56.

211. For a discussion of the Supreme Court’s “color-blind” jurisprudence, see Jerome McCristal Culp, Jr., Toward a Black Legal Scholarship: Race and Original Understandings, 1991 DUKE L.J. 39.
THOREAU’S PENCIL: SHARPENING OUR UNDERSTANDING OF WORLD TRADE

JAMES BACCHUS*

I brought my “power point” with me today. I even sharpened it.
This is a “Number Two” pencil.
This “Number Two” pencil makes the point I wish to make today about the power of trade.
This pencil belongs to me.
But in another, broader, truer sense, this pencil belongs to Henry David Thoreau.
In truth, we might rightly describe this “Number Two” pencil as “Thoreau’s pencil.”
Why? Why is this “Thoreau’s pencil”? And why does an understanding of why this is “Thoreau’s pencil” help sharpen our understanding of the significance of world trade?

Like the answers to so many other questions, the answers to these questions about Thoreau’s pencil are found where they keep the books. They are found in the New York Public Library.

A few years ago, on a visit to New York, my wife, Rebecca, and I went to the New York Public Library. There we saw a special exhibit about the best American books by the best American writers. One of the writers featured in the exhibit was Henry David Thoreau.

Among the items in the exhibit were the handwritten pages from Thoreau’s journals, the earliest drafts of Thoreau’s essays, and an early edition of Thoreau’s timeless classic, *Walden*—his lyrical account of the months he spent in self-imposed solitude in the 1840’s in a homemade hut in the woods beside Walden Pond.

* James Bacchus is the Chairman of the Appellate Body of the World Trade Organization in Geneva, Switzerland. He is the youngest Member, the longest serving Member, the only North American Member, and the only remaining original Member of the global tribunal, which hears final appeals in international trade disputes involving the ninety-five percent of world commerce conducted by the 145 countries and other customs territories that are Members of the WTO. Bacchus is a former Special Assistant to the United States Trade Representative in the Executive Office of the President. He is also a former member of the Congress of the United States, from Florida. He served two terms in the U.S. House of Representatives before choosing not to seek re-election in 1994. In 1995, he was appointed by the Members of the WTO to the Appellate Body. Bacchus received a Bachelor of Arts degree from Vanderbilt University, Magna Cum Laude and Phi Beta Kappa, with High Honors in History. He received a Master of Arts degree from Yale University, where he was a Woodrow Wilson Fellow in History. He was graduated with High Honors from the Florida State University College of Law, where he was Editor-in-Chief of the FSU Law Review. These are his personal views, and are not the views of his colleagues on the Appellate Body or the official views of the WTO.
Also among the items in the exhibit was a plain wooden pencil. It was Thoreau’s pencil. However, the pencil in the exhibit was not just a pencil that had been owned by Thoreau. It was also a pencil that had been made by Thoreau.

The exhibit explained that Thoreau’s father had owned a company that made pencils in their hometown of Concord, Massachusetts. The exhibit explained also that Thoreau had worked for a time in the family business of making pencils, and that his father had urged his son to make a career of laboring—not at writing essays—but at the more stable and more secure profession of making pencils.

Of course, like so many sons, Henry chose not to take his father’s advice. He chose to make his living as a writer. He chose to write the essays we still read today. He chose to go to Walden Pond.

This choice made long ago by the youthful Thoreau helps sharpen our understanding of the significance of world trade today. It does so in a way that helps clarify what is really at stake for all of us in world trade.

Here is why. Some time after my visit to the New York Public Library, I was re-reading Thoreau’s Walden for the “umpteenth” time when I was struck by this question Thoreau posed in the provocative first essay in Walden, on “Economy”: “Where is this division of labor to end? and what object does it finally serve? No doubt another may also think for me; but it is not therefore desirable that he should do so to the exclusion of my thinking for myself.”

I wondered then, and I wonder now: What would have happened if Thoreau had acted in his own life on the basis of his own obvious reservations about a division of labor? What would have happened if he had taken his father’s advice, ignored the call of his own unique talents as a writer of essays, and chosen to minimize the division of labor in his own life by spending all his time making all his pencils? Would Walden and all of Thoreau’s other enduring essays even have been written if Thoreau had not been able to benefit in his own life from the very division of labor that he denounced? Whether consciously or not, many of those who oppose world trade today are only echoing Thoreau’s reservations about a division of labor. They need to know more about Thoreau’s pencil. For the choice Thoreau faced in his life is the same choice we all face in our lives every day with every kind of trade. Will we do it ourselves? Or will we pay someone else to do it—whatever “it” is—so that we can have more time and more freedom to do something else? Will we choose to pay the “opportunity cost” of doing it ourselves, or will we choose a division of labor?

Trade is nothing more than the consequence of this choice. Trade is nothing more than a division of labor. Trade is about pencils. Trade is about making pencils. Trade is about buying and selling pencils. Trade is about the division of labor that is evidenced in every pencil that is made and bought and sold in every part of the world.

In all our debates about “globalization”—in all our discussions about the “pros” and “cons” of the World Trade Organization—in all our understandable attention to all the “ins” and “outs” of international trade negotiations, international trade agreements, and international trade disputes—in all our day-to-day attention to all the many arcane details of world trade—we have a tendency at times to forget what trade really is.

Trade is simply the exchange that results from a division of labor. Trade is simply the exchange of pencils. Some of our pencils are called “goods.” Some are called “services.” Some are the ideas that we call “intellectual property.” Yet, whatever we may call what we trade, everything we trade is some kind of a pencil.

Sometimes it is only a simple matter of the teenager next door mowing your lawn. Sometimes it is the clerk at the corner grocery selling you a carton of milk. Sometimes it is the bookseller at the local bookstore selling you a book.

Sometimes the trade resulting from the division of labor is only a simple local exchange. But, many times, it only seems that way. Often times, it is really much more. Often, the division of labor is a complex matter of buying a complicated “high-tech” instrument such as this “Number Two” pencil.

This pencil is the end result of the application of centuries of increasingly sophisticated technology. This pencil is a combination of highly-crafted parts from all the far corners of the world. This pencil has been assembled and finished and brought into the marketplace through the unique talents of many different individuals who all came together and worked together to make it. Often, our pencils do come from just next door. But, more often, our pencils come from somewhere else. Frequently, and increasingly, trade is “world trade.” “World trade” is called “world trade” only because some part of the pencil that is traded happens to cross some arbitrary and artificial political border. Apart from that, “world trade” is no different, economically, from hiring the kid next door to mow your lawn, or maybe make your pencil. I am not the first to use a pencil to illustrate this point. In an essay written in the 1950’s entitled “I, Pencil,” Leonard Read assumed the persona of a pencil. His pencil tells the tale of all
its many parts and many makers, and declares proudly that “not a single person on the face of this earth knows how to make me.”

Here is the point. Even the simple pencil is the complex product of the individual talents of many different people in many different places working together in many different ways. All of them know something about making a pencil. But not one of them knows everything that needs to be known to make even one pencil. And the same is true for virtually every other good and service that is exchanged in world trade.

Milton Friedman praised Read’s essay for illustrating, with the simple pencil, the possibilities of cooperation without coercion through the workings of the “invisible hand” of the market economy, and the impossibilities of an insular self-sufficiency that depends for success on having a breadth of knowledge that often is so dispersed among so many people in so many places that it is only available to everyone through cooperation.

I agree. As I see it, this is the key to understanding the significance of world trade. And, as I see it, this key can be turned only when we understand the indispensability of an international division of labor to all that we hope to achieve in the individual lives of all humanity.

Thoreau told us, in Walden, that he “went to the woods” because he “wished to live deliberately.” That is the common aim of all humanity. That is what we all wish for. Whoever we may be, wherever we may be, however we may define the good life, we all wish “to live deliberately.”

We all want to make life happen for us, and not just let life happen to us. We all want to use the unique talents that God gave each of us in ways that will help each of us give our lives more real and lasting meaning. And we all ask ourselves: how best can we do this?

In search of meaning in his life, Thoreau sought solitude. He sought the simplicity of self-sufficiency in the woods beside Walden Pond. His sojourn in the woods seems to suggest that the best way “to live deliberately” is to live alone, to work alone, to be alone. The message of Walden seems to be that we are most likely to find meaning in human isolation.

I keep reading and re-reading the writings of Henry David Thoreau because I believe that Thoreau understood human freedom. He

---

4. THOREAU, supra note 1, at 101.
understood that human freedom is, ultimately, about the autonomy, the integrity, and the dignity of the free individual. He understood that human freedom is individual freedom.

Thoreau understood also the real purpose of individual freedom. Human life is about living “deliberately.” It is, as he wrote, about confronting “the essential facts of life,” about learning what life has “to teach,” so that we will not, in the end, discover that we have not lived.5

But Thoreau did not understand how we can each best secure the meaningful individual freedom that will enable each of us “to live deliberately.” We cannot “live deliberately” if we choose to live in isolation. We can “live deliberately” only if we choose to live in ways that further cooperation.

If we take Thoreau’s advice and think for ourselves, if we think clearly, and if we think things through, then surely we will realize that we need the human cooperation of a division of labor.

Thoreau urged each of us to find meaning in life by trying to do it all ourselves. He thought it best “to oversee all the details yourself in person; to be at once pilot and captain, and owner and underwriter; to buy and sell and keep the accounts; to read every letter received, and write or read every letter sent; to superintend the discharge of imports night and day.”6

Thoreau would have us all seek self-sufficiency—down, as in Walden, to the last half-cent. He would not have us depend on other people. He would have us depend only on ourselves. He would have us all make all our own pencils. But, the truth is, we can never be entirely self-sufficient, either economically or otherwise. The truth is, we must depend on other people. We need other people. We each need other people if we hope to be able to fulfill the divine promise that is embedded deep within each and every one of us.

This was true of Thoreau—who needed someone else to make at least some of his pencils, so that he would have more time and more freedom to make all his magical essays. This is true of all of us—in as many individual ways as we have individual talents and individual dreams of using them.

Thoreau was inclined—as he once put it—to “measure distance inward and not outward.” He was right in thinking that it is essential for each of us to look inward to learn who we are and who we hope to become—as individuals. Yet, as I measure it, we must then

5. Id.
6. Id. at 51.
look outward, and use what we have learned about ourselves in cooperative ways that diminish our distance from others.

Thoreau was fond of paradox. He was intrigued by all the apparent contradictions in life. But one paradox he failed to see is this. Very often in life, simplicity needs complexity. The simplicity we seek as a way of inspiring self-discovery and self-fulfillment can only be found through the complexity of mutual cooperation.

For none of us can ever become all we might become unless all of us are able to develop fully all the “special” individual ways in which we all are unique. And none of us can ever develop fully as unique individuals without economic and other associations with other people.

This means we need the “specialization” of a division of labor. This means we need trade. The division of labor that is trade is a liberating force that is essential to unleashing the unique power of human thought through human initiative, incentive, invention, innovation, inspiration, imagination, ingenuity, and enterprise. It is a force for freeing the vast untapped potential for the singular creativity of humankind.

In one of his later essays, “Life Without Principle,” Thoreau lamented that—as he expressed it—“we are warped and narrowed by an exclusive devotion to trade and commerce and manufactures and agriculture and the like, which are but means, and not the end.”

Thoreau was right in concluding that the chief ends of life are not producing and consuming. He was right to urge us to look beyond mere materialism to see everything that truly can give life real meaning. But he was wrong to think that the true ends of life are not served by the means of trade.

Trade is a means to all the many ends of human freedom. Trade is a means of making more choices available to more people so they can make more personal choices about how they wish to live. Freedom is about choices. Freedom is choosing. The equation between trade and freedom is this. More trade equals more choices equals more freedom.

The division of labor multiplies human productivity, and, thus, human prosperity, and, thus, human opportunity. It multiplies human choices, and, thus, multiplies human freedom. It empowers more of us “to live deliberately.” This is equally so whether we hire the teenager next door to mow our lawn, or the worker on the other side of the planet to make our pencil.

By dividing our labor, by creating an ever-widening and ever-deepening international division of labor through world trade, we are establishing an economic foundation for uniting all the world in the deliberate life of freedom.

We are limited only by the reach of the market. And, more and more, the reach of the market is limited by less and less. More and more, we have more world trade in what is more truly a world economy characterized by the ever-dividing subdivisions of a more truly international division of labor.

In the midst of all the many controversies about world trade, we tend to forget why we trade. We tend to forget all the positive effects of an increasingly international division of labor in furthering and facilitating human freedom. We tend to forget what the world would be like if we all had to make all our own pencils.

In his definitive study of the history of the humble pencil—entitled The Pencil: A History of Design and Circumstance—Henry Petroski had much to say about Thoreau’s pencil and all it illustrates. He pointed out that, before Thoreau left to go to the woods, he made an exhaustive list of everything he needed to take with him—but he forgot to mention his pencil. And, yet, Thoreau kept his pencil with him, in his pocket, all the time.9

Trade, too, is with us all the time, and we tend to take the positive effects of trade for granted. We take the positive effects of trade for granted when trade is the kid next door who wants to mow our lawn. And yet we fret about trade when trade is someone in some other country who wants to help make our pencil. We fret endlessly about the competition that is an inevitable part of world trade.

I have this “Number Two” pencil only because my eleven-year-old daughter, Jamey, has not yet “borrowed” it. Pencils often disappear mysteriously from the pencil box on my desk at home. Jamey likes to sharpen the pencils she “borrows” in the pencil sharpener on the corner of my desk. She sometimes sharpens a whole handful of pencils at once. I have read that the average pencil can be sharpened seventeen times, but I’m not sure this is so with the pencils that Jamey grinds so eagerly in my pencil sharpener.10

The competition in world trade is like the grinding of a pencil sharpener. It makes a lot of noise. It uses up a lot of pencils. Yet it is absolutely necessary in order to make proper use of a pencil. Only with the relentless spur of free and fair competition can trade suc-

---

ceed. Only with the sharpened pencils of trade can we write all the words of human freedom.

But there must be rules to ensure free and fair competition in world trade. There must be rules to help free us to make more trade so that we can make more freedom. And that is why we have created the World Trade Organization. That is why we have created the much-needed but much-debated, much-maligned, and much-misunderstood “WTO.”

One common misunderstanding about the WTO is much like one common misunderstanding about the pencil. We commonly speak about a pencil “lead.” Yet the “lead” of a pencil is not really made of lead. It is made of a mixture of graphite and clay. Likewise, we commonly speak of the “WTO” as if it were some all-powerful, supranational organization that is somehow able to impose its arbitrary will on us by telling the sovereign countries of the world what to do. Yet the WTO is really nothing more than those very sovereign countries working together as the “WTO” to provide the right mix of rules the world needs to make more trade and, thus, more freedom.

The WTO is a cooperative effort by 145 countries and other customs territories to ensure the best mix of all the graphite and all the clay that will be needed to make all the many pencils that are needed by the world. This mix is made in a world trading system that serves five billion people in ninety-five percent of the world economy.

This is where I come in. I help the Members of the WTO clarify the rules of world trade, and I help them uphold the rules of world trade, so that the grinding competition of world trade will be freer, will be fairer, and will continue to make more trade and, thus, more freedom.

There are seven of us who have been appointed by the Members of the WTO to serve on the Appellate Body of the WTO. We seven work for all the Members of the WTO. We seven are independent and impartial.

When there are disputes about what the rules mean, the countries that make the rules, and that are bound by the rules, have the right to resolve those disputes in what we call the WTO dispute settlement system. The seven of us on the Appellate Body help the Members of the WTO uphold the rules by assisting them in their efforts to decide what the rules mean by drawing all the right lines in WTO dispute settlement.

Pencils are made for drawing lines, and the world needs rules for trade that draw all the right lines. The average pencil can write
45,000 words.\textsuperscript{11} The average pencil can draw a line thirty-five miles long.\textsuperscript{12} We need rules for trade that will have all the right words to draw all the right lines to take us all the long miles to freedom.

We need rules for trade on which the countries of the world have agreed. We need rules for trade that can help lower the barriers to trade, and help resolve trade disputes. We need rules for trade that can help provide the stability, the security, and the predictability that are needed for trade to be as successful as it can be in creating more freedom.

We already have many of the rules the world needs for trade. There are 30,000 pages of trade rules in the WTO Treaty. There are 13,000 pages of reports on these trade rules that have resulted thus far from WTO dispute settlement. Along with my six colleagues on the Appellate Body, I have helped the Members of the WTO draw many of the lines we need by writing many of those pages.

The many countries in the world are busy now trying to agree on the additional lines that need to be drawn in world trade. In the new worldwide trade negotiations under the auspices of the WTO, many countries are working together to make new rules for trade in the same way that many people work together to make a pencil.

But having all the rules we need for trade will not be enough.

The rules must be fair. The rules must be the same for everyone, and they must be applied to everyone in the same way. This is what is called the “rule of law.” And the rules must be upheld. Rules are not really rules unless they are upheld. In upholding the rules for trade, we are upholding the “rule of law.” Freedom is only possible within the “rule of law.” We cannot write without some kind of pencil, and we cannot be free without the lines that form the freeing framework of the law.

Only if we draw all the right lines, only if we have all the right rules, only if those rules are fairly written and fairly applied, and only if those rules are upheld, will we be able to \textit{maximize} all the many gains that can be made for freedom through the many gains from trade. Only then will we be able to make it possible for millions more people in every part of the world “to live deliberately.”

The more trade we have, the more gains from trade we will have to maximize, and the more “deliberately” we will all be able to live. And we will have the most trade—we will have the most personal choices—we will have the most freedom—if we trade in ways that enable—that empower—each of us to do what we each do the best

\begin{itemize}
  \item[12.] \textit{Id.}
\end{itemize}
when compared to others. The world will work best if we each do what we each do the best, when compared to others.

This is what the economists call our relative, “comparative advantage.” This is also the way for each of us to have the best chance to choose “to live deliberately.” Because a world in which we are each free to do what we each do the best when compared to others, is the world in which we will each be the freest to make a deliberate choice about how best to live.

At some basic level, Thoreau understood this. At some fundamental level, he understood the need for a division of labor that enables each of us to meet and beat competition by using and making “comparative advantage.” The evidence is not in what he said. It is in what he did. It is in what Thoreau did about Thoreau’s pencil.

Thoreau rejected his father’s advice to make a career of making pencils. He preferred to wander in the woods and write essays at Walden Pond.

But Thoreau was an American writer of American books. He was very much an American. And, when the going gets tough, we Americans don’t go live alone in the woods. We don’t flee the world. And we don’t fear the world. We face the world, and we face all the tough challenges the world presents.

Sometimes those challenges are in trade. Sometimes those challenges are in space. Sometimes those challenges are in some far dark corner of the world. Wherever and whatever those challenges are, we Americans always—always—meet them.

Thus, when competition from foreign pencils threatened to drive his father out of business, Thoreau went back to work at his father’s pencil company. In the face of the foreign competition, he showed up to save the family livelihood.

Thoreau spent long hours in the Harvard library studying the finer points of pencil technology. He developed a new grinding mill, a new pipe-forming machine, new water wheel designs, and all sorts of new processes for making pencils. He discovered a new way of mixing clay with graphite to make a superior pencil lead. But, most important of all, he discovered a way of varying the mix so that he could vary the hardness of the pencil lead.

This discovery saved his father’s company by making it the first American company to produce pencils with grades that varied ac-

cording to their hardness. The Thoreau pencils were numbered one, two, three, and four.\textsuperscript{14}

Thoreau knew what we need to do to meet the challenge of world trade.

We need to make a better pencil.

Make one he did. And, in making a better pencil, Thoreau showed why this pencil I hold today truly is “Thoreau’s pencil.”

By making it possible to produce pencils that vary in hardness, Henry David Thoreau helped give us this—and every other—“Number Two” pencil.

\textsuperscript{14} Petroski, supra note 9, at 118-19.
I. INTRODUCTION

It is often said that there are two sure things in life: death and taxes. Sometimes the two converge, and together form what is hatefully referred to as the “death tax.” For as much as the average American grimaces at the twenty or thirty percent taken from their paycheck each month, they downright despise the fact that the same money could be taxed again at their death. With that, some Americans fear the knocking of the Internal Revenue Service (IRS) at their deathbed just as voraciously as they do the grim reaper. As a result, taxpayers and their very able tax advisors have toiled to formulate
numerous trap doors and escape routes through the endless maze of tax codes, revenue rulings, and case law, all in the name of avoiding the dreaded “death tax.”

One of the more ubiquitous of these tax avoidance tools is the Crummey power. Named after the 1968 Ninth Circuit case, Crummey v. Commissioner,2 Crummey powers have morphed from a mechanism originally designed to level the estate tax playing field3 into a powerful estate planning tool, used by some to siphon off large portions of their estate free and clear of any estate or gift tax implications. On its face, such a tool seems to be plainly at odds with the purpose and design of the estate and gift tax structure. Despite this seemingly obvious abuse, in the thirty-four years since Crummey, the IRS has been relatively ineffective in leveling an attack of any significance against the massive estate tax avoidance precipitated by the advent of Crummey powers. As a result, the Crummey power has become a firmly entrenched tool in the estate planners’ arsenal; and without any significant structural change to the Crummey mechanism itself, the federal fisc will continue to watch as large chunks of potential estate tax revenue are siphoned away, $11,000 at a time.

II. OVERVIEW OF THE ESTATE AND GIFT TAXES

First enacted in 1916, the federal estate tax was intended to serve a two-fold purpose. In addition to the fundamental goal of raising revenue, it was hoped that the imposition of the estate tax would break up large concentrations of wealth. Although some have argued that the estate tax has failed on both counts,4 its presence nonetheless continues to be a major target of disdain and abuse.

Despite imposition of the estate tax in 1916, taxpayers were relatively unaffected by the tax, for it was easily avoided by gift transfers made during one’s life. To correct this formidable loophole, Congress enacted the gift tax in 1932. As the House Ways and Means Committee and the Senate Finance Committee aptly explained:

The gift tax will supplement both the estate tax and the income tax. It will tend to reduce the incentive to make gifts in order that distribution of future income from the donated property may be to a number of persons with the result that the taxes imposed by the higher brackets of the income tax law are avoided. It will also tend to discourage transfers for the purpose of avoiding the estate tax.5

2. 397 F.2d 82 (9th Cir. 1968).
3. See infra Section V.
In practice, the federal gift tax was established to provide a “safety net” for the estate tax. To prevent taxpayers from depleting their estates prior to death, thereby considerably reducing their estate tax liability, the gift tax, subject to certain limitations, imposes a tax upon all transfers or gifts, whether “direct or indirect, and whether the property is real or personal, tangible or intangible.”

Notwithstanding the enactment of the gift tax, significant advantages remained available to taxpayers who chose to use inter-vivos transfers to convey their estate to third parties. Initially, the gift tax was levied using a completely different schedule than the estate tax (set at roughly three-fourths the estate tax maximum) and was tax-exclusive while the estate tax remained tax-inclusive. Most important, for the purposes of this Comment, the gift tax contained an annual exclusion provision that allowed donors to make tax-free, inter-vivos transfers, while the estate tax contained no such provision, thereby providing considerable incentive for taxpayers to “gift away” their estates during their lifetime.

In an attempt to remove the preferences within the gift tax that served to thwart the purpose and utility of the estate tax, Congress adopted a unified transfer tax system in 1976. As part of this unified transfer tax system, Congress enacted the Unified Credit against the estate tax. Intended to mitigate some of the exposure faced by small and medium-sized estates subject to the estate tax, the Unified Credit serves to remove most taxpayers from the reach of both the estate and gift taxes. Under the current Unified Credit, the first $1,000,000 of estate property ($2,000,000 for married couples) is

6. See Dickman v. Comm’r, 465 U.S. 330, 338 (1984) (holding that “one of the major purposes of the federal gift tax statute is to protect the estate tax and the income tax”).
9. Id. The tax-exclusive/tax-inclusive distinction, although subtle, nonetheless creates an inherent advantage to those taxpayers who choose to transfer wealth during their life as opposed to at their death. As a tax-inclusive levy, the estate tax is imposed upon the total amount of the decedent’s gross estate (including those assets used to pay the tax). Whereas, the gift tax, which is tax-exclusive, is imposed only upon those assets transferred to a beneficiary, and not upon the assets used to pay the tax. For example, assuming a flat transfer tax of 50 percent, if Wyatt were to make an inter-vivos transfer of $1,000,000 to his brother Chet, the gift tax would be imposed on the amount that Chet received after the tax was applied (50% of $500,000) thereby giving Chet a net gift of $750,000. On the other hand, if Wyatt were to wait and transfer that same amount at his death, the estate tax would be imposed on the $1,000,000 before it is transferred to Chet (50% of $1,000,000) thus leaving Chet with a net transfer of only $500,000. See IRA MARK BLOOM ET AL., FEDERAL TAXATION OF ESTATES, TRUSTS AND GIFTS § 2.03 (3d ed. 2002).
10. Smith, supra note 4, at 375-77.
12. Id. Currently, the Unified Credit under I.R.C. § 2010 is set at $1,000,000. However, that amount is set for a phased increase up to $3,500,000 by 2009. I.R.C. § 2010(c) (2002).
exempted from the estate tax.\textsuperscript{13} Despite the imposition of the unified transfer tax system, limitations and loopholes remained abundant, thereby affording the savvy taxpayer ample opportunities to gift away their estate free from gift taxation, and reduce their subsequent estate tax burden in the process.

One of the more obvious limitations on the reach of the gift tax is the exclusion that applies to payments made by a taxpayer to satisfy the education or medical expenses of another.\textsuperscript{14} This exclusion does not require that the taxpayer and beneficiary have any specific relationship; nor is there any limitation on the amount that may be exempt from taxation when used for the specific purposes outlined in section 2503(e). Further, the removal of these “basic needs” from the reach of the gift tax effectively increased the value of the annual exclusion for those taxpayers who had previously exhausted their available annual exclusion to meet the financial needs of their adult children in school and/or elderly parents or relatives in need of medical care.\textsuperscript{15}

\section*{III. Overview of the Annual Exclusion}

Although the stated purpose of the gift tax was to prevent estate tax avoidance, the taxation of all gifts, no matter how inconsequential or small, would certainly prove unduly burdensome to the average taxpayer. To avoid that result, Congress enacted an annual exclusion to the gift tax:

\begin{quote}
[t]o obviate the necessity of keeping an account of and reporting numerous small gifts, and, on the other [hand], to fix the amount sufficiently large to cover in most cases wedding and Christmas gifts and occasional gifts of relatively small amounts.\textsuperscript{16}
\end{quote}

The rationale behind this exclusion is rather basic: If the government required every small, customary gift (i.e., Christmas, birthday, etc.) to be accounted for and subsequently taxed, massive noncompliance and collection problems would result, thereby eroding the credibility of the voluntary tax system as a whole.\textsuperscript{17}

Originally set at $5,000 per donee, per year in 1932, the annual exclusion was later reduced to $4,000 in 1939, and $3,000 in 1942.\textsuperscript{18} The exclusion remained at $3,000 until 1981, when Congress enacted the Economic Recovery Tax Act (ERTA) which, in one fell swoop, in-

\begin{footnotesize}
\begin{enumerate}
\item Id.
\item See I.R.C. § 2503(e) (2002); Steinkamp, supra note 8, at 116.
\item Steinkamp, supra note 8, at 116.
\item S. REP. NO. 72-665, at 41 (1932); accord H.R. REP. NO. 72-708, at 29 (1932).
\item See Steinkamp, supra note 8, at 118.
\item Id. at 118-19.
\end{enumerate}
\end{footnotesize}
creased the exclusion to $10,000. The legislative history behind the reductions of the annual exclusion in 1939 and 1942 indicate Congressional concern that taxpayers could use the exclusion as a device to transfer significant portions of their estate free from the reach of the estate and gift taxes. Thus, the reduction of the exclusion amount in those years was intended to maintain the exclusion at a level that kept with the spirit behind its original enactment, without allowing taxpayers to avoid the estate tax with sizeable annual tax-free gifts given to heirs and potential beneficiaries. Given the legislative history behind the original passage of the annual exclusion in 1932, and the subsequent reductions in 1939 and 1942, it is relatively clear that Congress did not intend for the annual exclusion to be used as an estate-planning tool. Rather, from its inception, the annual exclusion was designed as nothing more than a means of reducing the potential administrative burden and taxpayer discontent associated with the accounting and taxation of the most basic, everyday gift transfers.

A. Present Interest Requirement

Because Congress intended for the annual exclusion to cover only minor, routine gifts, which are customarily transfers of present interests, it follows that the exclusion “does not apply with respect to a gift to any donee to whom is given a future interest.” Whether or not a given transfer will qualify under the exclusion is often determined by how the Service will characterize the transfer: If the transfer is deemed a “present interest,” the annual exclusion will apply; if a “future interest,” the exclusion will not be applicable and the taxpayer must pay gift tax on the transfer.

The present interest requirement arose out of concern over the applicability and abuse of the exclusion. On the one hand, Congress recognized the potential difficulty that could arise in “determining the number of eventual donees and the value of their respective gifts” if gifts of future interest were allowed to qualify for the exclusion. More importantly, by limiting the exclusion to only present interest transfers, Congress intended to ensure that its use would apply only to those transfers it was originally meant to cover—routine, ordinary gifts. However, when using the exclusion as an estate-planning tool, taxpayers often “double dip” by using the full amount

19. See Smith, supra note 4, at 383-86. Currently, the rate established under section 2503(b) has been indexed for inflation and is now set at $11,000. I.R.C. § 2503(b) (2002).
22. Given this, it is rather apparent that Congress did not intend for the annual exclusion to be used as an estate planning device.
of their exclusion for estate planning purposes, while at the same time giving those same beneficiaries the normal, everyday gifts meant to be covered by the exclusion.\textsuperscript{23} This practice is a clear subversion of the intent behind the exclusion, and is exacerbated by the use of family members in Crummey trust situations (who are most likely to receive the normal, ordinary gifts meant to be covered by the exclusion).

Despite the importance placed upon the characterization of the transfer, neither “future interest” nor “present interest” are defined in the Internal Revenue Code. Treasury Regulations have defined “future interest” as “includ[ing] reversions, remainders, and other interests or estates, whether vested or contingent, and whether or not supported by a particular interest or estate, which [is] limited to commence in use, possession, or enjoyment at some future date or time.”\textsuperscript{24} However, courts have interpreted the distinction between present and future interests in a manner that is vague and contradictory, thereby leaving the taxpayer without a consensus as to what truly qualifies as a future interest for the purpose of the annual exclusion.

In most situations, the distinction between whether an interest is considered “present” or “future” for purposes of the annual exclusion is directly correlated to the type of interest transferred to the donee. For example, outright gifts of property will almost always be considered transfers of present interest because the beneficiary is given the ability to immediately enjoy the gift. In contrast, transfers in trust are almost always intended to provide for a beneficiary’s future welfare, and often contain restrictions that strictly prohibit a beneficiary’s immediate enjoyment of the gift.\textsuperscript{25} Thus, a transfer in trust will

\begin{itemize}
\item \textsuperscript{23} Smith, \textit{supra} note 4, at 394-95; see infra Section VIII.A.
\item \textsuperscript{24} Treas. Reg. § 25.2503-3(a) (as amended in 1983).
\item \textsuperscript{25} One exception to this rule are gifts made in trust for the benefit of a minor that meet the requirements of I.R.C. § 2503(c). Under I.R.C. § 2503(c), transfers made in trust will be considered present interests, and therefore subject to the annual exclusion, so long as: (1) the trust is limited to a single beneficiary under the age of twenty-one, and the trustee must be authorized to use the trust income and principal for the benefit of the minor beneficiary during his minority; (2) upon reaching age twenty-one, any portion of the trust corpus not expended during the beneficiary’s minority must be distributed to the beneficiary in full upon turning twenty-one; and (3) if the minor dies before reaching age twenty-one, the trust property is distributable to the minor beneficiary’s estate, or to whomever he appoints the property to under a general power of appointment. I.R.C. § 2503(c) (2002); Treas. Reg. § 25.2503-4 (a) (1958). Although I.R.C. § 2503(c) would seem to alleviate the need for Crummey powers, the transaction costs associated with creating a 2503(c) trust (e.g., separate trusts for each beneficiary), coupled with a beneficiary’s right to absolute control over the entire trust corpus upon turning twenty-one, renders use of 2503(c) impractical when viewed in terms of the estate-planning goals of Crummey grantors. See Smith, \textit{supra} note 4, at 391 (“Because the donor can control how property transferred into trust is used by selecting the trustee and trust terms, [Crummey] withdrawal rights encourage use of the annual exclusion to make transfers of property in a way that does not really give the donee control.”).\
\end{itemize}
generally be treated as a future interest for the purpose of qualifying for the annual exclusion.\textsuperscript{26} Beyond the two previous examples, what will and will not qualify as a present interest for the exclusion is often determined by the amount of control that a donee exercises over the transferred property.

In \textit{Fondren v. Commissioner},\textsuperscript{27} the United States Supreme Court established that determination of a present interest for the annual exclusion is:

\begin{quote}
[A] question . . . of time, not when title vests, but when enjoyment begins. Whatever puts the barrier of a substantial period between the will of the beneficiary or donee now to enjoy what has been given him . . . that enjoyment makes the gift one of a future interest within the meaning of the regulation.\textsuperscript{28}
\end{quote}

In contrast, the Seventh Circuit in \textit{Kieckhefer v. Commissioner},\textsuperscript{29} left no room for discretion when it established that it is not the actual enjoyment of an interest “which marks the dividing line between a present and future interest, but it is the right conferred upon the beneficiary to such use, possession or enjoyment.”\textsuperscript{30}

The best way to illustrate the subtle distinctions that often separate present from future interests is through example. If a beneficiary were to receive an annual income interest from the principal of a trust, it will likely be considered a present interest (and therefore qualify as a valid 2503(b) exclusion) under both \textit{Fondren} and \textit{Kieckhefer} because the beneficiary’s right to demand immediate payment of that income interest is absolute. However, in an extreme example, if the same beneficiary was held hostage in a Chinese prison camp, then it can be said that a “barrier of a substantial period between the will of the beneficiary” and his enjoyment of that income interest\textsuperscript{31} has been created by way of the beneficiary’s unfortunate incarceration. Under \textit{Fondren}, the once-present interest must now be characterized as a future interest, and thus is not subject to the annual exclusion.\textsuperscript{32} Nevertheless, according to \textit{Kieckhefer}, this unlikely scenario does nothing to the actual “right” conferred upon the beneficiary to use, possess, or enjoy the income interest (aside from the fact that it is difficult to withdraw money from a trust from within a bamboo cage on the other side of the globe). In turn, so long as that “right” stays intact, the transfer will remain a present interest and

\textsuperscript{26} See Treas. Reg. § 25.2503-3(c), ex.2 (as amended 1983); Phillips v. Comm’r, 12 T.C. 216 (1949); Rev. Rul. 79-47, 1979-1 C.B. 312.
\textsuperscript{27} 324 U.S. 18, 20-21 (1945).
\textsuperscript{28} Id. (citation omitted).
\textsuperscript{29} 189 F.2d 118, 121 (1951).
\textsuperscript{30} Id.
\textsuperscript{31} See \textit{Fondren}, 324 U.S. at 20-21.
\textsuperscript{32} See id.
Most taxpayers do not want to have to make outright gifts in order to qualify for the annual exclusion. Because minors are the most likely target for lifetime gifting, it is understandable that the donor in such a situation would want to maintain some control over the property to ensure that the funds are not wasted on youthful eccentricities. For this reason, gifts in trust are extremely useful to donors, in that they allow the donor to retain some control (via a dependable trustee and good tax advisor) over how the funds are managed and subsequently released.

Yet despite this apparent advantage to giving gifts in trust, a twofold problem remains: First, gifts in trust are almost always considered transfers of future interest and therefore fail to qualify for the annual exclusion; and second, because these gifts are considered future interests, the funds will be subject to the gift tax when distributed to the trust. However, if a taxpayer is somehow able to characterize a gift in trust as a present interest, the taxpayer can take advantage of the annual exclusion, avoid both estate and gift taxation on the amount transferred, and maintain control over how the property is managed. But how, precisely, can this be done? The answer lies in **Crummey**.

**IV. HOW DOES A CRUMMEY POWER WORK?**

As mentioned earlier, a Crummey trust allows a taxpayer to maintain control over a gift (through a retention of power granted in the trust instrument itself), while qualifying the gift as a present interest. But how is this done

---

33. Fondren and Kieckhefer represent only some of the subtle judicial distinctions between present and future interests.
34. Oftentimes, individuals will gift away most of their estates during their lifetime in an attempt to avoid the penal nature of the estate tax. See I.R.C. § 2001(c)(1) (2002).
35. In order for a gift in trust to be excluded from a donor's estate at death, the gift must be deemed "complete." See Treas. Reg. § 25.2511-2(b) (as amended in 1999). A completed gift is one where the donor has relinquished all dominion and control over the property transferred. See id. If the donor retains any interest whatsoever (no matter how small), that transfer will be considered incomplete, and the property will be included in the donor's estate upon death. Helvering v. Hutchings, 312 U.S. 393, 396 (1941); Rev. Rul. 67-396, 1967-2 C.B. 351, clarified by Rev. Rul. 84-25, 1984-1 C.B. 191.
36. This is still subject to the $11,000 cap on the annual exclusion in section 2503(b), as well as the “five or five” limitation under I.R.C. §§ 2041(b)(2), 2514(c). Under I.R.C. § 2041(b)(2), a beneficiary, upon lapse, is deemed to have released the general power of appointment granted to him via the Crummey power on any withdrawal amount greater than $5000 or five percent of the trust corpus. I.R.C. § 2041(b)(2) (2002).
37. See supra Section III.
38. There are many different ways that a donor may retain an interest over the trust corpus. For instance, a donor may preserve the power to add and remove beneficiaries of a trust at will or maintain power over the management of the funds kept in trust. Ultimately, the methods by which a donor may retain control over a Crummey trust, and still qualify for the annual exclusion, are virtually limitless.
interest, thereby allowing the taxpayer to utilize the annual exclusion. To accomplish this estate planning “two-step,” the donor establishes a trust (usually in the name of the intended beneficiaries); and in each year that the exclusion is sought, confers upon the donee a lapsing power to withdraw a specified amount (usually up to either the “5 or 5” limitation, or the annual exclusion amount—whichever is greater). Simply granting the beneficiary the lapsing right to withdraw only gets the donor to the dance floor. To perform the “two-step” properly, the beneficiary must allow the power to lapse, thereby allowing the gift to revert to the trust and the donor to take the annual exclusion. If, however, the donee were to exercise his/her power to withdraw, the music stops and the purpose of the transaction is thwarted because the funds do not revert to the trust, but rather into the immediate possession of the beneficiary. Although the donor in such a situation does not lose the ability to claim the annual exclusion, any significant control over the funds is lost to “youthful eccentricities.”

So, how does a donor ensure that the withdrawal right is not exercised? An easy answer is to simply not inform the potential beneficiary that they have a right to begin with. After all, you cannot demand what you do not know exists. Another potential method to ensure non-exercise is to make the demand period so unreasonably short that a beneficiary could not exercise his right even if he wanted to. To combat these types of abuse, the IRS, in numerous post-Crummey private letter rulings, established that a beneficiary must be given proper, written notice and a reasonable opportunity to exercise his demand right.

Even with proper notice and opportunity to exercise, donors may nonetheless control the situation through an agreement, either express or implied, that makes certain the beneficiary will not exercise his demand power. Within the context of these “lapse agreements” lies the crux of the Crummey problem from the viewpoint of both the

---

39. See supra Section III.
40. Prior to Revenue Ruling 81-7, it was common for a beneficiary to never receive notice of his right to withdraw. See Crummey v. Comm’r, 397 F.2d 82, 88 (9th Cir. 1968) (admitting that some, if not all, of the beneficiaries of the trust had no idea that they had a demand power). Cf. Rev. Rul. 81-7, 1981-1 C.B. 474.
41. See Rev. Rul. 81-7, 1981-1 C.B. 474 (establishing a minimum period for a demand right to remain available to a beneficiary in order for the “transfer” amount to be excluded from the grantor’s estate under I.R.C. § 2503(b)).
42. Id. In Revenue Ruling 81-7, the IRS addressed a trust that granted an adult beneficiary a demand right scheduled to lapse on December 31. Id. However, the donor did not make the actual gift until two days prior to the lapse (December 29) and failed to notify the beneficiary regarding the existence of the power until after the demand period had expired. Id. As a result, the IRS deemed the power to be “illusory,” and refused to recognize the transfer as a present interest and the annual exclusion was denied. See id.
If the donor and beneficiary have a prearranged agreement to allow the demand power to lapse, then the power is “illusory” and the annual exclusion should be denied. However, since these arrangements are informal, and often implied, their existence is very difficult to prove with any degree of objectivity.

V. PRE-CRUMMEY PRECEDENT—NON-LAPSING DEMAND POWERS

One of the most significant and oft-cited cases dealing with demand powers and the annual exclusion is the United States Supreme Court decision in Fondren v. Commissioner. In Fondren, the Court was faced with the question of whether taxpayers who had established seven separate trusts for their seven infant grandchildren, could exclude the first $5000 of contributions made to each trust in the years 1935, 1936, and 1937. To the extent that no present interest is given to a beneficiary, gifts to an irrevocable trust are future interests, and thus cannot be excluded. However, the taxpayers in Fondren argued that the withdrawal power given to the trustee to distribute funds from the trust corpus to provide for the support, education, and maintenance of the beneficiaries on the basis of need was sufficient to establish that the beneficiaries had, at the moment of each gift, a present right of enjoyment.

The Fondren Court acknowledged that the exclusion cannot apply simply because the “donee has vested rights.” Rather, the donee must receive a “right to [a] substantial present economic benefit,” as determined by the present right to “use, possess or enjoy the property,” in order for the exclusion to apply. Furthermore, the Court took special note of the language in the trust instrument which explicitly stated that the donor did not foresee the trustee having to invade the income or corpus of the trust because the beneficiaries already had “adequate and sufficient means of support.” The Court recognized that the “circumstances surrounding the donors and the donees . . . made enjoyment [of the gifts] contingent upon the occurrence of future events, not only uncertain, but by the recitals of the instrument itself improbable of occurrence.” Consequently, the Court in Fondren found that the

---

44. See infra note 143.
46. See infra Sections VIII.E, IX.D.
47. 324 U.S. 18 (1945).
48. Id. at 19-20.
49. Id. at 22-23.
50. Id. at 24.
51. Id. at 20.
52. Id.
53. Id. at 23.
54. Id. at 24.
circumstances surrounding the creation of the trust, coupled with the language of the trust itself, ultimately rendered the withdrawal power unlikely to be exercised. Thus the gifts were considered future interests and, therefore, non-excludable.55

Despite the Supreme Court’s recognition that surrounding circumstances were to be considered in determining the probability of withdrawal,56 the circuit courts remained split on whether to accept such a test. For instance, the Seventh Circuit in Kieckhefer v. Commissioner57 recognized that the court should not look beyond the “explicit terms of the trust”58 to determine whether there exist sufficient “conditions and restrictions”59 to establish that the donee had not received a present interest gift. The court focused upon the distinction between restrictions and contingencies that are imposed by the donor via the trust itself, and those that result from the incapacity of the beneficiary as a minor.60 Ultimately, the Kieckhefer court found that if the restrictions that prevent withdrawal are imposed by law (as is the case with minority), that is not enough to transform a present interest into a future interest. Out of the same concern for reliability and administrative convenience, many courts, including the Ninth Circuit in Crummey,61 used a similarly simple and predictable Kieckhefer-like approach to deal with the muddled distinction between present and future interests as applied to donee withdrawal rights.62

55. Id. at 29. In a related case, Commissioner v. Disston, 325 U.S. 442 (1945), the Court was faced with a trust problem similar to Fondren and subsequently decided the case in favor of the Service using the same surrounding circumstances test used in Fondren. See Fondren, 324 U.S. at 24.

56. See Fondren, 324 U.S. at 24.

57. 189 F.2d 118 (7th Cir. 1951). The taxpayer in Kieckhefer established a trust for the benefit of his grandson (a minor) and named his son (the father of the beneficiary) as trustee. The trust instrument provided that the beneficiary could demand any or all of the trust estate at any time through his father (the trustee) or a legal guardian. The taxpayer sought to claim an exclusion for the initial gift made to the trust. The case ultimately hinged upon two questions: First, could the beneficiary (an infant at the time the trust was created) really make an effective demand? Second, will the fact that no legal guardian was appointed to the child at the time of the trust effectively hinder the beneficiaries’ ability to make a demand so that the transfers will be considered future interests?

58. Id. at 120.

59. Id.

60. The court noted that the Service’s argument—the fact that no guardian had yet been appointed who could transform the gift into a future interest—was fallacious due to “the [Service’s] failure to distinguish between restrictions and contingencies imposed by the donor (in this case the trust instrument), and such restrictions and contingencies as are due to disabilities always incident to and associated with minors and other incompetents.” Id. at 122.

61. See Crummey v. Comm’r, 397 F.2d 82 (9th Cir. 1968).

However, not all courts were willing to sacrifice reality for "simplicity and predictability." Most notably, the Second Circuit in *Stifel v. Commissioner* closely followed the approach taken by the Supreme Court in *Fondren* and *Commissioner v. Disston*, noting that "the Court, in reaching its determinations, did not irrevocably lock itself inside the 'four corners' of the writings [of the trust] but held that the key might lie outside." The court clairvoyantly warned that if the surrounding circumstances were not considered, "a donor could make gifts which on paper were 100% present but in practice were 100% future." In turn, the *Stifel* court found that because the minor beneficiaries of the trust had not been appointed guardians through whom they could legally exercise their withdrawal rights, their power to make a demand was impractical, thereby making the withdrawal rights future interests, rather than excludible present interests.

VI. **CRUMMEY V. COMMISSIONER—THE LAPSING DEMAND POWER**

The most significant of the circuit court decisions dealing with characterization of demand powers for the purpose of annual exclusion was handed down by the Ninth Circuit in *Crummey v. Commissioner*. The dispute in *Crummey* revolved around the creation of a trust instrument that provided that each of the four beneficiaries could demand, at any time up until December 31st of that year, up to $4,000, or the amount transferred in to trust in that year, whichever was less. What made *Crummey* different from the cases previously

---

63. Id. at 98 (quoting Jeffery G. Sherman, *'Tis a Gift to be Simple: The Need for a New Definition of "Future Interests" for Gift Tax Purposes*, 55 U. CIN. L. REV. 585, 589 (1987)).
64. 197 F.2d 107 (2d Cir. 1952).
66. 325 U.S. 442 (1945).
67. *Stifel*, 197 F.2d at 110.
68. Id.
69. Id. at 110-11. The court, in dicta, noted that "it would then seem to be proper to consider the actual facts as to the father's influence on the guardian appointed" to determine whether the child was given a legitimate right to exercise their demand power or not. *Id.* at 110 n.5. One can logically infer that if a parent allows their power to lapse, inevitably they will do the same for their child. Thus, the benefit that is said to inure to the child is never actually realized by the child because the decision to exercise has been predetermined by their parent's decision to allow lapse to occur. If you grant an exclusion based upon the demand power given to these children, that right merely serves as a vehicle for multiplying the number of annual exclusions given to the parents because the powers are "conferred" upon the parent, not only for their own share, but for the share of their children as well. In turn, donors are thus able to mitigate the potential for any "strays" that may exercise their powers, because the Crummeys powers are essentially consolidated into the hands of a few. See infra Section VIII.E.
70. 397 F.2d 82 (9th Cir. 1968).
71. At the time the trust was created, the ages of the four beneficiaries were twenty-two, twenty, fifteen, and eleven years-old.
72. *Id.* at 83.
discussed\textsuperscript{73} was that the demand power granted to the beneficiaries did not continue for the life of the trust, but rather, lapsed at the end of each year it was given. Thus, by virtue of the trust language that afforded the beneficiaries the right to compel immediate distribution of the trust funds within a specified period of time, the taxpayer in \textit{Crummey} claimed that under \textit{Kieckhefer}\textsuperscript{74} the transfers were gifts of present interest, and therefore qualified for the annual exclusion.\textsuperscript{75} The IRS countered that the demand rights granted to the minor children, in light of the surrounding circumstances,\textsuperscript{76} were not likely to be exercised, and thus the transfer could not constitute a present interest.\textsuperscript{77}

In wrestling with how to characterize a lapsing demand power, the Ninth Circuit in \textit{Crummey} opted for a middle ground between the \textit{Kieckhefer} approach urged by the taxpayer, and the \textit{Stifel} test argued by the Service. Closely following the 1956 Tax Court decision of \textit{Perkins v. Commissioner},\textsuperscript{78} the Ninth Circuit determined that “all that is necessary [for the exclusion to apply] is to find that the demand could not be [legally] resisted.”\textsuperscript{79} Using this “right to enjoy”\textsuperscript{80} criterion, the \textit{Crummey} court found that the lapsing demand powers given to the beneficiaries in 1962 and 1963, although never exercised, “could not be [legally] resisted”\textsuperscript{81} and thus the taxpayers were entitled to the annual exclusion for the gifts made to the four beneficiaries in both years at issue.\textsuperscript{82} In its reasoning, the \textit{Crummey} court rejected the approach taken by the IRS (relying on \textit{Stifel}\textsuperscript{83}) and established that whether or not a guardian had been expressly appointed by the donor has no relevance to the application of the annual exclusion, so long as the “demand power was valid and enforceable.”\textsuperscript{84}

\textsuperscript{73.} See supra Section V. 
\textsuperscript{74.} See \textit{Kieckhefer v. Comm’r}, 189 F.2d 118 (7th Cir. 1951) (holding that so long as the explicit terms of the trust afford a beneficiary a present right to receive income from a trust, that alone is enough to establish a present interest gift). 
\textsuperscript{75.} \textit{Crummey}, 397 F.2d at 84-85. 
\textsuperscript{76.} The IRS noted that no guardian had been appointed to make the demand for the minor beneficiaries thereby transforming the demand power into a non-excludible gift of future interest. \textit{Id.} 
\textsuperscript{77.} The IRS relied heavily upon \textit{Stifel} to make this point. See \textit{Stifel v. Comm’r}, 197 F.2d 107 (2d Cir 1952). 
\textsuperscript{78.} 27 T.C. 601 (1956). 
\textsuperscript{79.} \textit{Crummey}, 397 F.2d at 88. 
\textsuperscript{80.} See \textit{Gilmore v. Comm’r}, 213 F.2d 520 (6th Cir. 1954). 
\textsuperscript{81.} \textit{Crummey}, 397 F.2d at 88. 
\textsuperscript{82.} \textit{Id.} 
\textsuperscript{83.} See \textit{Stifel v. Comm’r}, 197 F.2d 107 (2d Cir. 1952). 
\textsuperscript{84.} Bradley E.S. Fogel, \textit{The Emperor Does Not Need New Clothes—The Expanding Use of “Naked” Crummey Withdrawal Powers to Obtain Federal Gift Tax Annual Exclusions}, 73 TUL. L. REV. 555, 573 (1998). Moreover, the Ninth Circuit tacitly repudiated the \textit{Kieckhefer} test by declining to acknowledge that postponed enjoyment due only to the minority of a beneficiary will nevertheless be considered a present interest. See \textit{Crummey}, 397 F.2d at 88; \textit{Kieckhefer v. Comm’r}, 189 F.2d 118, 119 (7th Cir. 1951).
The \textit{Crummey} court expressly declined to even entertain the IRS argument that the likelihood of the beneficiary’s demand should be considered when determining whether the transfer is a present or future interest. In doing so, the court noted the inconsistency and unfairness to the taxpayer that could result from the IRS implementing such an approach due to the arbitrary nature of trying to interpret the subjective intent behind each decision to allow a demand power to lapse. Ultimately, the court opted for the lesser of two evils, and determined that predictability for all taxpayers was of greater importance than the risk of occasional estate tax avoidance.

A. \textit{Public Policy Rationale for Crummey Decision}

Despite what has become of \textit{Crummey} today, there nevertheless exists a sound policy rationale for the Ninth Circuit’s decision. The desire to provide for one’s offspring in the future is a time-honored and sacred human instinct throughout all levels of society. Often, the estate and gift taxes present a significant (but necessary) roadblock to an individual’s realization of this cherished human goal. In some cases, the estate tax can mean the difference between truly providing for one’s offspring in the future, and merely supplying a temporary windfall. With that in mind, the Ninth Circuit sought to level the playing field between taxpayers who did not have secondary concerns over their beneficiaries (i.e., beneficiaries who can be trusted to control the funds given to them), and taxpayers with beneficiaries who cannot be trusted to effectively manage the gift in the same manner as beneficiaries without secondary concerns. By allowing donors to

85. \textit{Crummey}, 397 F.2d at 87. The court went so far as to acknowledge that some, if not all, of the beneficiaries had no idea that they had the right to demand funds from the trust, and therefore it was “highly unlikely that a demand [would] ever be made” at all. \textit{Id.}
86. \textit{Id.} at 88.
87. The use of Crummey trusts as an estate planning tool has become a firmly entrenched in the tax planning arena largely because the IRS has failed to wage a significant attack upon the abusive uses of Crummey. See generally Fogel, supra note 84 (arguing that the Service’s treatment of Crummey has been inconsistent and ineffective largely because of their insistence on taking a piecemeal approach of attacking only those decisions that expanded Crummey powers, rather than attacking the use of lapsing withdrawal powers on the basis of the Crummey decision itself).
88. The term “secondary concerns” is generally meant to encompass those situations where the donor is unable to freely make outright gifts to the beneficiary due to the fact that the beneficiary cannot be trusted to effectively manage the funds by virtue of the beneficiary’s minority or other incapacity. On the other hand, those donors who wish to provide for beneficiaries without secondary concerns are free to transfer their estate to their intended beneficiary via outright gift and qualify for the annual exclusion, because those beneficiaries can be trusted to look after their own best interest in managing the funds.
89. Prior to \textit{Crummey}, a taxpayer with minor children had only three routes to take: (1) make a complete gift in trust for the beneficiary (in which case the donor must relinquish total control over management of the funds to qualify for the annual exclusion—a risky proposition when your child’s future well-being is at stake); (2) make an incomplete
utilize the lapsing demand power to establish a present interest gift, the Crummey court remedied the inequity faced by taxpayers with young children. To take advantage of the annual exclusion prior to Crummey, these taxpayers had to either relinquish control over the funds meant to provide for their children’s future support, or wait until those children were mature enough to manage those funds on their own. Thus, after Crummey, taxpayers were finally afforded a mechanism by which they could make a gift to trust for the benefit of a minor child, maintain some control over the funds in trust (through the language in the trust instrument), and still claim the transfer under the annual exclusion.

When framed in this manner, Crummey can certainly be seen as having served an important and justifiable policy goal. However, with the benefit of hindsight, it is now fairly obvious that the court in Crummey may have misconstrued the vastness of the beast it had unleashed. Although the policy behind Crummey was sound, the mechanism left behind was broad and pliable, thereby affording tax planners a significant amount of “wiggle room” with which the estate and gift tax could be avoided by means never envisioned by the Ninth Circuit. As a result, the Crummey power has become an effective estate-planning tool that affords taxpayers the ability to legally and systematically drain the value of their estates while side-stepping both estate and gift tax consequences on the amount transferred.

B. Naked and Semi-Naked Crummey Powers

The discussion thus far has operated under the assumption that the beneficiary of a withdrawal right maintains an interest in that right, even after lapse, because they have a vested interest in the trust corpus to which the demand right has reverted as a result of the lapse. Thus, one can logically presume that a beneficiary may refuse to exercise the demand power and take the money immediately

---

90. Leaving such a responsibility in the hands of another (even if a trusted fiduciary with strict guidelines) is a “dicey” proposition for any parent. In turn, it can be inferred that the Ninth Circuit’s purpose was to limit its remedy to the problem before it, and not to create a mechanism by which individuals could utilize the annual exclusion through lapsing demand powers given to beneficiaries with a remote contingent interest in the trust corpus. See infra Section VII.

Thus, the Ninth Circuit in Crummey implicitly recognized the need for parents to maintain control over the funds to be used for their children’s future, especially in those cases where the children are relatively young and cannot be expected to manage the funds in a responsible manner.
knowing that the gift is not lost, but postponed until the trust is payable to him.\textsuperscript{91}

However, the same logic does not follow when the beneficiary of a demand power does not have any interest in the trust outside of the Crummey power, but nevertheless allows the withdrawal right to lapse. In such a situation, the beneficiary retains what is commonly referred to as a “naked” Crummey power,\textsuperscript{92} that is, a lapsing demand right without any corresponding interest in the corpus of the trust. With naked Crummey powers, the beneficiary has no conceivable economic incentive to allow a lapse to occur because once the demand power has lapsed, the income from that demand right is gone forever.\textsuperscript{93} Given this, it logically follows that the beneficiary of a naked Crummey power will not allow the lapse to occur, and thus the transfer will amount to nothing more than an outright gift, thereby affording the grantor with none of the estate-planning benefits normally associated with a Crummey power.\textsuperscript{94} Yet, if a grantor could somehow ensure that a naked Crummey beneficiary will not exercise their demand right, then it follows that the number of annual exclusions that can be claimed under naked Crummey powers is limited only by the number of potential beneficiaries who can be entrusted to let the power lapse.\textsuperscript{95} Without any limitation regarding the beneficiary's interest, or lack thereof, in the trust corpus, donors can merely pick and choose as many beneficiaries as necessary to reduce their taxable estate to zero. In addition, if notice and opportunity to exercise were not required (as was the case prior to Revenue Ruling 81-7) then a donor could simply open the phone book and transfer their entire estate into trust in one year without any gift tax consequences whatsoever. For example, assume a donor with an estate worth $11,000,000 has set up a trust with his only heir as the sole beneficiary. Without notice and opportunity to exercise, the donor can simply name as many naked Crummey power holders as it takes to relieve his estate from any estate tax consequences at his death. Any concern that the donor may have over potential exercise is remedied by the fact that the power holders cannot exercise what they don’t know exists (because they never received notice). In this particular situation, the donor would simply divide $11,000 (the annual exclusion maximum) by the net worth of his estate (minus the unified credit

\textsuperscript{91}. In such a situation, a beneficiary will likely have an economically rational motive for allowing his power to lapse. For instance, a beneficiary may, and often will, allow lapse simply because the trustee will better manage the trust property, thereby assuring the beneficiary a long-term source of future income.

\textsuperscript{92}. See generally Fogel, supra note 84. The IRS has generally failed in its attempts to argue that naked Crummey powers are illusory. See id.

\textsuperscript{93}. See Smith, supra note 4, at 390-91.

\textsuperscript{94}. See supra Section IV.

\textsuperscript{95}. See Fogel, supra note 84, at 583.
amount available to the donor) and the number he is left with is the amount of naked Crummey power holders he must name to siphon off his entire taxable estate in one year. Thus, if the donor were to die the following year, his entire estate will have passed to his sole heir free of any estate tax implications. Ultimately, the most telling indicator of whether or not an agreement for such assurances has been made is the economic motivation of the beneficiary in allowing the Crummey power to lapse. As explained earlier, basic economic rationality establishes that naked Crummey beneficiaries will not let their power of withdraw lapse, unless there exists some other motivating factor external to economic considerations.

As is the case with most areas of the law, there is a gray area between beneficiaries with an absolute, vested interest in the trust corpus, and naked beneficiaries without an interest in the trust whatsoever. Semi-naked Crummey beneficiaries are those individuals whose interest in the trust corpus, external to their Crummey power, is contingent upon the occurrence of some uncertain future event. Because the exact interest that semi-naked Crummey beneficiaries hold in the trust corpus is difficult to determine, so too is their motive for lapse. Unlike naked Crummey beneficiaries, whose motives to allow lapse are presumably non-economic, semi-naked beneficiaries are often motivated by a mixed bag of economic and non-economic concerns (which are ultimately dependant upon surrounding circumstances) and thus a presumption cannot be so easily reached with regard to their decision to lapse. Whether or not semi-naked beneficiaries decision to lapse is economically motivated, and therefore viable, is ultimately dependent upon the likelihood of the contingency occurring.

96. Assuming the donor in this example has not made any prior gifts against his unified credit during his lifetime and has no other credits or deductions available to him, he will avoid a total estate tax liability of $4,700,800 by engaging in the naked Crummey transaction.

97. External factors can include, but are certainly not limited to: the desire to maintain good favor with the donor to protect the donee’s separate interests in the donor’s estate, apart from the trust or demand power; intra-family pressure; and prearranged agreements between the donor and donee. See Tech. Adv. Mem. 87-27-003 (Mar. 16, 1987).

98. Although the Ninth Circuit in Crummey refused to engage in such a likelihood analysis, other courts have been willing to engage in a determination of the probability of contingency. Most notably, the Supreme Court in Commissioner v. Disston, 325 U.S. 442 (1945), addressed this very issue when the taxpayer in that case created a trust equally divided for the benefit of each of his five children, including his nineteen year-old son, who was given a one-half interest in the corpus of the trust. The nineteen-year-old’s interest, however, was held in trust until he reached forty-five, at which time he was to receive his entire one-half share. The trust instrument itself failed to provide for a beneficiary of the one-half share should the nineteen year-old fail to reach the age of forty-five. Thus, the Court concluded that it was too tenuous to assume that the nineteen year-old would live to forty-five, and as such, the gifts that comprised that half of the corpus were gifts of future interest and therefore did not qualify for the annual exclusion. Moreover, the trust failed to provide for the future beneficiary of the second half of the trust corpus. The taxpayer argued that because that portion of the trust could be invaded (up to ten percent of the cor-
For example, a common Crummey trust scenario will provide lapsing demand powers for the grantor’s children and grandchildren. Often, the children of the grantor are the sole primary beneficiaries of the trust, leaving the grandchildren with only a remainder interest contingent upon the death of their parent beneficiary before a certain age. Moreover, even in cases where the contingency is remote, at best, these transfers are made between family members with significant economic interests external to the lapsing withdrawal power and the trust corpus, thereby making maintenance of good favor with the donor a much more valuable commodity than the relatively small $11,000 annual gift. In these situations there often exists an implied “you scratch my back, I'll scratch yours” understanding between the grantor and beneficiary where the grantor gets the present benefit of the annual exclusion through the lapse, and the beneficiary maintains their future interest in the grantor’s estate by allowing the lapse. From an economic rationality standpoint, because the contingency upon which these grandchildren’s interest relies is unlikely to occur, the grandchild, like their naked Crummey power counterparts, should exercise their demand power due to the fact that they will likely never see that money again. Assuming that a donor has five grandchildren, the net estate of that donor is decreased by $55,000 annually without having to give away anything “real.” On its face, the $55,000 a year seems a bit paltry in the grand scheme of things. However, if the number of beneficiaries with semi-naked Crummey demand rights are increased to twenty or thirty, and then multiplied over twenty-five years; what at first blush seemed to be an insignificant loss for the fisc, suddenly becomes quite a hemorrhage in estate tax revenue.

VII. CRISTOFANI V. COMMISSIONER: ABUSE OF THE ANNUAL EXCLUSION

As discussed earlier, the Crummey court left tax planners with broad latitude to use the annual exclusion as a device to avoid the estate tax. As such, tax planners did not disappoint, and continued to
push the limits of Crummey\textsuperscript{102} beyond its originally intended parameters.\textsuperscript{103} The most significant Tax Court case to deal with the expansion of the Crummey precedent\textsuperscript{104} was Cristofani v. Commissioner.\textsuperscript{105} The controversy in Cristofani\textsuperscript{106} was centered around annual exclusions claimed by Maria Cristofani for gifts made to her two children and five grandchildren (all minors) in 1984 and 1985. As is the case in a Crummey trust situation, the two children and five grandchildren were each given a lapsing power to demand up to $10,000 of the trust income in the year it was given.\textsuperscript{107} In Cristofani, however, only the two children of the grantor were named as primary beneficiaries of the trust.\textsuperscript{108} The five grandchildren, on the other hand, retained a future interest in the trust only if their parent failed to outlive the donor by one-hundred-twenty days.\textsuperscript{109} Due to the tenuous nature of the Cristofani grandchildren’s contingent interest,\textsuperscript{110} the IRS refused to recognize the transfers made to them as gifts of present interest and denied the annual exclusions for all five grandchildren in both years in controversy.\textsuperscript{111} Relying on the definition of a present interest promulgated in Crummey,\textsuperscript{112} the taxpayer in Cristofani argued that, so long as the grandchildren’s right to demand payment “could not be [legally] resisted,”\textsuperscript{113} the annual exclusion should apply.\textsuperscript{114} The IRS countered by establishing that Crummey was inapplicable to the situation in Cristofani because the beneficiaries in Crummey possessed “substantial future economic benefits’ in the trust corpus and income,” while the grandchildren in Cristofani retained no such future interest because the contingency was unlikely to occur.\textsuperscript{115} Ultimately, the Cristofani court sided with the taxpayers and found that the demand power held by each of the grandchildren was indeed a present interest and therefore subject to the annual exclusion.\textsuperscript{116}

\textsuperscript{102} See Crummey v. Comm’r, 397 F.2d 82 (9th Cir. 1968).
\textsuperscript{103} See supra Section VI.A.
\textsuperscript{104} Crummey, 397 F.2d 82.
\textsuperscript{105} 97 T.C. 74 (1991).
\textsuperscript{106} Id.
\textsuperscript{107} Id. at 75. The beneficiaries were only given fifteen days within which to exercise their demand power. Id. at 76. That is fifteen days less than the amount mandated under the notice requirements in Revenue Ruling 81-7. See supra note 41 and accompanying text.
\textsuperscript{108} Cristofani, 97 T.C. at 74.
\textsuperscript{109} Id.
\textsuperscript{110} The Cristofani court recognized the unlikelihood of the contingency ever being met, noting that both of the donor’s children were in good health upon her death. Id. at 83.
\textsuperscript{111} Id. at 78.
\textsuperscript{112} Crummey v. Comm’r, 397 F.2d 82, 88 (9th Cir. 1968).
\textsuperscript{113} See Cristofani, 97 T.C. at 80 (quoting Crummey, 397 F.2d at 88).
\textsuperscript{114} Id. at 80-81.
\textsuperscript{115} Id. at 82.
\textsuperscript{116} Id. at 84-85.
In reaching its decision, the Cristofani court expanded the reach of Crummey far beyond what was contemplated by the Ninth Circuit, reading Crummey as follows:

We do not believe, however, that Crummey requires that the beneficiaries of a trust must have a vested present interest or vested remainder interest in the trust corpus or income, in order to qualify for the section 2503(b) exclusion.\textsuperscript{117}

Although this seems to be a rather egregious expansion of Crummey, it can nevertheless be argued that the Cristofani court made its decision within the parameters of the procedural structure left by the Ninth Circuit. Crummey involved a trust instrument that provided:

[With the exception of the yearly demand provision, the only way the corpus can ever be tapped by a beneficiary, is through a distribution at the discretion of the trustee.\textsuperscript{118}]

If construed broadly, because the beneficiaries’ interest in the trust corpus was dependent upon the discretion of the trustee, then it would seem that Crummey too involved beneficiaries with contingent interests in the trust corpus, thereby opening the door for courts to uphold Cristofani-like abuses of the annual exclusion. However, to read Crummey in such a broad manner distorts the reality of the circumstances surrounding the beneficiaries’ interest in the trust corpus.

On the one hand the contingency in Crummey was predicated upon the discretion of a trustee, who, as a fiduciary, must give effect to the donor’s will (as outlined in the trust instrument). Given the language in the Crummey trust instrument, coupled with the legal obligation of a fiduciary to carry out the terms of that trust instrument, it was likely, if not certain, that the beneficiaries would receive a share of the trust corpus at some point in the future.\textsuperscript{119} In contrast, the contingency in Cristofani was based upon the occurrence of an uncontrollable outside event (the death of both trustees within 120 days of the death of the decedent) that, given the nature of probabili-

\textsuperscript{117} Id. at 83.
\textsuperscript{118} Crummey, 397 F.2d at 88.
\textsuperscript{119} The trust instrument in Crummey provided that “the trustee was authorized to invade the trust corpus of a beneficiary’s trust ‘up to the whole thereof’ . . . [for the] ‘proper care, maintenance, support and education’ of each beneficiary. Crummey v. Comm’r, 25 T.C.M. (P-H) 144, 146 (1966). Given this language, it is fair to say that the beneficiaries in Crummey retained a vested economic interest in the trust corpus, even after lapse of their demand power. Furthermore, the Crummey trust language comports with the Ninth Circuit’s implicit policy goal in Crummey, that is, to allow grantors the benefit of the annual exclusion when providing for the future support of their minor children. See supra Section VI.A.
ties, would likely never occur. To assume that a contingency based upon the death of two healthy individuals within four months of another is as likely to vest as one based upon a trustee’s discretion (which is guided by specific parameters outlined in a trust instrument) is the logical equivalent of saying that you are as likely to get struck by lightning as you are to stub your toe. Given this distinction, Cristofani improperly applied the Crummey precedent to a class of beneficiaries never contemplated by the Ninth Circuit’s opinion. In turn, Cristofani’s reliance upon the reasoning in Crummey without a corresponding analysis of the probability of vesting in each case was unjustifiable and warranted much greater scrutiny than it received.

More fundamental than the Cristofani court’s failure to recognize the different vesting probabilities was its overly broad reading of the Crummey decision. By focusing its analysis upon the facts presented before it, the Ninth Circuit in Crummey implicitly intended to limit its decision to those facts. The purpose of the Crummey opinion was not to create a new mechanism of estate tax avoidance, but to eliminate an inherent inequity in the tax treatment of gifts to minors and adults. Despite this goal, the Tax Court in Cristofani chose to engage in an analysis and application of the Crummey decision without recognizing the Ninth Circuit’s implicit attempt to control the breadth of its opinion. The Cristofani court in turn interpreted the Ninth Circuit’s silence in the broadest sense possible, finding that Crummey did not require “that the beneficiaries of a trust must have a vested present interest or vested remainder interest in the trust corpus or income, in order to qualify for the 2503(b) exclusion.”

120. The Cristofani court went so far as to recognize that each of the primary beneficiaries had, and at all times prior to and after the death of the decedent, been in perfect health. Cristofani, 97 T.C. at 84-85.

121. Despite this obvious factual inconsistency, the Cristofani court, relying heavily upon the reasoning in Crummey, flatly rejected any consideration of the probability of future vesting. Rather, the court focused solely upon the beneficiary’s legal right to exercise their demand power as the only relevant factor in determining whether a transfer of present interest was made. (In fact, the Cristofani court even went so far as to recognize the fact that each of the primary beneficiaries had, and at all times prior to and after the death of the decedent, been in perfect health.) Such a simplistic test may have been warranted regarding the rather benign contingency in Crummey, but the court’s application of the same test, without more, to the contingency interests in Cristofani represents a complete disregard for the fairly blatant factual distinctions between Crummey and Cristofani.

122. The Ninth Circuit rejected the opportunity to establish a bright line rule (taken from a lower court decision), believing that it did not fit the facts in its case. See Crummey, 397 F.2d at 88 (“In another case we might follow the broader Kieckhefer rule since it seems least arbitrary and establishes a clear standard.”). Moreover, throughout the Crummey opinion, the Ninth Circuit made constant references to “our” case, implicitly limiting the scope of its holding to the specific facts presented before the court. Id.

123. Id.

124. See supra note 89.

125. Cristofani, 97 T.C. at 83.
By turning a blind eye to the obvious circumstantial distinctions between its case and *Crummey*, as well as the implicit limits placed upon the *Crummey* precedent by the Ninth Circuit, the *Cristofani* decision expanded the class of eligible Crummey beneficiaries far beyond what was contemplated by the court in *Crummey*. Affording a grantor the opportunity to take advantage of the annual exclusion in a situation where the beneficiary of the Crummey power has little or no opportunity to receive a share of the trust corpus (other than the annual demand power) is a subversion of both the policy rationale behind *Crummey* and the general purpose of allowing an annual exclusion in the first instance.\(^{127}\)

### VIII. Problems With the Current System

Although thirty-six years of jurisprudence has solidified acceptance of the Crummey power as a viable estate-planning tool, the practical use of the Crummey power is fundamentally at odds with the underlying goals of the tax system as a whole.

#### A. Crummey Powers Are a Subversion of the Annual Exclusion

The first, and most basic flaw recognized by many commentators is that Crummey powers represent a clear subversion of the Congressional purpose behind allowing an annual exclusion. In creating an annual exclusion to the gift tax, Congress intended to only cover small, customary gifts to “obviate the necessity of keeping account” of such gifts for tax purposes.\(^{128}\) When viewed in this context, Crummey powers, even when granted to vested beneficiaries, certainly tend to subvert the spirit behind the exclusion. If the annual exclusion was intended to cover only relatively small, customary gifts one normally provides for friends and family during the course of a year, it is hard to imagine how Crummey powers fall within those parameters.

For instance, picture a suburban Chicago Christmas, and little, four year-old Maggie makes her way down the stairs to find two gifts from her Grandma Willie under the tree: a shiny red bicycle with a bow on it and a wrapped box with a gleaming piece of paper notifying her that she has thirty days to exercise an $11,000 Crummey withdrawal right. Although Maggie is intelligent beyond her years, it is safe to say that she only views one as a “gift.” Why? Because a bike is something that the average person would see as a regular, run of the mill Christmas present. A bike at Christmas is no different than a doll for her birthday, a $100 savings bond for her first communion, or $5,000 for her college graduation. The $11,000 withdrawal right, on

---

126. See supra Section VI.A.
127. See supra Section III.
the other hand, is clearly not considered a “gift” in the same sense by Maggie—or anyone else for that matter. Yes, Maggie has a thirty day “right” to that $11,000, but what is “ordinary” about such a right? Even an $11,000 bicycle at Christmas and a thirty-day lapsing withdrawal right are certainly not “ordinary” in the same sense. Using this reasoning, many argue that the Crummey power is not a “gift” recognized under the annual exclusion, and as such should be considered a gift of future interest.

B. Horizontal Inequity Created Between Grantors Who Use and Do Not Use Crummey Gifts

Maggie’s situation illustrates a related problem with Crummey powers. Assuming that Grandma Willie’s $11,000 “Crummey” Christmas present to Maggie is considered a gift for purposes of the annual exclusion, then the threshold amount for annual gifts allowable under the exclusion is met, thereby rendering any additional gift from Grandma Willie to Maggie during the same year taxable as an inter-vivos transfer. Under this scenario, Grandma Willie would owe gift taxes on the value of the bicycle given to Maggie (in addition to the total value of any other gifts to Maggie made throughout the year for her birthday, Easter, and other holidays or special occasions). Gift taxes are rarely, if ever, paid on the value of incidental gifts given to a beneficiary in excess of the annual exclusion amount used to cover the value of the Crummey withdrawal right. Thus, in an ironic twist of fate, the very same gifts that Congress intended to cover with the annual exclusion are not taxed when, in fact, they should be. This problem of Crummey grantors “double dipping” at the annual exclusion effectively blurs the amount excludable under I.R.C. § 2503(c) into an indeterminate number, amenable to the needs of each individual Crummey grantor.

As a result of “double dipping,” a horizontal inequity problem is created between those who grant lapsing withdrawal rights and still give those beneficiaries “ordinary” gifts (Crummey taxpayers), and those taxpayers that only give “ordinary” gifts (regular, non-Crummey taxpayers). Because Crummey taxpayers can effectively increase the amount of the annual exclusion available to them by “double dipping,” they are placed at a significant tax advantage when

129. After all, if Maggie wanted to be careful and leave her new bike untouched until spring, would her parents take it back after thirty days? Of course not. That is not what a gift is about. On the other hand, if Maggie wanted to wait until March to exercise the withdrawal right she received at Christmas, she would find that the “gift” was removed from her reach until some point in the distant future. When Maggie’s right to the $11,000 is not absolute, how can she say that she has “received” anything?

130. See I.R.C. §§ 2501(a), 2001(c) (2002).
compared to taxpayers who do not utilize a lapsing withdrawal power.

For example, imagine two taxpayers, Arnold and Willis, each with $5,000,000 estates. During the course of a year, Willis gives his granddaughter Kimberly three gifts totaling $16,000. At tax time, Willis claims all three gifts to Kimberly. Because of the annual exclusion, Willis will not be taxed on the first $11,000 of the gifts to Kimberly, but will incur gift tax liability on the $5,000 in excess of the exclusion amount. Not to be outdone, Arnold gives his grandson Abraham incidental gifts (on his birthday, Christmas, Arbor Day, and so forth) totaling $5,000. In addition, Arnold has set up a Crummey trust naming Abraham as sole beneficiary. Each year Arnold grants Abraham an $11,000 withdrawal right that Abraham allows to lapse. When tax time rolls around, Arnold claims the $11,000 lapsing demand power under the annual exclusion, but never claims the $5,000 in incidental gifts to Abraham because, as he says, “How can normal gifts be taxable?” Arnold is partially right. Ordinary gifts are not supposed to be taxable; they are meant to be captured by the annual exclusion. However, when the full annual exclusion amount has already been exhausted by a lapsing demand right to a beneficiary, the value of ordinary gifts to that same beneficiary should theoretically be considered a taxable inter-vivos transfer.\footnote{Id.}

Unfortunately for the fisc, the Arnolds of the world do not count the ordinary gifts made to beneficiaries in excess of the $11,000 claimed for the lapsing demand power; and therein lies the inequity to taxpayers like Willis. While Willis is taxed on the excess $5,000 given to Kimberly, Arnold, who gifted the same total amount to Abraham, is not taxed because incidental gifts are never claimed and, by extension, never taxed. As a result, two similarly situated taxpayers are afforded different tax treatment due to one taxpayer’s utilization of the Crummey power.

C. Horizontal Inequity Between Crummey Grantors With Differing Beneficiary Pools

In addition to the horizontal inequity created by “double dipping,” Crummey powers also create horizontal inequity between similarly situated Crummey grantors. Again, by way of example, assume that two similarly situated taxpayers, Bo and Luke, each want to gift $55,000 into a trust for the benefit of their Uncle Jesse (who cannot be trusted with a demand right) while claiming the entire amount under the annual exclusion. Bo, who has five children, can simply grant each of his children a contingent remainder interest in the
trust and a corresponding demand power to achieve his of annually adding $55,000 into the trust under the I.R.C. § 2503(b) exclusion. Luke, on the other hand, has only two minor children and, thus, can only claim $22,000 under the annual exclusion (with the remaining $33,000 still includible in his estate). As a result, two similar taxpayers with the same goal are treated differently under the Tax Code simply because one individual has more children that he can trust not to exercise the demand power.

D. Parent Beneficiaries Who Act as Guardians for Minor Beneficiaries of the Same Trust

One of the more abusive situations involving Crummey powers occurs when a beneficiary of a Crummey trust simultaneously serves as guardian over his minor child who has a contingent remainder interest in the same trust. In such a situation, the minor beneficiary cannot be expected to make an effective demand, and thus the parent, as guardian, makes the minor child’s decision to lapse for him. The problem in this scenario is that the parent cannot be expected to make an impartial decision for his minor children because the parent retains a vested interest in the trust corpus.

To illustrate, assume that a Crummey grantor, Howard Cunningham, grants a $10,000 lapsing Crummey withdrawal right to his two sons, Richie and Fonzi. In addition, he grants a lapsing withdrawal right to Richie’s two children, Joanie, age four, and Chachi, age six. Under the trust instrument, Richie and Fonzi are the primary beneficiaries, each to receive half of the accumulated trust corpus upon the death of the grantor. Further, if the grantor outlives both primary beneficiaries, Richie and Fonzi, then Joanie and Chachi share in Richie’s half of the corpus and Fonzi’s surviving issues get his share.

In most situations, all four beneficiaries will allow their demand powers to lapse. Richie and Fonzi each make their own decision to lapse, while Joanie and Chachi’s decision is made through their parent and legal guardian, Richie. By making the decision to allow Joanie and Chachi’s Crummey power to lapse, Richie has effectively

132. This hypothetical assumes that both Bo and Luke will only trust members of their immediate family with the demand power. For Luke to accomplish the goal of annually gifting away $55,000 tax-free, he would have to confer a demand power upon an extra-family beneficiary, incurring a greater risk of exercise because an extra-family beneficiary does not have the same type of external pressures to allow lapse that a family member does (i.e., to stay in the good graces of the grantor to maintain a potential inheritance).

133. This hypothetical illustrates yet another problem presented by lapsing demand powers: parental control over a minor child’s Crummey power. See infra Section IX.D.

increased his share of the trust corpus by $10,000.135 Because Richie’s interest in the trust corpus is vested, it is likely that he will receive his share of the corpus at some point in the future. That being the case, economic rationality demands that when an individual is presented with an opportunity to add to his overall wealth, he will exercise the choice that affords him with the greatest increase in that wealth.136 For example, if an individual is presented with three bills on a table ($1, $20, and $100), the economically rational decision is to opt for the $100 bill. In much the same way, when Richie is presented with the opportunity to increase the amount in trust that will be payable to him in the future, economic rationality demands that he exercise that opportunity.

Thus far, the discussion has centered around the economic motivation behind Richie’s decision to lapse Joanie and Chachi’s demand power. What about Joanie and Chachi’s economic motivation? After all, if their economic motivation to lapse is the same as Richie’s, then what harm is done by allowing Richie to do what Joanie and Chachi would have done anyway? Unfortunately, it is not that simple. Remember, Joanie and Chachi are contingent beneficiaries whose interest in the trust corpus is reliant on an outside event (the death of Richie within three months after the death of Howard) that in all likelihood will never occur, thereby making their probability of vesting unlikely at best. Once again, economic rationality tells us that, when placed in the situation of choosing between sure money in the present, or unlikely money in the future, an individual will always choose sure money in the present.137 Therefore, under normal circumstances, Joanie and Chachi will always choose to exercise their de-

135. To further illustrate how Richie’s share is increased by allowing Joanie and Chachi’s Crummey power to lapse, assume that Howard gifts away $10,000 per beneficiary for ten years and subsequently dies after the tenth year. If all four beneficiaries allow their power to lapse, the entire trust corpus would be composed of $400,000 ($200,000 for Richie and $200,000 for Fonzi). However, if Joanie and Chachi were to exercise their demand power in each of those years, then, at the end of the ten years Richie and Fonzi would only be left with $100,000 each. Thus, by allowing Joanie and Chachi’s power to lapse, Richie effectively doubles the amount available for distribution between Fonzi and himself upon Howard’s death.

136. An increase in wealth need not be immediate to be considered economically rational. For instance, a vested beneficiary will often allow a lapse of his Crummey power, and thus forgo the immediate increase in wealth, in order to keep the demand amount in the trust (which will likely produce a greater long-term rate of return on the $11,000 gift than could be achieved by a beneficiary if he were to invest it himself). Thus, the decision to lapse in that situation is economically rational and thus justified. See supra note 91; cf. Fogel, supra note 84, at 606-07 (arguing that the actuarial value of a vested beneficiary’s lapsed demand right will always be less than the present value of the Crummey gift because when the same lapsed demand power is placed in trust, it must be discounted to reflect the possibility that the beneficiary will not reach the age requirement specified in the trust language).

137. Of course, this assumes that there are not any external circumstances that would warrant a lapse.
Having established that economic rationality dictates that Richie will always seek to increase his share in the corpus by allowing Joani and Chachi’s demand power to lapse, and Joani and Chachi will always choose to exercise their demand power (being contingent beneficiaries), how do we reconcile these conflicting desires? Given that Richie’s economic interests are at odds with Joanie and Chachi’s, it simply cannot be said that Richie’s decision to lapse on behalf of Joanie and Chachi is completely objective. Why, then do we allow Richie to make that decision for Joanie and Chachi? The answer for some lies, in part, on the nature of intra-family relationships.

Although it is a maxim of human existence that most every parental decision is made with the child’s best interests in mind; as with everything else in life, there are plenty of exceptions to that rule. As such, a minor child’s best interests are not always represented when a parent beneficiary is the one responsible for making the decision whether to exercise a child’s demand right.

E. Collusive Agreements Between Grantor and Beneficiary Used to Assure Lapse

A related problem with allowing a parent beneficiary to exercise his child’s demand right is that such an arrangement effectively allows a donor to multiply the number of exclusions available to them without increasing the corresponding risk of exercise. In granting Crummey powers to beneficiaries to claim the annual exclusion, the only real risk to any donor is that a beneficiary will exercise their withdrawal right, thereby eliminating the primary advantage of a Crummey trust (i.e., gifts in trust being considered present interest gifts in order to take advantage of the annual exclusion). Thus, above all else, Crummey donors want to be sure that their beneficiaries will not exercise their withdrawal right. Using the same hypothetical above, by granting withdrawal rights and corresponding contingency interests to Joanie and Chachi, Howard effectively doubled the amount he could siphon off from his estate to his children under the annual exclusion. Moreover, because Richie effectively exercises the demand rights for not only himself, but Joanie and Chachi as well, Howard is able to “kill three birds with one stone” and multiply the amount excludible from gift tax without a corresponding risk of exercise.

138. This is only from a purely economic standpoint, without any consideration given to intra-family externalities.

139. See, e.g., Rev. Rul. 87-4 (where a parent, as the sole primary beneficiary of a trust, consistently allowed her minor child’s withdrawal right to lapse and upon the parent’s death, she left nothing to her child (who was an adult by that time)).
Nevertheless, in order to preserve the tax effectiveness of his Crummey gifting plan, Howard must be assured that Richie and Fonzi will allow their Crummey powers to lapse. Otherwise, the gift is never placed in trust, and the transaction ultimately amounts to a roundabout method of gifting outright. As mentioned earlier, because Richie and Fonzi are the primary beneficiaries of the Cunningham Crummey trust, the decision to allow their Crummey powers to lapse has an economically rational justification because their future access to that property is not eliminated by the lapse, but merely postponed.140

However, if the facts in the Cunningham hypothetical are changed, and Fonzi’s interest in the trust principal is made contingent upon his outliving Richie, Joanie, and Chachi by three months, then it follows that Fonzi will not allow his Crummey power to lapse because he will almost certainly never have access to that money again. Given this, how can Howard overcome the economic rationality that demands Fonzi exercise the $11,000 withdrawal right, and thus be assured that Fonzi will allow lapse? The easiest, and thus most often employed, solution to Howard’s dilemma is to simply come to a prearranged “understanding” with Fonzi whereby Fonzi guarantees Howard that he will not exercise his demand right.141 Thus, once he is assured that Fonzi will allow his Crummey power to lapse, Howard can safely convey an additional $11,000 (along with the $33,000 transferred to Richie, Joanie and Chachi) under the annual exclusion into the Cunningham trust meant to benefit only Richie and his heirs.

From the Service’s perspective, the problem with these prearranged agreements is that Howard is afforded the benefit of the annual exclusion for the demand power given to Fonzi when, given the preexisting “understanding” not to exercise, Fonzi was not “really” granted the present right to the $11,000. Because Fonzi’s “use, possession or enjoyment”142 of his right to $11,000 is, for all practical purposes, eliminated by the prior agreement, then the demand right cannot be considered a present interest and the annual exclusion should be denied. Thus, the IRS would contend, the Crummey power given to Fonzi in this situation is illusory and should not qualify for the annual exclusion.

Despite the logic behind the Service’s position, any attempt to prove the existence of such an “understanding” has proven difficult. This problem is exacerbated by the fact that these agreements are of-

140. See supra notes 91 and 136.
141. Often this “understanding” is secured by a beneficiary’s belief that he will be “penalized”—via a reduction or elimination of their interests in future gifts and/or bequests from the grantor—if he exercises his withdrawal right.
ten implied between family members, leaving the government with very little tangible evidence with which to prove its case. As a result, courts have been reluctant to recognize the validity of the “lapse agreement” argument due to the subjectivity involved in making such a determination.143

IX. SUGGESTIONS FOR REFORM

Given the potential for abuse of the annual exclusion established by Cristofani and its progeny, it is relatively clear that a change to the structure of the annual exclusion is necessary if the IRS is to maintain the integrity of the gift tax. Only one question remains: What do we change?

A. Elimination of the Crummey Power

As an answer to Crummey’s subversion of the purpose behind the annual exclusion, some argue that the availability of the annual exclusion to all lapsing demand powers should be eliminated.144 After all, if the donor truly intended to give a beneficiary a present interest in the property subject to the demand power, he would have simply given that beneficiary an outright gift. Moreover, such a solution would subsequently solve the horizontal equity problem faced by our friends Arnold and Willis.145 However, such a remedy undermines the

143. See Holland v. Comm’r, 73 T.C.M. (CCH) 3236 (1997). Despite having found that there was an “informal agreement” regarding the beneficiaries’ decision to lapse their Crummey demand right, the Holland court held that so long as an agreement between the parties is not legally binding, the annual exclusion will not be denied on the basis of a prearranged “understanding” between the parties. Id. at 3237. Given that the vast majority of Crummey powers are conveyed in the intra-family context, any agreement between donor and donee to ensure lapse of that power will often, if not always, be informal in nature. As a result, the Holland court’s insistence on legal enforceability is particularly damaging to the Service’s ability to combat Crummey abuse through their prearranged “understanding” argument.

Cf. Kohlsaat v. Comm’r, 73 T.C.M. (CCH) 2732 (1997). Decided only a few months prior to Holland, the Kohlsaat court took a much more accepting view of the Service’s prearranged “understanding” argument. Although the Kohlsaat court ultimately sided with the taxpayers and refused to infer an agreement from the surrounding circumstances, its decision was not based upon the legal enforceability of the agreement between donor and donee. Rather, the court refused to accept the Service’s prearranged “understanding” argument because the record lacked sufficient evidence to prove the existence of such an agreement. Id. at 2734. Thus, the Kohlsaat decision seems to imply that in cases where there is sufficient evidence of a prearranged “understanding” not to exercise a Crummey demand right, the annual exclusion should be denied, regardless of the legal enforceability of such an “understanding.” See also Trotter v. Comm’r, 82 T.C.M. (CCH) 633, 637 (2001) (finding that the terms of the trust, coupled with the circumstances surrounding its operation, were “supportive of an implied understanding that the withdrawal rights would not be exercised” (emphasis added)).


145. See supra Section VIII.B.
sound policy rationale for allowing donors the ability to take advantage of the annual exclusion through Crummey powers. As discussed earlier in this Comment, the purpose of the Crummey power is to right the inequity between taxpayers who can provide for beneficiaries through annual outright gifts (because those beneficiaries are mature enough to manage the property effectively) and those donors who wish to do the same for beneficiaries who cannot be trusted to manage the gift in an efficient manner. Thus, when limited to those beneficiaries with a vested interest in the trust of equal or greater value than their cumulative demand powers, the Crummey power serves a rational public policy function that should not be eliminated on account of abusive situations that exist beyond that purpose.

B. Change the Annual Exclusion From a Per Donee Model to a Per Donor Model

Another suggestion espoused by many commentators to eliminate abuse of the annual exclusion has been to base the exclusion on a “per donor” model rather than the current “per donee” system. Under the per donor model, rather than cap the exclusion amount on the basis of gifts given to a single donee, the exclusion would be based upon the total amount gifted by a donor in a given year. Although the overall amount of the exclusion would likely increase, this solution would nevertheless create an absolute threshold on annual gifts and thus place a significant limit on potential abuse of the exclusion. Assuming that the annual exclusion is bumped up to $50,000 per donor, the horizontal inequity between our friends Bo and Luke would effectively be eliminated given that each would now be able to claim the entire $50,000 under the annual exclusion no matter how many children each had.

The per donor model has significant merit when placed in the context of Crummey powers. Capping the amount of the annual exclusion based upon the donor, while raising the amount excludible in a reasonable manner, not only eliminates the potential for significant abuse, but also maintains the integrity of the policy rationale behind allowing Crummey powers. In turn, such a reform serves as a justi-
fiable middle ground between complete elimination of the Crummey power and the current system.

C. Creation of an Independent Fiduciary for Minor Beneficiaries

The examples presented earlier concerning the Cunningham clan are some of the more abusive and ubiquitous abuses of the annual exclusion through the use of Crummey powers.151 By permitting a parent with a vested economic interest in the trust to allow his minor child’s right to withdrawal lapse, the economic interests of the child are not met. In other legal arenas, when a child’s best interests are not met by a parent, a court or statute will step in and appoint an agent to ensure that the minor child is properly represented.152 Minor Crummey beneficiaries should be treated no differently. Therefore, rather than relying upon a parent beneficiary with an economic interest in their minor child’s lapse, a statutorily mandated independent fiduciary should be appointed in cases where the exercise of a minor child’s withdrawal right is dependent upon a parent beneficiary with a vested interest in the same trust. The fiduciary, guided by the circumstances surrounding the demand power, may be seen as the gatekeeper of economic rationality, ensuring that the integrity of the withdrawal power is kept intact.153 Moreover, this remedy provides an added benefit to courts in that it removes the burden of having to subjectively determine why a particular demand was not made because the court may assume that a decision to lapse made by an independent fiduciary was in the absolute best interests of the child.

In addition, the presence of an independent fiduciary serves to curb abuse of the annual exclusion through the multiplication of contingent beneficiaries without any corresponding risk of exercise. As indicated by the Cunningham hypothetical, a Crummey grantor can increase the number of annual exclusions claimed in a given year simply by granting the children of each primary beneficiary a meaningless contingent interest in the trust and giving them a lapsing withdrawal right. Therefore, so long as the Crummey grantor can ensure that the parent will not withdraw, there is no increase in risk that the contingent beneficiaries will exercise their demand power either (because the parent will make the decision for them). However, the situation is drastically changed when it is no longer an economi-

---

151. See supra Section VIII.D.
152. For example, in large tort settlements in which the plaintiff is a minor, a court will often step in and appoint a guardian to oversee distributions of the settlement to ensure that the child’s economic best interests are maintained.
153. After all, if the Crummey grantor truly intended to give the beneficiary an absolute right to demand, then actual withdrawal will have no unexpected consequence to the donor.
cally interested parent making the decision to lapse, but rather an independent fiduciary who will make a determination based upon the best interests of the child. In such a situation, a Crummey grantor cannot be certain that the minor beneficiary’s demand power will be lapsed, thereby defeating the purpose behind granting the demand power to begin with. As a result, with the fear that an independent fiduciary will actually exercise a minor beneficiary’s demand right, Crummey grantors will be less likely to grant minor beneficiaries demand powers coupled with meaningless contingency rights in trust, thereby reducing the number of “sham” annual exclusions claimed by grantors such as Howard Cunningham.

D. Establish a Five-Year Window of Review for All Trusts With Lapsing Demand Powers

Due to the reluctance of most courts to recognize the validity of the pre-arranged “understanding” argument, the Service has been relatively unsuccessful in combating the use of lapsing Crummey powers to multiply the amount that grantors may transfer into trust under the annual exclusion. For most courts, the problem with the pre-arranged “understanding” argument lies in the fact that it relies primarily upon a subjective assumption by the IRS—that is, absent an agreement between the grantor and beneficiary, the decision to lapse is illogical. In turn, to remove the taint of subjectivity from its attack on abusive Crummey powers, the IRS should employ an objective mechanism that will not only effectively determine the validity of a Crummey beneficiary’s decision to allow lapse, but also confine the availability of the annual exclusion to only those Crummey beneficiaries that are in keeping with the policy rationale underlying the Ninth Circuit’s decision in Crummey.

One method by which the IRS may accomplish this two-pronged goal is through the creation of a five-year “look back” window that the Service can use to determine the validity of the annual exclusions claimed under a Crummey trust. Under this approach, the IRS will review all Crummey trusts five years after their creation and, in doing so, utilize three objective factors to determine whether the annual exclusions claimed by way of Crummey withdrawal powers were bone fide present interest gifts, and therefore justified.

First, the Service must look to the relationship between the grantor and the beneficiaries that received lapsing withdrawal rights during the first five years of the trust. Any decision to lapse made by a beneficiary that is a direct descendent (son or daughter) of the grantor will be considered genuine per se, so long as that beneficiary has

154. See Fogel, supra note 84, at 612.
an irrevocable vested interest in the trust principal. In turn, the annual exclusions claimed with regard to the Crummey powers granted to those beneficiaries will stand, thereby exempting them from any future “look back.”

If, however, an annual exclusion is claimed for a lapsed demand power granted to a beneficiary other than a direct descendent, a five-year “look back” window will be triggered. Using this “look back,” the Service will review the actual operation of the withdrawal rights granted to those beneficiaries. If it is determined that any of the non-exempted beneficiaries have failed to exercise their Crummey demand rights in more than two of the first five years of the trust’s existence, then the annual exclusions claimed for those beneficiaries will be deemed illusory, and the grantor will no longer be afforded the benefit of the annual exclusion for any future lapsing demand powers given to that beneficiary. To prevent grantors from circumventing the purpose of the five-year “look back” by simply complying with the requirements for the first five years of the trust, the Service will again review the trust every ten years thereafter to ensure compliance.155

Finally, regardless of whether a beneficiary is exempted from the “look back” or not, at each point of review (five years from date of creation and every ten years thereafter) the Service will review each lapsing demand power to ensure that it fully complied with the notice and opportunity to exercise requirements set forth in Revenue Ruling 81-7.156 Any lapsing withdrawal right granted to a beneficiary that fails to meet the notice and opportunity-to-exercise requirements will be deemed illusory and thus fail to qualify for the annual exclusion in each year where the requirements were not met.

Although far from perfect, this proposed solution avoids the pitfalls of the Service’s current pre-arranged “understanding” argument by employing a mechanical test that is reliant only upon objective criteria to prove the genuineness of a Crummey demand right. Moreover, this answer maintains the spirit of Crummey by limiting the use of the Crummey mechanism to the beneficiaries that were of primary concern to the Ninth Circuit—the grantor’s children.

X. CONCLUSION

When Crummey powers are utilized outside of their originally intended context, the potential for abuse of the annual exclusion is exponentially increased to a point that borders on inevitability. This abuse comes in all shapes and sizes, with the greatest opportunity for

155. The ten-year review will require a beneficiary to have exercised his withdrawal right in six of the ten years reviewed in order for the annual exclusions to be deemed valid.
156. See Rev. Rul. 81-7, 1981-1 C.B. 474; supra text accompanying note 42.
abuse arising in the context of beneficiaries with a “naked” demand right. Despite this abuse, the Crummey power is a firmly entrenched legal tool with a sound policy rationale behind it. Due to the justifiable policy goals furthered by Crummey powers, Crummey need only be fixed, not eliminated altogether. In turn, the IRS and the legal community are best served by looking for specific solutions to the narrow problems created by the Crummey power, rather than pushing for an all out repeal of the Crummey decision.

By far, the most equitable of these specific reforms is the donor annual exclusion model. It is the only solution that can both maintain the integrity of the annual exclusion, while at the same time preserve the justifiable policy rationale behind Crummey. Once instituted, such a policy will finally afford parents the opportunity to pass on to their children what is rightfully theirs, and render unto Caesar what is rightfully his.
I. INTRODUCTION..................................................................................................... 957
A. Historical Origin................................................................................................ 957
B. Current Crisis..................................................................................................... 959
II. THEORIES OF CIVIL LIABILITY ............................................................................. 961
A. Breach of Fiduciary Duty.................................................................................. 961
B. Intentional Infliction of Emotional Distress.................................................. 963
C. Clergy Malpractice.......................................................................................... 965
D. Respondeat Superior......................................................................................... 967
E. Negligent Hiring, Supervision, and Retention.............................................. 969
III. NEGLIGENT HIRING, SUPERVISION, AND RETENTION AND THE FIRST
AMENDMENT ........................................................................................................ 971
A. Split of Authority............................................................................................ 971
B. Analysis Under the First Amendment............................................................ 972
C. The Free Exercise Clause................................................................................ 972
D. The Establishment Clause............................................................................... 975
IV. MALICKI V. DOE AND DOE V. EVANS................................................................. 979
V. CONCLUSION ........................................................................................................ 983
A. Mandatory Reporting and Clergy Privilege.................................................... 984
   1. Reconciliation and Waiver of the Shield....................................................... 985

I. INTRODUCTION

A. Historical Origin

Papal courts gained popularity in the Middle Ages as a major source of the Roman Church's power and wealth; and the Pope, as the head of the papal court, often decided cases in collaboration with a king, a duke, or an archbishop.1 Ironically, early concerns over the papal court centered on the Pope's entanglement in the secular and political life of Europe. “[T]o think chiefly in legal, was to think chiefly in secular, terms,”2 a practice ill-suited to Peter's successors. Increasingly, the Popes of the Middle Ages acted like kings and became entrenched in diplomatic, international relations. Efforts to reform the Church by improving ethical standards and “disengaging the clergy from their role as supporters of the State, ended, by a kind of helpless logic, in thrusting the Church far more deeply and completely into the secular world [and] . . . the Church became a secular

* B.A. Haverford College, 1994; J.D. Florida State University, expected May 2003. Thank you to Professor Vinson and Bill, who managed to dredge through the “legaldegooky garbage” to offer valuable comments; to my friends and family, for their encouragement and faith; and to Steve, whose love makes roller coasters worth riding.
2. Id.
world of its own.” Like many of the rival kingdoms in Europe, the Church and the State were bound for war.

The personal and political conflict between Thomas Becket (1118-1170) and King Henry II (1154-1189), one of the most competent of English kings, demonstrates early tensions between conflicting authorities. In the early years of King Henry II’s reign, lawlessness ran rampant among members of the clergy; but clergymen were tried in the less severe legal system of the Church. Hoping to rid his kingdom of this double standard, King Henry looked to his friend Thomas Becket, who had served England as a “seemingly complaisant chancellor.”

On June 2, 1162, King Henry secured the election of Becket as archbishop of Canterbury. Henry hoped that he and Becket might end the conflict created by the existence of two legal systems. But as archbishop, Becket found in God a higher authority than the king. When a Worcestershire cleric raped a young girl and murdered her father:

Becket had [him] branded. This was open to four objections: it was inadequate; it was a sentence unknown to canon law; it was, indeed, a usurpation of royal authority; and it flatly contradicted Becket’s own argument that clerks should not suffer mutilation, normal in royal courts, ‘lest in man the image of God be deformed.’

King Henry responded by enacting the Constitutions of Clarendon, which limited the power of the Church in a number of ways, including the subjection of clergy to civil courts and the weakening of the Church’s power of excommunication. As a servant to God, no longer to the king, Becket resisted these changes. Tensions grew between Church and State, and King Henry, temporarily enraged, called for Becket’s death. Norman knights carried out the King’s request, and “Henry was forced to abandon the Constitutions of Clarendon and to do public penance at Becket’s grave.” Despite the apparent victory for the Church, the King of England, rather uneventfully, continued his control over ecclesiastical affairs.

3. Id.
4. See id. at 208 (“Perhaps one in fifty people could make some claim to be considered in orders. And of these about one in six could expect to get into trouble with the law.”).
6. Id.
7. See JOHNSON, supra note 1, at 207.
8. See id. at 209.
9. Id.
10. WALKER ET AL., supra note 5, at 367.
11. See id. at 368.
12. Id.
13. Id.
B. Current Crisis

Church and State have collided once again; and once again, the Church has been entirely too forgiving under the legal standards of the State. In June 2001, Cardinal Bernard F. Law admitted that in 1984 he appointed Reverend John J. Geoghan parochial vicar of a suburban parish two months after learning that Geoghan allegedly molested seven boys.\(^\text{14}\) Law’s admission prompted investigative reports—reminiscent of Woodward and Bernstein’s uncovering of Watergate—by editor Walter V. Robinson and reporters Matt Carroll, Sacha Pfeiffer, and Michael Rezendes of The Boston Globe’s “Spotlight Team.”\(^\text{15}\) The Spotlight Team gained access to confidential documents and discovered written proof that the archdiocese had known about Geoghan’s abuse of children for decades.\(^\text{16}\) In over 300 newspaper articles on clergy sexual abuse, the Globe reported that cardinals and “bishops had known about [numerous incidents of] abuse [by members of the clergy] but failed to remove the priests from their jobs . . . [and] that over the past decade the Archdiocese of Boston had secretly settled cases in which at least seventy priests had been accused of sexual abuse.”\(^\text{17}\) But the problem was much larger than events in Boston. From January to April 2002, 176 priests across the country were removed from their positions, and bishops in the United States, Poland, and Ireland resigned.\(^\text{18}\)

Reports of abuse and of the Church’s failure to protect minors led infuriated Catholics to withhold contributions to the Church and demand reform.\(^\text{19}\) Some state legislators amended mandatory reporting laws to include clergy among the list of mandatory reporters, and

\(\text{15. See id.}\)
\(\text{16. Id.}\)
\(\text{17. Id. For a full text of Cardinal Law’s June 5th and June 7th depositions, detailing the Cardinal’s appointment of known pedophiles to positions within the Church, see Pam Belluck, \textit{Cardinal Told How His Policy Shielded Priests}, N.Y. TIMES, Aug. 13, 2002, available at http://www.nytimes.com/2002/08/14/national/14CARD.html (last visited Jan. 7, 2003) (on file with author); see also Fred Bayles, \textit{Abuse Victims Flock to Lawyers}, USA TODAY, July 30, 2002, available at http://www.usatoday.com/news/religion/2002-07-30-abuse-litigate_x.html (last visited Jan. 7, 2003) (“Attorneys who defend the church say the swift, quiet settlements of the past are now impossible with the flood of cases and news reports of church coverups.”); Margaret Cronin Fisk, \textit{Church’s New Trial}, NAT’L L.J., May 10, 2002, available at http://www.law.com (last visited Nov. 25, 2002) (explaining that previously settled claims are resurfacing in litigation where it is alleged that settlements were fraudulently induced, “the church failed to live up to agreements to remove the priests from the active ministry,” or the Church falsely “testified that there were no other victims”).}\)
\(\text{18. THE INVESTIGATIVE STAFF OF THE BOSTON GLOBE, supra note 14, at 4.}\)
\(\text{19. Id.}\)
prosecutors began issuing arrest warrants for priests. Cardinals of
the United States and the leadership of the United States Conference
of Catholic Bishops (the “USCCB”) traveled to the Vatican in April
for a meeting with the Prefects of the Roman Congregations, where
Pope John Paul II issued a statement recognizing that the sexual
abuse of minors “is by every standard wrong and rightly considered a
crime by society [and that] it is also an appalling sin in the eyes of
God.” As a result of this meeting, on June 14, 2002, the USCCB
adopted its first national policy on dealing with sexual abuse.

The Charter for the Protection of Children and Young People
states that “[t]here is no place in the priesthood or religious life for
those who would harm the young.” But this statement comes more
than a decade too late, and the Church is now called to defend itself
against hundreds of lawsuits in which church victims are filing tort
claims ranging from breach of fiduciary duty to negligent hiring, su-
ervision, and retention. To fight this war, the Church has replaced
its once-trusted weapon of excommunication with a more modern
shield: the First Amendment. Behind this shield, battles are being
won and lost.

In examining third party tort actions against the Church, Part II
next surveys the causes of action asserted to hold a religious organi-
sation liable for the sexual misconduct of a member of its clergy.
Part III focuses on the viability of one of those causes of action—negligent hiring, supervision and retention—under First Amendment analysis. Part IV discusses the Florida Supreme Court’s decisions in *Malicki v. Doe* and *Doe v. Evans*, where the court held that the First Amendment did not bar a claim against a religious institution, at least at the initial pleading stage, for harm caused by the sexual misconduct of a cleric. 25 Finally, this Comment identifies common issues faced by those who represent victims of clergy sexual misconduct and explores potential avenues to reconciling church doctrine with justice.

II. THEORIES OF CIVIL LIABILITY

Over the last few years, allegations of clergy sexual misconduct have been asserted in greater numbers, and it has become increasingly common for plaintiffs to sue religious organizations for their clerics’ misconduct. Not coincidentally, courts increasingly hold religious organizations liable for clerical torts. Theories of liability—including imputed or direct negligence, breach of a fiduciary duty, respondeat superior or agency, intentional infliction of emotional distress, clergy malpractice, and negligent hiring, supervision, and retention—have been received by the courts with varying degrees of success. A sampling of these causes of action is discussed below.

A. Breach of Fiduciary Duty

Some state courts have recognized a cause of action against the Church for breach of fiduciary duty with respect to sexual misconduct of clergy. 26 To recognize this cause of action, a court must first find the presence of a fiduciary relationship between the Church hierarchy and the victim of abuse. Under section 874 of the *Second Restatement of Torts*, “[a] fiduciary relation exists between two persons when one of them is under a duty to act for or to give advice for the benefit of another upon matters within the scope of the relation.”


Fiduciary relationships are “founded upon trust or confidence re-posed by one person in the integrity and fidelity of another.” While not every clergy-parishioner relationship is a fiduciary relationship, many give rise to fiduciary duties because of the parishioner’s dependence on, and faith in, the Church.

Florida courts are among the state courts that recognize a cause of action for breach of a fiduciary duty against a religious institution or members of the clergy, and a cleric “who commits a breach of his duty . . . is guilty of tortious conduct to the person for whom he should act.” Fiduciary duties may be the product of personal, moral, or social relations; “liability is not dependent solely upon an agreement or contractual relation between the fiduciary and the beneficiary.” For example, the Florida Supreme Court has held that a church that promotes its clergy as qualified marriage counselors has a fiduciary duty to its “counselees,” despite the fact that most parishioners do not contract with the Church prior to receipt of counseling. Because liability is based on factual distinctions rather than legal distinctions, juries—not judges—decide: (1) whether a fiduciary relationship exists between the plaintiff-parishioner and the defendant-clergy, and (2) whether the defendant breached his or her fiduciary duty. Jury instructions that properly and simply define the characteristics of a fiduciary relationship are therefore essential.

In Moses v. Diocese of Colorado, the Supreme Court of Colorado agreed that the existence of a fiduciary relationship is a question of fact for the jury and held that the trial court had “properly instructed [the jury] on the requisite elements of a fiduciary relationship.” The trial court had asked jurors to evaluate the level of trust that the parishioner placed in the clergy-defendants; whether the trust was justified; whether the defendants knew or should have known that the parishioner relied on them to act in her best interest; whether the defendants “invited, accepted, or acquiesced in” the parishioner’s trust; and whether the defendants attempted to protect the parishioner’s interests. The high court upheld the jury’s finding of a fiduciary relationship and affirmed the trial court’s holding that the Diocese of Colorado and the defendant bishop breached their duties “to deal

30. See, e.g., Doe v. Evans, 814 So. 2d 370, 376 (Fla. 2002); Shealey v. Masters, 821 So. 2d 342, 345 (Fla. 4th DCA 2002); Palafrugell Holdings, Inc. v. Cassel, 825 So. 2d 937, 940 (Fla. 3d DCA 2001).
32. Id.
33. Evans, 814 So. 2d at 375.
34. Id.
35. 863 P.2d 310, 322 (Colo. 1993).
36. Id. at n.14.
‘with utmost good faith and solely for the benefit’ of the dependent party.\textsuperscript{37} In reaching this conclusion, the court never addressed church doctrine nor relied on a layperson’s definition of reasonableness.

Yet some courts have been unwilling to recognize this action for breach of fiduciary duty because the inquiry would present “constitutional difficulties” in defining the standard of care for a clergyman.\textsuperscript{38} In \textit{H.R.B. v. J.L.G.}, for example, the Missouri Court of Appeals identified a fiduciary relationship similar to the one discussed in \textit{Moses} but refused to recognize a cause of action for breach of fiduciary duties against the defendant priest, archbishop, and Church of the Immaculate Conception School and Parish.\textsuperscript{39} Although it conceded that the priest’s sexual misconduct was not motivated by religious beliefs, the court somehow concluded that an inquiry into the religious aspects of the relationship between the Church authorities and the plaintiff parishioner was constitutionally improper.\textsuperscript{40}

Beyond the bond shared between a parent and child, it is difficult to imagine a more sacred relationship than the one shared by a faithful parishioner and his or her church. Even the term “Father,” as used by members of the clergy, invites trust and lulls loyal followers into feeling safe. This relationship must be guarded at all costs, and the Church hierarchy should be liable for failing to protect it.

\textbf{B. Intentional Infliction of Emotional Distress}

Intentional infliction of emotional distress is a relatively new cause of action that has slowly gained recognition in state courts. Perhaps one reason for this slow pace is the potential for abuse that haunts a cause of action based on emotional rather than physical injuries. Nevertheless, modern science has provided courts—and more importantly, juries—with greater understanding of the human mind. Americans have become sensitized to the reality of psychological trauma, and the judicial system has acknowledged that legitimate claims for mental suffering exist.

The Florida Supreme Court, for example, first recognized the tort of intentional infliction of emotional distress in 1985 in \textit{Metropolitan Life Insurance Co. v. McCarson}.\textsuperscript{41} Two years later, in \textit{Dependable Life Insurance Co. v. Harris}, the court identified the four elements re-

\textsuperscript{37} Id. at 323 (quoting Destefano v. Grabrian, 763 P.2d 275, 284 (Colo. 1988)).
\textsuperscript{39} 913 S.W.2d 92, 98-99 (Mo. Ct. App. 1995).
\textsuperscript{40} Id. at 99. \textit{But see} Doe v. Evans, 814 So. 2d 370, 376 (Fla. 2002) (holding that Doe’s breach of fiduciary duty claim was not barred by the First Amendment, as the court is not being called upon to interpret ecclesiastical doctrine, and the “claim is governed by neutral tort law principles of general application”).
\textsuperscript{41} 467 So. 2d 277, 278 (Fla. 1985).
quired to establish a cause of action for intentional infliction of emotional distress: (1) the defendant must have engaged in the deliberate or reckless infliction of mental suffering; 42 (2) the conduct must be outrageous; (3) the conduct must have caused the emotional distress; and (4) the distress must be severe. 43 In helping to define the terms "deliberate," "reckless," and "outrageous," the Fifth District Court of Appeal explained that the defendant had to act with purpose or recklessness, meaning that the defendant knew or should have known that the alleged actions would cause severe distress, and the defendant’s behavior must have been conduct that any reasonable person would call indecent or intolerable. 44 Combined, these elements form a high standard designed to safeguard against fraudulent claims and to reflect the evolution of public opinion.

In stating a cause of action for intentional infliction of emotional distress against a religious institution, a plaintiff must allege active conduct by the Church that directly caused his or her severe distress; it is not enough to allege the sexual misconduct of a cleric. 45 For example, in Elders, the Third District Court of Appeal affirmed the trial court’s dismissal of a claim for intentional infliction of emotional distress against the Florida Conference, Saint John’s on the Lake United Methodist Church, and the District Superintendent of the Methodist Church. 46 The court explained that “the plaintiffs’ allegations boil[ed] down to a claim of negligent failure to supervise Pastor Rivers[, which was] legally insufficient to establish a claim” based on outrageous conduct. 47

Few plaintiffs, if any, have been successful in imposing liability on a religious institution for the intentional infliction of emotional distress because it is difficult to prove that the Church caused the plaintiff’s distress by taking actions separate and distinct from the cleric’s sexual misconduct and to prove that the Church’s actions were “ou-

42. See Williams v. City of Minneola, 619 So. 2d 983, 987 n.2 (Fla. 5th DCA 1993) (“Gross negligence does not meet the standard for an award of punitive damages, . . . and, thus, certainly cannot meet the standard to establish the tort of outrageous and reckless conduct.”). But see Champion v. Gray, 478 So. 2d 17, 18 (Fla. 1985) (recognizing a cause of action for negligent infliction of emotional distress under limited circumstances when there is “death or significant discernible physical injury, when caused by psychological trauma resulting from a negligent injury imposed upon a close family member within the sensory perception of the physically injured person”).

43. Dependable Life Ins. Co. v. Harris, 510 So. 2d 985, 986 (Fla. 5th DCA 1987) (citation omitted).

44. See Williams v. City of Minneola, 575 So. 2d 683, 690-91 (Fla. 5th DCA 1991); see also Metro. Life, 467 So. 2d at 278-79 (explaining that the conduct must be “so outrageous in character, and so extreme in degree, as to go beyond all possible bounds of decency”).

45. See Elders v. United Methodist Church, 793 So. 2d 1038, 1042 (Fla. 3d DCA 2001).

46. Id.

47. Id.
rageous.” The *Boston Globe*’s investigative reporting on clergy sexual abuse educated the public about a scandal that seemed almost too horrible to be true. The reports showed that members of the Church hierarchy—including Cardinal Bernard F. Law of Boston, the most influential American Catholic prelate in the Vatican—were not only aware of the abuse but had gone to enormous lengths to hide the scandal from public view . . . .

Most shocking to everyday Catholics, and most damaging to the Church, was the incontrovertible evidence that Cardinal Law and other leaders of his archdiocese had engaged in such a massive cover-up. Rather than protect its most vulnerable members, the Church had been putting them in harm’s way.48

Unlike *Elders*, where allegations against the Church defendants were founded on the Church’s failure to act, the *Boston Globe*’s account of Cardinal Law’s behavior describes outrageous conduct, distinct from the sexual misconduct of Reverend Geoghan, that purposefully and recklessly caused the severe distress experienced by hundreds of victims of sexual abuse. These allegations would clearly state a cause of action for the intentional infliction of emotional distress.

**C. Clergy Malpractice**

The term *malpractice* refers to “a professional’s improper or immoral conduct done either intentionally or through carelessness or ignorance.”49 It is often used to denote a physician’s or lawyer’s “unskillful performance of duties resulting from such person’s professional relationship with patients or clients.”50 Unlike an intentional tort or ordinary negligence—which is actionable regardless of a person’s profession—malpractice, by definition, is “the breach of a professional duty unique to that profession.”51 Thus, while the Church may be liable for an intentional tort or ordinary negligence, most courts have not recognized clergy malpractice because it calls for setting a standard for the duties of a clergyman.52 And while courts greatly disagree on whether claims for torts like negligent hiring are

50. *Id.*
52. *See*, e.g., *Dausch* v. *Ryske*, 52 F.3d 1425, 1428 (7th Cir. 1994) (stating that the plaintiff’s claims could be heard under traditional professional negligence claims and declining to recognize a cause of action for clergy malpractice); *Destefano* v. *Grabrian*, 763 P.2d 275, 285 (Colo. 1988) (en banc) (“We do not recognize the claim of ‘clergy malpractice.’”); *Leary* v. *Geoghan*, 2000 WL 1473579, at *2 (Mass. Super. 2000) (“[I]t is safe to say that there is no such thing as ‘clergy malpractice’ in Massachusetts, or most other places.”); *Hester*, 723 S.W.2d at 550 (“[M]inisterial malpractice is a tort not known in Missouri law.”); *Byrd* v. *Faber*, 565 N.E.2d 584, 587 (Ohio 1991) (“[T]here is no basis for recognizing [the] claim for clergy malpractice.”).
barred by the First Amendment, they seem to agree that the states, through their judicial systems, may not set such a standard. 53

The court’s rejection of the plaintiff’s claim for clergy malpractice in Schmidt v. Bishop is typical. 54 In Schmidt, Christine Schmidt claimed that Reverend Joseph Bishop, a pastor at the Rye Presbyterian Church, sexually abused her under the auspices of providing family counseling when she was twelve years old. 55 The court acknowledged that the alleged facts stated a cause of action for battery, but then explained that Ms. Schmidt had to rely on alternative causes of action since the statute of limitations had run on her claim for an intentional tort-like battery. 56 Malpractice was one of the many alternatives pleaded in the complaint.

According to the court, Ms. Schmidt “deliberately avoided asserting that the case involve[d] clergy malpractice, but rather style[d] [the] claim artfully as one for malpractice by a ‘youth worker and counselor.’” 57 One reason Ms. Schmidt attempted to fit her claim into the category of counselor malpractice may have been because the first and most cited case on clergy malpractice, Nally v. Grace Community Church of the Valley, left open the question of whether the First Amendment barred a claim of clergy malpractice for negligent counseling. 58 Moreover, asserting a more established cause of action


55. Id. at 324.

56. Id. New York has not adopted an extended statute of limitations (“SOL”) for sexual abuse or tolling provisions for delayed realization cases. If the abuse is treated as an intentional tort, New York’s SOL is one year. N.Y. CIVIL PRAC. LAW § 215 (McKinney 2002). An action based in negligence, rather than criminal conduct, has a SOL of three years. Id. § 214. Other states, like Florida and Massachusetts, have adopted the delayed discovery doctrine. See Herndon v. Graham, 767 So. 2d 1179, 1181 (Fla. 2000) (stating that the SOL does not begin to accrue until the victim is aware that the abuse occurred and that accrual avoids the seven-year statute of repose); MASS. GEN. LAWS ch. 260, § 4C (2000) (setting the SOL for sexual abuse of a minor at three years from the time of the alleged acts or “within three years of the time the victim discovered or reasonably should have discovered that an emotional or psychological injury or condition was caused by said act”).


58. 763 P.2d 948, 960 (Cal. 1988) (addressing several claims by parents against a church and its pastors for the wrongful death of a son who committed suicide after counseling by the pastors).
for malpractice by a youth worker and counselor was not a complete stretch. Indeed, “[t]he duties of a clergyman most nearly approximate to an existent professional practice, and hence most accountable to minimum professional standards, . . . include that of spiritual counseling.” The Missouri Court of Appeals, nevertheless, took the position that despite Ms. Schmidt’s broad use of the term “malpractice,” the real issue had to be clergy malpractice; otherwise, by analogy, “a medical malpractice action against a surgeon might well be characterized as one for ‘counseling malpractice,’ since surgeons often engage in pre- and post-operative ‘counseling.’” After calling the horse a “horse,” so to speak, the court addressed the viability of clergy malpractice theories.

Typically, clergy malpractice claims have been denied either because they fail to allege distinct facts pertinent to the clergy-parishioner relationship that are not already actionable, or because the court is concerned that the alleged conduct is within the purview of the First Amendment. In Schmidt, the court stated that “[i]t would be impossible for a court or jury to adjudicate a typical case of clergy malpractice, without first ascertaining whether the cleric . . . performed within the level of expertise expected of a similar professional . . . .” To measure a cleric’s conduct, for example, the court would have to create a hypothetical, reasonably prudent priest, pastor, or bishop. And while it may be possible to find that reasonably prudent clerics resist sexually abusing children, the court worried that the next case would be less clear. Thus, the court found that an analysis dependent on professional standards within the Church violates the Supreme Court’s prohibition against excessive entanglement with religion.

D. Respondeat Superior

Even if a court finds that the Church hierarchy, meaning the governing organization or officials responsible for the cleric’s employment within the Church, has committed no wrongs distinct from clergy misconduct, the Church may be held vicariously liable for its employee’s misconduct. Under the doctrine of respondeat superior, an employer is held liable for the tortious or criminal acts of an em-

60. Schmidt, 779 F. Supp. at 326.
61. Hester, 723 S.W.2d at 551.
63. See id. at 327-28.
64. See id. at 328 (citing Lemon v. Kurtzman, 403 U.S. 602 (1971)). For further discussion on excessive entanglement, see Part III of this Comment.
65. See RESTATEMENT (THIRD) OF TORTS § 13, cmt. b (1999) (“Perhaps the most popular justification for vicarious liability is that the costs of an agent’s torts should be borne by the enterprise [and] that a financially responsible party will respond if damages occur.”).
ployee if the acts “were committed during the course of the employment and to further a purpose or interest, however excessive or misguided, of the employer.” Most clergy sexual abuse cases that address respondeat superior focus on whether the sexual misconduct was committed within the cleric’s scope of employment.

In defining scope of employment, Florida’s Third District Court of Appeal explained that:

[a]n employee’s conduct is within the scope of his employment, where (1) the conduct is of the kind he was employed to perform, (2) the conduct occurs substantially within the time and space limits authorized or required by the work to be performed, and (3) the conduct is activated at least in part by a purpose to serve the master.

Sexual assaults and batteries are generally held to be outside the scope of an employee’s employment, but an exception to the rule may be found where the tortfeasor used his or her employment to commit the tort. Notwithstanding the fact that allegations of a cleric’s sexual misconduct often include situations where the cleric used his or her position in the Church to gain the trust of and access to a victim, most courts have been unwilling to apply this exception to clergy sexual abuse cases.

In Byrd v. Faber, for example, the Supreme Court of Ohio addressed the viability of a claim against the Ohio Conference of Seventh Day Adventists based on the doctrine of respondeat superior. There the plaintiff claimed that her pastor forced her into a sexual relationship. The court held the doctrine inapplicable because the

66. Iglesia Cristiana La Casa Del Señor, Inc. v. L.M., 783 So. 2d 353, 356 (Fla. 3d DCA 2001) (citation omitted).
67. Id. at 357.
68. Id.; Hennagan v. Dep’t of Highway Safety & Motor Vehicles, 467 So. 2d 748, 751 (Fla. 1st DCA 1985) (finding that it could not be said, as a matter of law, that a trooper’s sexual misconduct, which was committed under the guise of detaining a shoplifting suspect, was not within the trooper’s scope of employment).
69. See, e.g., Rita M. v. Roman Catholic Archbishop of L.A., 187 Cal. App. 3d 1453, 1461 (1986) (“It would defy every notion of logic and fairness to say that sexual activity between a priest and a parishioner is characteristic of the Archbishop of the Roman Catholic Church, [and the court cannot conclude that] the Archbishop ‘ratified’ the concupiscent acts of the priests.”); Elders v. United Methodist Church, 793 So. 2d 1038, 1041 (Fla. 3d DCA 2001) (“As a matter of common sense, having sexual relations with a counselee is not part of the job responsibilities of a minister.”); N.H. v. Presbyterian Church (U.S.A.), 998 P.2d 592, 599 (Okla. 1999) (“Ministers should not molest children. When they do, it is not a part of the minister’s duty nor customary within the business of the congregation. Rather than increasing membership, the conduct would assuredly result in persons spurning rather than accepting a faith condoning the abhorrent behavior.”); Pritzlaff v. Archdiocese of Milwaukee, 533 N.W.2d 780, 783 (Wis. 1995) (affirming that the alleged sexual misconduct of a priest “could not have been in the scope of his employment”).
70. 565 N.E.2d 584, 585 (Ohio 1991).
71. Id.
Adventists “did not hire [the pastor] to rape, seduce, or otherwise physically assault members of his congregation,” and the plaintiff did not allege that they “should reasonably have foreseen that [the pastor] would behave in this manner toward his parishioners.”

Similarly, in Iglesia Cristianoa La Casa Del Señor v. L.M., Florida’s Third District Court of Appeal refused to apply respondeat superior because the plaintiff failed to allege that the sexual misconduct took place on Church property, that the pastor’s behavior was motivated by a desire to serve the Church, or that the pastor had been counseling her on the day of the crime.

Note what these cases do not hold. These cases leave open the prospect that the Church may, under some circumstances, be held vicariously liable for the sexual misconduct of a cleric. These cases merely find that the allegations, as set forth in the individual complaints, fail to support the theory of respondeat superior. A plaintiff-parishioner may, on the other hand, be successful by proving the abuse took place on Church property, or that the cleric characterized his or her behavior as an act of God, or that the abuse took place under the auspice of a counseling session. In addition to the causes of action already discussed in this Comment, the Church may be directly liable for the negligent hiring, supervision, and retention of its clergy.

E. Negligent Hiring, Supervision, and Retention

While the doctrine of respondeat superior generally fails to survive the Church defendant’s motion to dismiss in a clergy sexual misconduct case, religious institutions are often held liable for a cleric’s misconduct under the doctrines of negligent hiring, negligent supervision, and negligent retention (collectively referred to as “negligent hiring, supervision, and retention”). The theory of negligent hiring, supervision, and retention is similar to the doctrine of respondeat superior in that the employer is held liable for the employee’s conduct; but unlike respondeat superior, negligent hiring, supervision, and retention does not require a plaintiff to allege that the employee’s misconduct was within the scope of his or her employment.

Of course, this lowered hurdle does not mean that an employer is strictly liable for any acts committed by his or her employee, against any person and under any circumstances; limitations have developed

72. Id. at 588.
73. Iglesia Cristiana La Casa Del Señor, Inc. v. L.M., 783 So. 2d 353, 357-58 (Fla. 3d DCA 2001).
74. See Destefano v. Grabrian, 763 P.2d 275, 287 (Colo. 1988) (holding that although the priest’s acts did not create a basis for holding the diocese vicariously liable, the diocese may be directly liable for negligently supervising the priest).
75. Garcia v. Duffy, 492 So. 2d 435, 438 (Fla. 2d DCA 1986).
on a case-by-case basis. As the Second Restatement of Torts explains:

A master is under a duty to exercise reasonable care so to control his servant while acting outside the scope of his employment as to prevent him from intentionally harming others or from so conducting himself as to create an unreasonable risk of bodily harm to them if (a) the servant (i) is upon the premises in possession of the master or upon which the servant is privileged to enter only as his servant, or (ii) is using a chattel of the master, and (b) the master (i) knows or has reason to know that he has the ability to control his servant, and (ii) knows or should know of the necessity and opportunity for exercising such control.

Under this definition, it seems that many religious institutions have a duty to exercise reasonable care in hiring, supervising, and retaining members of the clergy because clerics often live on church premises; and unlike secular employers, the Church hierarchy exerts control over every aspect of its “employees” lives. In fact, the Church’s blending of personal and professional lives is motivated, at least in part, by the Church’s desire to know its clergy, and this knowledge is a critical element of the tort. The court’s determination of when the Church knew or should have known of the cleric’s unfitness—in other words, the timing of the Church’s knowledge—determines whether the Church is liable for negligent hiring, negligent supervision, negligent retention, or any combination of the three. The tort is committed at the point in the employee’s tenure when the employer first had knowledge or constructive knowledge of foreseeable harm.

Florida recognized a cause of action for negligent hiring, supervision, and retention in Mallory v. O’Neil. The Florida Supreme Court stated that an employer may be liable for negligence where he or she knowingly employed a dangerous person and knew or should have known that the person “was dangerous and incompetent and liable to do harm.” In establishing a prima facie case, a plaintiff must prove

---

76. See Iglesia Cristiana La Casa Del Señor, 783 So. 2d at 358 (“For the Church to be held liable for negligent supervision, it must have had constructive or actual notice that [the defendant] was unfit to work as a pastor at the Church.”); Byrd, 565 N.E.2d 584, 590 (raising the pleading standards for negligent hiring and holding that “the plaintiff must plead facts which indicate that the individual hired had a past history of criminal, tortious, or otherwise dangerous conduct about which the religious institution knew or could have discovered through reasonable investigation.”).

77. Restatement (Second) of Torts § 317 (1965).


79. 69 So. 2d 313, 315 (Fla. 1954).

80. Id.
that (1) the employer was required but failed to make an appropriate investigation of the employee; (2) an appropriate investigation would have revealed the unfitness of the employee for the particular duty to be performed or for employment in general; and (3) it was unreasonable for the employer to hire the employee in light of the information he knew or should have known. These elements contribute to the broader notion of reasonable foreseeability and enable courts to evaluate an employer's employment practices to determine whether the employer breached a duty of care.

Third party tort claims based on a religious institution's negligent hiring, supervision, and retention of a member of its clergy are highly contentious because the inquiry requires a court to evaluate the reasonableness of a Church's employment decisions. The Church's greatest defense against such claims is that the court's investigation of a religious institution's employment practices is barred by the First Amendment. Although the United States Supreme Court has recognized that "the appointment [of clergy] is a canonical act, [and that] it is the function of the church authorities to determine what the essential qualifications of a chaplain are and whether the candidate possesses them," it has not addressed whether the First Amendment bars a claim for negligent hiring, supervision and retention.

III. NEGLIGENT HIRING, SUPERVISION, AND RETENTION AND THE FIRST AMENDMENT

A. Split of Authority

In seeking to hold a religious institution responsible for the sexual misconduct of a cleric and for the resulting injury to third parties, the tort of negligent hiring, supervision, and retention has emerged as one of the most viable theories of civil liability, and in light of the Boston Globe's reports on Cardinal Law's practice of protecting felons, it could be one of the most useful tools for holding the Church liable. But there is at least one problem: a split of authority. While:

[s]ubstantial authority in both the state and federal courts concludes that . . . the First Amendment is not violated by permitting the courts to adjudicate tort liability against a religious institution[, ] . . . contrary authority . . . concludes that any tort claim against a religious institution founded on negligent hiring or su-

82. Gonzalez v. Roman Catholic Archbishop of Manila, 280 U.S. 1, 7 (1929).
The United States Supreme Court should explicitly hold that the First Amendment does not bar a claim for negligent hiring, supervision, and retention against a religious institution based on the sexual misconduct of its clergy.

B. Analysis Under the First Amendment

The First Amendment of the United States Constitution states that “Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof.” The Supreme Court has made the First Amendment applicable to the states through incorporation into the Fourteenth Amendment. Therefore, state tort law must avoid restraining the free exercise of religion. In applying the First Amendment to state and federal laws, courts have recognized that the Free Exercise Clause and the Establishment Clause serve distinct purposes.

C. The Free Exercise Clause

The Free Exercise Clause—“Congress shall make no law . . . prohibiting the free exercise” of religion—guarantees, “first and foremost, the right to believe and profess whatever religious doctrine one desires.” Nevertheless, as the United States Supreme Court explained in Cantwell v. Connecticut, the First Amendment “embraces

---

83. Malicki, 814 So. 2d at 358. Compare Schmidt v. Bishop, 779 F. Supp. 321, 332 (S.D.N.Y. 1991) (finding that the First Amendment bars a claim of negligent hiring, supervision, and retention because the inquiry “might involve the Court in making sensitive judgments about the propriety of the Church Defendants’ supervision in light of their religious beliefs”), Swanson v. Roman Catholic Bishop of Portland, 692 A.2d 441, 445 (Me. 1997) (holding that the First Amendment bars a claim of negligent supervision because “imposing a secular duty of supervision on the Church and enforcing that duty through civil liability would restrict its freedom to interact with its clergy”), and Pritzlaff v. Archdiocese of Milwaukee, 533 N.W.2d 780, 790 (Wis. 1995), cert. denied, 516 U.S. 1116 (1996) (concluding that the First Amendment bars a claim of negligent hiring or retention because the court would be required to interpret “church canons and internal church policies and practices”), with Doe v. Hartz, 970 F. Supp. 1375, 1431-32 (N.D. Iowa 1997) (finding that the First Amendment does not bar a negligent supervision claim because the tort is based on neutral principles of law), Isley v. Capuchin Province, 880 F. Supp. 1138, 1151 (E.D. Mich. 1995) (concluding that the First Amendment does not bar a claim of negligent supervision because the court would not be required to interpret church doctrine), and Moses v. Diocese of Colo., 863 P.2d 310, 320-21 (Colo. 1993) (holding that the First Amendment does not bar a claim for negligent hiring and supervision because analysis would “not require interpreting or weighing church doctrine and neutral principles of law can be applied”), cert. denied, 511 U.S. 1137 (1994).

84. U.S. CONST. amend. I.


86. U.S. CONST. amend. I.

two concepts,—freedom to believe and freedom to act. The first is absolute but, in the nature of things, the second cannot be. Conduct remains subject to regulation for the protection of society. 88 Since the freedom to act according to one’s belief is not absolute, the court must first ask whether the regulated conduct is “rooted in religious belief.” 89 Even if the behavior is religiously based, neutral laws of general application do not violate the First Amendment. 90 But “[a] law failing to satisfy these requirements must be justified by a compelling governmental interest and must be narrowly tailored to advance that interest”; the court will apply strict scrutiny to determine whether the law violates the First Amendment. 91 Consequently, if the conduct being regulated is religiously based, the court must identify the law’s purpose and function.

The Supreme Court engaged in this neutral-law-of-general-application inquiry in Employment Division v. Smith 92 and Church of the Lukumi Babalu Aye, Inc. v. City of Hialeah. 93 In Smith, the Court examined an Oregon statute that proscribed the use of controlled substances, including the religious use of peyote. 94 The plaintiffs conceded that the criminal law had not been designed to inhibit their religious practices and that it was constitutional as applied to the general public. 95 Astounded by this concession, the Court reminded the parties that it had “never held that an individual’s religious beliefs excuse him from compliance with an otherwise valid law prohibiting conduct that the State is free to regulate. On the contrary, the record of more than a century . . . contradicts that proposition.” 96 Because the statute did not violate the First Amendment, Oregon could deny unemployment compensation to unemployed workers who had been discharged as a result of their religious drug use.97

Three years later, in Lukumi Babalu Aye, members of the Santee Church challenged the constitutionality of the City of Hialeah’s ban on the ritual sacrifice of animals. 98 The Supreme Court acknowledged that a neutral law of general applicability “need not be justi-

88. 310 U.S. 296, 303-04 (1940).
94. Smith, 494 U.S. at 876.
95. See id. at 878.
96. Id. at 878-79; see, e.g., Jones v. Wolf, 443 U.S. 595, 606 (1979) (explaining that “[t]he neutral-principles approach cannot be said to ‘inhibit’ the free exercise of religion, any more than do other neutral provisions of state law governing the manner in which churches own property, hire employees or purchase goods”).
97. Smith, 494 U.S. at 890.
fied by a compelling governmental interest even if the law has the incidental effect of burdening a particular religious practice; but a law that targets the practices of a particular religion and that selectively imposes burdens on religious conduct requires a compelling governmental interest to support it.\textsuperscript{99} Accordingly, the Court initiated an inquiry to determine first, whether the ban on ritual animal sacrifice was neutral; second, whether the laws were of general application; and third, whether the laws were supported by a compelling governmental interest and narrowly tailored to serve that interest.\textsuperscript{100}

The Supreme Court found that although the texts of Hialeah's ordinances were facially neutral, the legislative histories revealed that, in drafting the ordinances, the City had been improperly motivated by a desire to restrict the religious practices of Santeria.\textsuperscript{101} Second, the laws were not generally applied to all members of the public who were engaged in similar activities.\textsuperscript{102} For example, the ordinances outlawed the religious sacrifice of animals but not the nonreligious killing of animals by hunters and fishermen.\textsuperscript{103} Finally, having determined that Hialeah's ordinances were not neutral laws of general application, the court applied strict scrutiny and found that even if the government's interests were compelling, the ordinances were not narrowly drawn to accomplish those interests; the laws were therefore unconstitutional under the Free Exercise Clause of the First Amendment.\textsuperscript{104}

Unlike Hialeah's ordinances, however, a cause of action for negligent hiring, supervision and retention does not violate the Free Exercise Clause. While it may be conceded, \textit{arguendo}, that the Church's employment decisions constitute religious conduct,\textsuperscript{105} common law torts are analogous to Oregon's outlawing of controlled substances in that they are neutral laws of general application and are therefore enforceable against a religious organization. As the court explained in \textit{Smith v. O'Connell}, “[i]t is easy to envision the kinds of ‘anomalies’ that could result from such an absolutist interpretation of the free exercise clause. For example, laws prohibiting murder would have no

\textsuperscript{99.} \textit{Id.} at 531-34 (citing \textit{Smith}, 494 U.S. at 887-89) (“Neutrality and general applicability are interrelated, and, as becomes apparent in [\textit{Lukumi Babalu Aye}], failure to satisfy one requirement is a likely indication that the other has not been satisfied.”).

\textsuperscript{100.} \textit{Lukumi Babalu Aye}, 508 U.S. at 531-34.

\textsuperscript{101.} \textit{See id.} at 534-36.

\textsuperscript{102.} \textit{Id.} at 545-46.

\textsuperscript{103.} \textit{See id.} at 545.

\textsuperscript{104.} \textit{Id.} at 546 (finding that “all four ordinances are overbroad or underinclusive in substantial respects” and that “interests could be achieved by narrower ordinances that burdened religion to a far lesser degree”).

\textsuperscript{105.} In \textit{Malicki}, the Church Defendants never alleged that their employment decisions were guided by religious doctrine, and the Florida Supreme Court concluded that the Free Exercise Clause had not been implicated. \textit{See Malicki v. Doe}, 814 So. 2d 347, 360-361 (Fla. 2002).
application to human sacrifices performed pursuant to some religious practice."\textsuperscript{106} Likewise, the compelling need to protect both adult and minor parishioners from clergy sexual misconduct, and to prevent the Church’s sustained efforts to hide or ignore that misconduct, meets the requirements of the court’s strict scrutiny. For these reasons, the Supreme Court should find that negligent hiring, supervision and retention, as applied to a religious institution, is constitutional under the Free Exercise Clause.

D. The Establishment Clause

The second component of the First Amendment, the Establishment Clause, states that “Congress shall make no law respecting an establishment of religion.”\textsuperscript{107} This mandate for the separation of church and state, a cornerstone of American democracy, prevents both state and federal governments from enacting laws that “aid one religion, aid all religions, or prefer one religion over another.”\textsuperscript{108} Several tests have emerged to guide the courts’ analysis of alleged violations of the First Amendment’s Establishment Clause.\textsuperscript{109} For instance, in addressing constitutional issues that arise within the scope of public education, the court has applied the \textit{Lemon} test set forth in \textit{Lemon v. Kurtzman},\textsuperscript{110} the “endorsement” test, which was first articulated by Justice O’Connor and later adopted by a majority of the Court in \textit{County of Allegheny v. ACLU},\textsuperscript{111} and the “coercion” test, which was born out of the Court’s decision in \textit{Lee v. Weisman}.

While courts are free to apply any or all of the three tests, and to invalidate statutes that fail any one of them,\textsuperscript{113} First Amendment analysis of third party tort claims against a religious institution most often begins with a discussion of the \textit{Lemon} test.\textsuperscript{114} In \textit{Lemon}, the Court established a three-part test to determine whether a neutral law violates the Establishment Clause: (1) the legislation must have a secular purpose; (2) the legislation must not have the primary ef-

\textsuperscript{106} Smith v. O’Connell, 986 F. Supp. 73, 80 (D.R.I. 1997).
\textsuperscript{107} U.S. CONST. amend. I.
\textsuperscript{109} See Newdow v. United States Cong., 292 F.3d 597, 606-07 (9th Cir. 2002).
\textsuperscript{110} 403 U.S. 602 (1971).
\textsuperscript{111} 492 U.S. 573 (1989).
\textsuperscript{112} 505 U.S. 577 (1992).
\textsuperscript{113} See Sante Fe Indep. Sch. Dist. v. Doe, 530 U.S. 290, 310-16 (2000) (applying all three tests to invalidate a school district’s practice of permitting student-led prayers at high school football games).
fect of advancing or inhibiting religion; and (3) the legislation must not promote an excessive government entanglement with religion.  

While the Supreme Court designed the Lemon test to bring greater uniformity to First Amendment analysis, the test—particularly in the 1990s—has been neither consistently applied nor formally overturned. For example, subsequent cases have used the third prong, excessive entanglement, as a factor in considering the primary effect of the law. Other cases have been decided on the basis of either of the first two prongs, and the courts have failed to address the excessive entanglement prong altogether.

Despite this trend to devalue the third prong, the few cases that have applied Establishment Clause analysis to holding a religious organization liable for the tortious conduct of its clergy have bypassed the first two prongs of the Lemon test and have focused on the opportunity for the State to become excessively entangled in the Church’s employment practices.

The Supreme Court of Wisconsin took this approach in Pritzlaff and found that the parishioner’s claims were barred by the First Amendment. Alleging that Father John Donovan used his position as a priest to coerce her into a sexual relationship with him, Ms. Judith Pritzlaff sued the Archdiocese of Milwaukee for negligent hiring, training, and supervision. Although the court found that Ms. Pritzlaff’s claims were time barred, it proceeded with its analysis under the assumption that a cause of action for negligent hiring or retention existed in Wisconsin at the time of the alleged relationship.

117. See Newdow, 292 F.3d at 611 (refusing to examine the second and third prongs of the Lemon test because it had been determined that the 1954 Act, which inserts the words “under God” into the Pledge of Allegiance, violated the first prong and was therefore unconstitutional).
118. See Smith v. O’Connell, 986 F. Supp. 73, 81 (D.R.I. 1997) (holding that “it is unlikely that exercising jurisdiction over a tort action against the Church will result in any ‘excessive entanglement’ between church and state”); Schmidt, 779 F. Supp. at 328 (“It may be argued that it requires no excessive entanglement with religion to decide that reasonably prudent clergy of any sect do not molest children.”); L.L.N., 563 N.W.2d at 440 (“It is well-settled that excessive governmental entanglement with religion will occur if a court is required to interpret church law, policies, or practices.”).
120. See id. at 782-84.
121. See id. at 789 (citing In re Guardianship of L.W., 482 N.W.2d 60 (Wis. 1992) (allowing a moot issue to be decided where it is likely to arise again and should be resolved by the court to avoid uncertainty)). The Pritzlaff court shifted from a discussion on “negligent hiring, training and supervision” to an analysis of “negligent hiring or retention” without
According to the court, “Ms. Pritzlaff would have to establish that the Archdiocese was negligent in hiring or retaining [Father] Donovan because he was incompetent or otherwise unfit”; but the First Amendment barred the court from setting standards for competent priests.\textsuperscript{122}

The court applied the third prong of the \textit{Lemon} test, though not explicitly, and expressed its concerns that inquiry into a Church’s hiring practices would improperly entangle the court in Church policies and practices since “traditional denominations each have their own intricate principles of governance, as to which the state has no right of visitation.”\textsuperscript{123} The majority seems to have been influenced by a law review article that rhetorically asked:

If negligent selection of a potential pedophile for the religious office of priest, minister or rabbi is a tort as to future child victims, will civil courts also hear Title VII challenges by the non-selected seminarian against the theological seminary that declines to ordain a plaintiff into ministry because of his psychological profile?\textsuperscript{124}

Had the justices been able to envision a situation where the Church knew that a seminarian was a pedophile but hired him anyway and continued to put children in his charge, perhaps their answer would have been different;\textsuperscript{125} instead, they insisted that “[a]ny award of damages would have a chilling effect leading indirectly to state control over the future conduct of affairs of a religious denomination, a result violative [sic] of the text and history of the establishment clause.”\textsuperscript{126} Fearing entanglement, the court granted the Church defendants’ motion to dismiss.\textsuperscript{127}

The Wisconsin Supreme Court revisited this issue one year later in \textit{L.L.N. v. Clauder}, where an adult patient became sexually involved with a priest who had counseled her while serving as a hospital chaplain; the patient later brought a claim for negligent supervi-
sion against the diocese that assigned the priest to the hospital. The court discussed its decision in *Pritzlaff* and reasserted its belief that secular remedies could offend church doctrine. As the court explained,

> The reconciliation and counseling of the errant clergy person involves more than a civil employer’s file reprimand or three day suspension without pay for misconduct. Mercy and forgiveness of sin may be concepts familiar to bankers but they have no place in the discipline of bank tellers. For clergy, they are interwoven in the institution’s norms and practices . . . . Therefore, due to this strong belief in redemption, a bishop may determine that a wayward priest can be sufficiently reprimanded through counseling and prayer. If a court was asked to review such conduct to determine whether the bishop should have taken some other action, the court would directly entangle itself in the religious doctrines of faith, responsibility, and obedience.

Accordingly, the court granted summary judgment in favor of the Church defendants.

The Wisconsin Supreme Court’s reasoning in *Pritzlaff* and *L.L.N.* is flawed. First, the court seems to have been skeptical of the facts alleged in each case and consequently held little sympathy for the adult plaintiffs. In *Pritzlaff*, the court noted that Ms. Pritzlaff claimed, twenty-seven years after her initial experience with Father Donovan, that “the sexual relationship was ‘without her consent’ and was a result of ‘force and coercion.’” Then the court quipped: “Of course, she must allege this as fact because a consensual sexual relationship between two adults is no longer actionable in Wisconsin.” In *L.L.N.*, the court’s skepticism rings in its statement that the “un-disputed facts demonstrate that Clau der, a single man, engaged in a consensual sexual relationship with an adult, single, female.” One can only wonder if the cases would have been decided differently if

---

128.  563 N.W.2d 434, 436-42 (Wis. 1997).
129.  *See id.* at 441.
132.  *Pritzlaff*, 533 N.W.2d at 786.
133.  *Id.*
the plaintiffs had been children or even adults with stronger allegations.

Second, the court’s opinions are flawed because they focus on punishment rather than prevention. Recall that the Second Restatement of Torts defines an employer’s duty under the tort of negligent hiring, supervision, and retention as a duty “to exercise reasonable care so to control his servant while acting outside the scope of his employment as to prevent him from intentionally harming others or from so conducting himself as to create an unreasonable risk of bodily harm.”135 This duty requires a religious organization to prevent foreseeable harm; it does not mandate a particular path of punishment. While it is true that the First Amendment prohibits state intervention in internal church disputes that concern religious matters and require interpretation of religious doctrine,136 the duty to prevent harm is consistent with church doctrine; the limits canon law places on the Church’s authority to discipline (redemption and forgiveness) are irrelevant in determining whether the Church acted to prevent harm. Thus, the court may avoid interpreting ecclesiastical law. Enforcing this duty will entangle the courts in no religious policies, and a cause of action for negligent hiring, supervision, and retention passes the Establishment Clause tests.

IV. Malicki v. Doe and Doe v. Evans

On March 14, 2002, the Florida Supreme Court contributed to the national dialogue on the relationship between clergy sexual abuse cases and the First Amendment of the United States Constitution by issuing two opinions, Malicki v. Doe137 and Doe v. Evans.138 These cases were factually similar, although Malicki included the abuse of a minor, and both cases required the courts to expressly construe the First Amendment.139 Despite factual similarities and identical law,
the decisions made by the Third and Fourth District Courts of Appeal contradicted each other. The Florida Supreme Court’s March 14 opinions represent a first step in ending the dispute.

In *Malicki*, a minor and an adult parishioner claimed that Father Jan Malicki sexually assaulted them on several occasions on the premises of St. David Catholic Church, and they jointly sued Father Malicki, St. David Catholic Church, and the Archdiocese of Miami (the latter two defendants are referred to collectively as the “Church Defendants”) on eight counts.140 The first two counts set forth claims of negligent hiring, supervision, and retention against the Church Defendants based on Father Malicki’s sexual misconduct and the Church’s failure “to make inquiries into Malicki’s background, qualifications, reputation, work history, and/or criminal history prior to employing him in the capacity of Associate Pastor.”141 Reaching for their shield, the Church Defendants claimed that the court’s inquiry was barred by the First Amendment because it would “involve the internal ecclesiastical decisions of the Roman Catholic Church required by Canon Law,” and they moved to dismiss the complaint.142

The trial court agreed that the First Amendment barred review of the Church’s employment practices and, prior to addressing the veracity of the plaintiffs’ factual allegations, granted the Church Defendants’ motion to dismiss with prejudice.143 On appeal, the Third District explained that its review was grounded in tort law and not religious doctrine, and went on to consider whether the Church Defendants knew or should have known about Father Malicki’s sexual misconduct and whether they failed to protect the parishioners from reasonable harm.144 In this manner, the Third District held that the First Amendment tolerates a claim for negligent hiring and supervision.145

In *Doe v. Evans*, a female parishioner filed a claim against the Reverend William Evans, the Church of the Holy Redeemer, Inc., the Diocese of Southeast Florida, Inc., and Bishop Calvin Schofield, Jr. (the latter three defendants are referred to collectively as the “Church Defendants”), consisting of several counts, including negligent hiring and supervision.146 The parishioner alleged that Reverend

---

140. *Malicki*, 814 So. 2d at 352.
141. Id. (quoting the Complaint jointly filed by the minor and adult parishioners).
142. Id. at 353 (quoting the Church Defendants’ motion to dismiss); see Serbian E. Orthodox Diocese for the U.S. & Can. v. Milivojevich, 426 U.S. 696, 713-14 (1976) (explaining that civil courts have no jurisdiction over “purely ecclesiastical” disputes concerning subjects such as “theological controversy, church discipline, ecclesiastical government, or the conformity of the members of the church to the standard of morals required of them”).
143. *Malicki*, 814 So. 2d at 353.
144. *Malicki*, 771 So. 2d at 548.
145. Id.
Evans initiated a counseling relationship with her and sexually exploited the difficulties she was having in her marriage.\footnote{147} She also claimed that although the Church Defendants had actual knowledge of prior incidents where Evans had become sexually involved with parishioners he had intended to counsel, and although the Church Defendants “had the right to exercise control over” Evans, they did nothing “to rectify the situation.”\footnote{148} Finally, the parishioner alleged that the Church Defendants’ actions, or more precisely their inactions, were not religiously motivated.\footnote{149} Arguing that the First Amendment barred the parishioner’s claim for negligent hiring and supervision, the Church Defendants moved to dismiss.\footnote{150}

The trial court granted the motion to dismiss, and the Fourth District Court of Appeal affirmed.\footnote{151} The court found that the First Amendment barred the negligent hiring and supervision claims but asserted that a more compelling state interest, such as protecting a child from sexual abuse, may have forced a different outcome.\footnote{152} The parishioner appealed.

Although the Florida Supreme Court decided \textit{Malicki v. Doe} and \textit{Doe v. Evans} separately, the opinions were issued on the same day, and the court expressly recognized that the outcome of \textit{Malicki} controlled \textit{Evans}.\footnote{153} In fact, the court dedicated an entire section of the \textit{Malicki} opinion to a summary of the facts and issues presented in \textit{Evans}.\footnote{154} Thus, a review of the court’s decision in \textit{Malicki} also serves to explain the court’s analysis of \textit{Evans}.\footnote{155}

After an extensive discussion on First Amendment analysis in general, the Florida Supreme Court examined the factual allegations in \textit{Malicki} under the Free Exercise and Establishment Clauses of the

\footnote{147} Id. at 372.\footnote{148} Id.\footnote{149} Id.\footnote{150} Id.\footnote{151} Doe v. Evans, 718 So. 2d 286, 287 (Fla. 4th DCA 1998).\footnote{152} Id. at 289-90. The Fourth District explained:

\text{[W]e are persuaded that just as the State may prevent a church from offering human sacrifices, it may protect its children against injuries caused by pedophiles by authorizing civil damages against a church that knowingly . . . creates a situation in which such injuries are likely to occur. We recognize that the State’s interest must be compelling indeed in order to interfere in the church’s selection, training and assignment of its clerics. We would draw the line at criminal conduct.}\footnote{Id. at 289 (emphasis omitted).}

\text{Id. at 289} (emphasis omitted).

\footnote{153} \textit{Evans}, 814 So. 2d at 371 (“For the reasons expressed in [Malicki], we hold that the First Amendment does not provide a shield behind which a church may avoid liability for harm caused to a third party arising from the alleged sexual misconduct by one of its clergy members.”).

\footnote{154} See \textit{Malicki v. Doe}, 814 So. 2d 347, 359-60 (Fla. 2002).

\footnote{155} \textit{But see Evans}, 814 So. 2d at 382 (Harding, J., dissenting) (“[T]he majority quashes the lower court’s decision in this case on the basis of a questionable extension of Malicki.”).
First Amendment and determined that the Church is not immune from a tort claim of negligent hiring, supervision, and retention. The court first noted that the allegations set forth in the complaint—namely that the Church Defendants were negligent in failing to research Malicki’s character and in allowing Malicki to supervise the parishioners when the Church Defendants “either knew or should have known that Malicki had the propensity to commit sexual assaults and molestations”—closely paralleled the “classic elements” of negligent hiring and negligent supervision claims. Justice Pariente distinguished claims for negligent hiring, supervision and retention from claims for illegal hiring or discharge of a minister: claims based on negligence turn on the reasonable foreseeability of the cleric’s misconduct and not on the religious institution’s reasons for hiring or firing the cleric. Since the question of foreseeability does not implicate religious doctrine, the court found judicial scrutiny consistent with the Free Exercise Clause of the First Amendment. Nor does the application of tort law to a religious institution by a secular court implicate the Establishment Clause. As the court explained, “imposing tort liability based on the allegations of the complaint neither advances nor inhibits religion.” Because the court found that judicial review of the parishioners’ negligence claims does not run afoul of the Free Exercise or Establishment Clauses, it explicitly disapproved the Fourth District’s compelling state interest requirement.

But Malicki does not end the discussion on the First Amendment’s control over tort claims against a religious institution in the state of Florida. In the third to the last sentence of its opinion, the Florida Supreme Court qualified its decision. The court restricted its holding to a statement that:

[T]he First Amendment cannot be used at the initial pleading stage to shut the courthouse door on a plaintiff’s claims, which are founded on a religious institution's alleged negligence arising from the institution's failure to prevent harm resulting from one of its

156. See Malicki, 814 So. 2d at 357 (“Although an entanglement inquiry is associated with the adjudication of an Establishment Clause claim, the extent to which the courts will be called upon to determine matters of church practice also implicates the Free Exercise Clause.”).

157. Id. at 362.

158. See id. at 363 (“[T]he court does not inquire into the employer’s broad reasons for choosing this particular employee for the position, but instead looks to whether the specific danger which ultimately manifested itself could have reasonably been foreseen at the time of hiring.”) (quoting Van Os dol v. Vogt, 908 P.2d 1122, 1132-33 (Colo. 1996)).

159. Id. at 363-64.

160. Id. at 364.

161. See id. (“[W]e reject the distinction that the Fourth District drew in Evans, 718 So. 2d at 289-90, that would apparently allow a negligent supervision claim against a Church defendant only if the underlying sexual misconduct involved criminal activity . . .”.


clergy who sexually assaults and batters a minor or adult parishioner.\textsuperscript{162}

The court’s qualification is important because the standard of review at the pleading stage is more relaxed. A court must assume that all facts alleged in the complaint are true, draw all reasonable inferences favorable to the plaintiff, and decide the issues on questions of law only.\textsuperscript{163} Thus, \textit{Malicki} and \textit{Evans} leave open the First Amendment’s role at later stages of the trial process, and it is unclear whether the Church will be able to lift its shield to avoid liability in the future.\textsuperscript{164}

\section*{V. CONCLUSION}

Beyond First Amendment concerns, litigation against the Church is fraught with political considerations.\textsuperscript{165} Attorneys representing victims of clergy abuse are mindful of the line between aggressive advocacy and an attack on religion.\textsuperscript{166} One attorney confessed that he ini-

\begin{footnotesize}
\begin{enumerate}
\item[162.] \textit{Id.} at 365 (emphasis added).
\item[163.] Connolly v. Sebeco, Inc., 89 So. 2d 482, 484 (Fla. 1956).
\item[164.] For example, several courts have held that certain document discovery requests create First Amendment problems. See, e.g., Seattle Times Co. v. Rhinehart, 467 U.S. 20, 31-34 (1984) (protecting information regarding the identity of contributors to and members of a religious foundation); Baldwin v. Comm'r, 648 F.2d 483, 488 (8th Cir. 1981) (finding that compliance with discovery requests would infringe on a religious organization’s freedom of association). For a general discussion on First Amendment issues in relation to discovery requests, see Nicholas P. Cafardi, \textit{Discovering the Secret Archives: Evidentiary Privileges for Church Records}, 10 J.L. & RELIGION 95 (1993/1994); Jeffrey Hunter Moon, \textit{Protection Against the Discovery or Disclosure of Church Documents and Records}, 39 CATH. LAW. 27 (1999).
\item[166.] While these cases, at least from the victim’s perspective, are seldom about money, settlement agreements and awards for compensatory and punitive damages threaten to bankrupt individual dioceses. See Stephen Kurkjian & Michael Rezendes, \textit{Bankruptcy Filing Called Option for Archdiocese}, BOSTON GLOBE, Aug. 2, 2002, at A1:

The bankruptcy option is also being considered at a time when the archdiocese is grappling with a fiscal crisis . . . and a downturn in the economy that have forced church officials to cut this year’s operating budget by as much as 40 percent, affecting urban parishes, parochial schools, and other church programs. \textit{See also} Olin, supra note 165, at 83 (“Many archdioceses’ insurance policies are tapped out or insufficient to cover the potential liabilities . . . . [T]he country’s 194 dioceses are le-
\end{enumerate}
\end{footnotesize}
tially had difficulty believing that he was “not suing God.” 167 Another attorney found it easier to make a distinction “between the faith and the men who are a part of the faith. Men make mistakes, they can be corrupted, they can be criminal.” 168 Media coverage has helped to define the line through increased public awareness of Church policies and practices. 169 As victims share their stories, calls for reform grow stronger.

A. Mandatory Reporting and Clergy Privilege

Although all states have adopted mandatory reporting laws that require childcare workers to report incidents of known or suspected child abuse to local officials, most states excuse members of the clergy as mandatory reporters. 170 In Florida, for example, mandatory reporters include physicians, health care professionals, school teachers, and social workers, but not members of the clergy. 171 In fact, although the Florida legislature recently amended Section 39.204, Florida Statutes, which abrogates certain privileged communications in cases involving child abuse, the amended statute maintains the penitent priest privilege. 172 Other states have responded to the clergy sexual abuse scandal more swiftly. On May 3, 2002, Massachusetts Governor Jane Swift signed legislation requiring clergy members to report allegations of child sex abuse. 173 The law requires priests, rab-


169. For complete coverage of the Church scandal in Boston, see http://www.boston.com/globe/spotlight/abuse/ (last visited Jan. 7, 2003); Belluck, supra note 17 (recognizing “the extraordinary public interest” in Cardinal Law’s deposition).


172. See 2002 Fla. Sess. Law. Serv. 174 (West) (amending FLA. STAT. § 39.204(3)); FLA. STAT. § 39.201 (“The privileged quality of communication . . . except that between . . . clergy and person . . . does not apply to any situation involving known or suspected abuse . . . .”) (emphasis added).

173. MASS. GEN. LAWS ANN. ch. 119, § 51A (West 2002).
bis, ordained ministers, and other leaders of religious bodies to report incidents of abuse to state officials within thirty days of discovery.\(^174\)

All states should follow Massachusetts's lead and, if necessary, amend their mandatory reporting laws to include members of the clergy. The decision to report sexual abuse must be taken out of the hands of Church officials, who have a conflict of interest when it comes to reporting clergy sexual misconduct. Ironically, the Archdiocese of Boston failed to notify proper authorities of the heinous crimes committed by members of its clergy because the Church hierarchy wanted to protect the priesthood and avoid scandal within the Church. Mandatory reporting laws, furthermore, eliminate First Amendment issues, making it easier to hold the Church responsible for its failure to protect children from known pedophiles. Mandatory reporting laws are neutral laws of general application, and a determination of whether the Church neglected to report known or suspected incidents of child abuse will not require inquiry into church doctrine. Finally, mandatory reporting laws are designed to protect children who are silenced by sexual predators or who are unable to speak for themselves. Spiritual counselors, like teachers or health care workers, must provide those children with a voice.

1. Reconciliation and Waiver of the Shield

After The Boston Globe's Spotlight Team reported on the Boston Archdiocese's protection of Reverend John J. Geoghan in January 2002, the world waited, and waited, for official word from the Vatican. Almost three months later, Pope John Paul II responded in the form of his annual pre-Easter letter.\(^175\) The Pope showed great concern for the world's priests, stating that “as priests we are personally and profoundly afflicted by the sins of some of our brothers who have betrayed the grace of Ordination in succumbing even to the most grievous forms of the [mystery of evil] at work in the world.”\(^176\) But his empathy for victims of abuse was limited to a brief appeal to Catholics to “commit” to Christ “[a]s the Church shows her concern for the victims and strives to respond in truth and justice to each of these painful situations.”\(^177\) The Pope avoided words like “clergy sexual abuse” or “pedophilia,” and critics wondered whether he understood the full extent of the American crisis.\(^178\)

\(^{176}\) Id.
\(^{177}\) Id.
In contrast, some of America’s Catholic leaders have expressed great remorse for the Church’s improprieties; their willingness to deal openly with clergy sexual abuse offers hope for reform. In a letter published in *The Miami Herald*, Archbishop John C. Favalora recognized that:

> [t]he sexual abuse of children and young people by some priests and bishops has caused great pain, anger and confusion, and these feelings have been compounded by the inadequate ways in which some Catholic Church leaders have dealt with these terrible acts . . . [He expressed] great sorrow and regret . . . for the suffering of victims of sexual abuse, their families and [the] Catholic community.179

Beyond showing sympathy, Bishop Wilton Gregory, President of the U.S. Conference of Catholic Bishops, encouraged priests who have engaged in sexual misconduct to “report this fact so that justice and the Church will be served, and [the priest] will be able to live honestly with [his] own conscience.”180 Confession and repentance, an integral part of the Christian faith, should also serve as cornerstones of the Church’s legal defense.

If the Church genuinely wishes to respond “in truth and justice,” then it must waive its right to assert the First Amendment as a shield to liability. It is unethical to preach confession and repentance in a house of God while practicing denial in a court of law. It is time for the Church to set down its shield. This is not war. Like the moment when King Henry II kneeled before Becket’s casket, this is the time to repent.

---

I. INTRODUCTION

[T]he School District was well aware of Mr. Doe’s history. Indeed, Mr. Doe’s mother had specifically cautioned the teachers and the principal of the need to keep a watchful eye on him. . . . Some-time in November, Mr. Doe took Ms. Jones to a secluded area and sexually assaulted her. Ms. Jones, who was menstruating at the time, bled and vomited during the course of the assault and battery. Upon discovering Mr. Doe and Ms. Jones, a janitor told them to clean up the mess, returned them to class, and advised the teachers where he had found them. . . . [T]he teachers had tied other clothing around her waist to hide it, but [her mother] was never . . . informed of any of the circumstances leading to the soil-ing of Ms. Jones’ clothing. . . . The teachers told Ms. Jones not to tell her mother about the incident and encouraged her to forget it.
had happened at all. . . . Because of these incidents and because she had begun to engage in self-destructive and suicidal behavior, Ms. Jones left school and entered a psychiatric hospital. . . . Following her release from the hospital, Ms. Jones attempted to return to school . . . but stayed for only one day because she was once again battered by Mr. Doe and ridiculed by other students for Mr. Doe's earlier sexual attacks on her.¹

Following the 1999 United States Supreme Court decision in Davis v. Monroe County Board of Education,² Verna Williams, lead counsel for the plaintiff, wrote that Davis “is a wake-up call to the nation’s educational institutions—elementary, secondary, and post-secondary alike—to make sure that they take seriously complaints about a student’s sexual harassment by a peer.”³ Ms. Williams succeeded in convincing the Court that educational institutions should be required to pay damages under Title IX of the Education Amendments of 1972 “if they turn their backs when students harass one another sexually.”⁴ While students were previously granted the right to seek damages against educational institutions if sexually harassed by a teacher,⁵ Davis was the first case in which the Court extended this right under Title IX to students sexually harassed by their fellow classmates.⁶

Like Ms. Williams, many women’s rights advocates declared a victory for young women as Davis appeared to finally acknowledge that students in federally funded educational institutions deserve protection and relief from sexual harassment.⁷ This victory seemed much needed, particularly after the American Association of University Women announced in 1993 that eighty-five percent of all female students experienced some form of sexual harassment, with sixty-five percent being harassed in the classroom, and seventy-three percent being harassed in their school hallways.⁸ The Davis decision appeared to give federally funded educational institutions the motiva-

¹. Murrell v. Sch. Dist. No. 1, 186 F.3d 1238, 1243-44 (10th Cir. 1999). Murrell is one of two successful Title IX peer sexual harassment claims to date and demonstrates the level of severity the harassment must reach to be considered actionable sex discrimination under Title IX.
⁴. Id.
⁶. 526 U.S. at 643.
⁸. AM. ASS’N OF UNIV. WOMEN EDUC. FOUND., HOSTILE HALLWAYS: THE AAUW SURVEY ON SEXUAL HARASSMENT IN AMERICA’S SCHOOLS (1993) [hereinafter AAUW].
tion needed to adopt effective policies to protect students from the potentially debilitating effects of sexual harassment.9

However, the majority opinion in Davis, written by Justice O'Connor, sets forth a standard under which students have had difficulty winning their Title IX peer sexual harassment claims.10 While reiterating the Court's rejection of the use of agency principles,11 Justice O'Connor concluded that federally funded educational institutions must have actual notice of, and act deliberately indifferent to, sexual harassment “that is so severe, pervasive, and objectively offensive that it can be said to deprive the victims of access to the educational opportunities or benefits provided by the school.”12 O'Connor stated that the notice requirement “cabins the range of misconduct that the statute proscribes”13 and that all of the required factors under this new standard “combine to limit a recipient’s damages liability to circumstances wherein the recipient exercises substantial control over both the harasser and the context in which the known harassment occurs.”14 While the symbolic significance of granting students the ability to be awarded monetary damages for being subjected to sexual harassment is great, the practical reality is that the federal circuit courts have been careful to construe O'Connor's standard narrowly, thus dismissing many students' Title IX claims even though they have been subjected to what should amount to actionable sexual harassment. Distillations of the Davis standard vary from circuit to circuit, and questions remain unanswered as to the level of control required, the form of actual notice needed to trigger deliberate indifference, and the extremity of post-notice harassment needed to show that an educational institution has been deliberately indifferent.

In this Comment, I consider how the federal circuit courts have reacted to the Davis decision and discuss whether the lower courts have consistently applied and interpreted the Davis standard. Specifically, I discuss in Part II the evolution leading to the Davis decision through which a cause of action has been recognized under Title IX for peer sexual harassment. In Part III, I discuss the standard arising from Davis. In Part IV, I provide a comparative analysis of the federal circuit court decisions applying the Davis standard. In Part V, I conclude by discussing several questions Davis left unre-

10. Id. at 650.
12. Davis, 526 U.S. at 650.
13. Id. at 644.
14. Id. at 645.
solved that have led to the federal circuit courts’ conservatively construing O'Connor's Davis standard.

II. RECOGNIZING A CAUSE OF ACTION UNDER TITLE IX\(^\text{15}\) FOR SEXUAL HARASSMENT OF STUDENTS IN FEDERALLY FUNDED EDUCATIONAL INSTITUTIONS

A. Legislative Intent of Title IX

Legal questions regarding the applicability of Title VII to protect students in federally funded educational institutions from discrimination were mooted when Congress passed Title IX of the Education Amendments of 1972 (Title IX).\(^\text{16}\) Congress crafted Title IX broadly with the intention of reaching all forms of sexual discrimination within the control of a school, as it states, “No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance.”\(^\text{17}\) Senator Birch Bayh, the congressional sponsor of the amendment, “recognized that discrimination can result from a school’s attitude, as well as its actions, toward women”\(^\text{18}\) and stated that “one of the great failings of the American educational system is the continuation of corrosive and unjustified discrimination against women.”\(^\text{19}\)

Title IX forms the basis for student complaints of sexual harassment by a member of his or her educational community.\(^\text{20}\) The Office for Civil Rights (OCR), which provides support to the United States Department of Education, is responsible for enforcing Title IX and other federal statutes that prohibit discrimination in education programs and activities receiving federal financial assistance.\(^\text{21}\) The OCR offers guidelines entitled, “Sexual Harassment Guidance: Harassment of Students by School Employees, Other Students, or Third Parties”\(^\text{22}\) and maintains the belief that “[w]hen a school makes it clear that sexual harassment will not be tolerated, trains its staff,
and appropriately responds when harassment occurs, students will see the school as a safe place where everyone can learn.23

However, in one Title IX lawsuit, the Court applied a narrow reading of Title IX’s enforcement ability.24 The Court found that the only enforcement mechanism for Title IX was the termination of federal money to the discriminatory institution.25 Because this enforcement mechanism only reached the specific program or activity receiving funds and not the entire institution, the Court ruled that Title IX could only apply to specific programs or activities receiving federal funds.26 To clarify the purpose of Title IX after this “ unacceptable decision,”27 Congress passed the Civil Rights Restoration Act of 1987 “to restore the prior consistent and long-standing executive branch interpretation and broad, institution-wide application” of certain civil rights laws, including Title IX.28

B. Progression of Case History Leading to Davis v.
Monroe County Board of Education

Although the language of Title IX does not expressly provide a private cause of action for student victims of sexual harassment, in Cannon v. University of Chicago the Court found that Title IX was sufficiently broad in nature to include an implied private right of action for victims.29 Subsequently, the Court held in Franklin v. Gwinnett County Public Schools that private plaintiffs could receive monetary damages for sexual harassment under Title IX.30 The Franklin Court extended the theories of the traditional line of Title VII sexual harassment cases31 to faculty-on-student harassment, reiterating that sexual harassment is sex discrimination; “when a supervisor sexually harasses a subordinate because of the subordinate’s sex, that supervisor ‘discriminate[s]’ on the basis of sex.”32 The Franklin Court ruled that the sexual harassment of a student by a teacher (or other agent of the school) is considered quid pro quo sexual harassment, a form of sex discrimination prohibited under Title IX.33 But as recently as 1998, the United States Supreme Court “clarified” the

25. Id.
26. Id.
30. 503 U.S. at 76.
32. Franklin, 503 U.S. at 75 (quoting Vinson, 477 U.S at 64).
33. Id.
standard for holding schools liable for damages under Title IX for sexual harassment of a student by a teacher in Gebser v. Lago Vista Independent School District.\textsuperscript{34} The Court held that schools are liable for damages when a school official with authority to take corrective action actually knew about the sexual harassment and acted with deliberate indifference.\textsuperscript{35}

Peer sexual harassment is the most common form of sexual harassment in schools, with over eighty percent of students who are sexually harassed reporting that a peer had sexually harassed them.\textsuperscript{36} The number of allegations of peer sexual harassment has steadily increased throughout the last decade.\textsuperscript{37} However, determining liability for student-on-student sexual harassment has been a comparatively arduous journey as determining standards of liability for workplace sexual harassment. Before Davis, numerous conflicting lower court decisions made the discussion of school liability for peer sexual harassment ripe for United States Supreme Court review. Such conflicts primarily stemmed from defining the appropriate standards for notice, authority, and responsibility.\textsuperscript{38}

As early as 1993, the Ninth Circuit heard the first case of student-on-student sexual harassment in Doe v. Petaluma.\textsuperscript{39} Jane Doe, a middle-school student in Petaluma, was subjected to sexually harassing remarks and behavior for many months by her peers.\textsuperscript{40} Doe reported the harassment to school officials, who promised to end the harassment but failed to do so.\textsuperscript{41} Nor did any school officials inform Doe or her parents of her rights under Title IX.\textsuperscript{42} In 1995, the Doe court theorized that Title VII principles might be applied in determining if a school had notice of peer harassment and failed to take appropriate corrective action.\textsuperscript{43}

However, cases that followed would further confound, rather than clarify, liability standards for schools faced with student-on-student sexual harassment claims. In Burrow v. Postville Community School District, one federal district court determined that a student may bring a Title IX cause of action against a school for its knowing failure to take appropriate remedial action in response to the hostile environment created by students at the school.\textsuperscript{44} In the same year, the

\begin{itemize}
\item \textsuperscript{34} 524 U.S. 274 (1998).
\item \textsuperscript{35} Id. at 292-93.
\item \textsuperscript{36} AAUW, supra note 8.
\item \textsuperscript{37} ERA, supra note 7.
\item \textsuperscript{38} 526 U.S. 629 (1999).
\item \textsuperscript{39} 54 F.3d 1447 (9th Cir. 1995).
\item \textsuperscript{40} Id. at 1449.
\item \textsuperscript{41} Id.
\item \textsuperscript{42} Id.
\item \textsuperscript{43} Id. at 1452.
\item \textsuperscript{44} 929 F. Supp. 1193, 1205 (N.D. Iowa 1996).
\end{itemize}
Fifth Circuit ruled in *Rowinsky v. Bryan Independent School District* that a school district should not be held liable for peer sexual harassment under Title IX unless the funding recipient directly committed the sexual harassment or the school district treated the sexual harassment of one gender more seriously than the sexual harassment of the other.⁴⁵ The Eleventh Circuit in *Davis v. Monroe County Board of Education* affirmed the district court’s decision to dismiss Davis’s claim on the ground that Title IX provides no private cause of action for peer sexual harassment.⁴⁶ This disagreement over the nature of school liability under Title IX gave the United States Supreme Court an opportunity to break the stalemate.

### III. The Davis Decision

The United States Supreme Court chose to address the issue of peer sexual harassment in the case of *Davis v. Monroe County Board of Education*.⁴⁷ As set forth in the complaint, a classmate of fifth-grader LaShonda Davis had subjected her to fondling, offensive comments, and abusive actions over a five-month period.⁴⁸ During that time, LaShonda’s mother pleaded for help from school officials, but no meaningful action followed.⁴⁹ One teacher allegedly refused, for more than three months, to allow LaShonda to change her assigned seat away from her tormentor.⁵⁰ The school lacked a sexual harassment policy and procedure that could have helped LaShonda find a way to remedy the sexual harassment.⁵¹ Eventually, LaShonda’s mother filed a criminal complaint against the harasser and filed suit against the school district.⁵² The harasser pleaded guilty to the criminal charge and finally, the harassment ceased.⁵³

In its *Davis* ruling, the Court followed the progression of most lower courts and decided against the view of the Fifth Circuit. The Court found that just as Title VII is violated if a sexually hostile working environment is created by co-workers and tolerated by the employer, Title IX is violated if a fellow student creates a sexually hostile educational environment and the supervising authorities knowingly fail to act to eliminate the harassment.⁵⁴

---

⁴⁵. 80 F.3d 1006 (5th Cir. 1996).
⁴⁶. 120 F.3d 1390, 1406 (11th Cir. 1997), rev’d and remanded, 526 U.S. 629 (1999).
⁴⁸. *Id.* at 633 (in reviewing the legal sufficiency of the cause of action, the Court “assume[d] the truth of the material facts as set forth in the complaint”) (quoting FED. R. CIV. P. 12(b)(6)).
⁴⁹. *Id.* at 633-34.
⁵⁰. *Id.* at 635.
⁵¹. *Id.*
⁵². *Id.*
⁵³. *Id.*
⁵⁴. *Id.* at 647.
The *Davis* ruling did not surprise the OCR. When the OCR issued its guidelines on the sexual harassment of students in March of 1997, it suggested that peer sexual harassment should be actionable under Title IX. The OCR criticized the Fifth Circuit as the odd-man-out in its deviation from other federal courts on the subject of school liability for student-on-student sexual harassment. The Fifth Circuit determined in *Rowinsky v. Bryan Independent School District* that a school district is not liable under Title IX for peer harassment unless the school district itself directly discriminates based on sex by responding differently to similar claims of sexual harassment by girls versus boys. The OCR and, ultimately, the United States Supreme Court believed that this decision was a misapplication of Title IX, as the OCR explained:

Title IX does not make a school responsible for the actions of the harassing student, but rather for its own discrimination in failing to take immediate and appropriate steps to remedy the hostile environment once a school official knows about it. If a student is sexually harassed by a fellow student, and a school official knows about it, but does not stop it, the school is permitting an atmosphere of sexual discrimination to permeate the educational program. The school is liable for its own action, or lack of action, in response to this discrimination. Notably, Title VII cases that hold that employers are responsible for remedying hostile environment harassment of one worker by a co-worker apply this same standard.

In *Davis*, the United States Supreme Court did not, however, ignore the potential for the courts to be flooded with peer sexual harassment cases. The growing body of research at that time showed that student-on-student sexual harassment was rampant in educational institutional settings, particularly in America’s colleges and universities. The Court thus narrowed the circumstances in which schools can be held liable and what actions constitute sex discrimination under Title IX. Justice O’Connor, writing for the majority, stated:

[F]unding recipients are properly held liable in damages only where they are deliberately indifferent to sexual harassment, of

---

56. Id.
57. 80 F.3d 1006, 1016 (5th Cir. 1996).
59. 526 U.S. at 648.
60. Id.
which they have actual knowledge, that is so severe, pervasive, and objectively offensive that it can be said to deprive the victims of access to the educational opportunities or benefits provided by the school.61

While Justice O'Connor narrowed the circumstances in which harassment may be actionable, it is clear that this ruling applies to all levels of education, including institutions of higher learning: “[R]ecipients of federal funding may be liable for ‘subject[ing]’ their students to discrimination where the recipient is deliberately indifferent to known acts of student-on-student sexual harassment and the harasser is under the school’s disciplinary authority.”62

Thus, women’s advocacy organizations had their victory: a recognized right under Title IX for student victims of peer sexual harassment to sue schools and districts that fail to respond to sex discrimination occurring under their noses.63 However, a carefully crafted standard was carved into the woodwork of Davis that so narrows the range of actionable conduct that only some victims are able to realize fully their after-the-fact right to be free from sexual harassment. The Davis holding can hardly be said to be a victory for student victims of sexual harassment, when the only victims who succeed under Davis are students utterly debilitated by the harassment. Davis does not equal the right to be free from sexual harassment, nor is Davis an effective tool to motivate educational institutions to participate in the effort to eliminate sexual harassment in our schools. Rather, Davis has been the glue that has held the educational status quo of general indifference in place.

A. The Actionable Right

The language of Title IX is short and sweet: “No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance.”64 Congress included an enforcement provision that empowered federal programs offering financial assistance to use “any . . . means authorized by law” to fulfill the congressional intent of Title IX.65

While the Court could have concluded that Title IX only bestows a federal right to terminate funds to a funding recipient in response to a violation of Title IX, the Court has interpreted the “any means”

61. Id. at 650.
62. Id. at 646-47.
63. ERA, supra note 7.
65. Id. § 1682.
provision as allowing the termination of funding as just one possible option. The Court has not, however, chosen to limit the federal government's power to enforce Title IX solely through the termination of funding. The Court justified its broader interpretation of Title IX by comparing the language in Title IX to the similar language in Title VI. Because the Court had already recognized an implied right of action in Title VI prior to congressional adoption of Title IX, it was appropriate to determine that Congress intended the similar language in Title IX to bestow the same private cause of action. Thus, Franklin's previous approval of the availability of monetary damages from such an action logically followed.

However, this right of a private cause of action is limited by the power under which Congress passed Title IX. The Court has treated Title IX as legislation passed under Congress's Spending Clause authority. While typically Congress cannot abrogate the States' Eleventh Amendment immunity under the Spending Clause, the United States Supreme Court has categorized Title IX funds as gifts to the States, through which States agree to waive their immunity from suit in exchange for the gifted funds.

However, the United States Supreme Court clarified that the "mere receipt of federal funds cannot establish that a State has consented to suit in federal court." Rather, Congress must manifest "a clear intent to condition participation" in the federal funding "on a State's consent to waive its constitutional immunity." Thus, to sue under Spending Clause provisions, funding recipients must have adequate notice that they could be liable for particular conduct. This contractual arrangement allows Congress to encourage certain behavior in exchange for federal funds, but requires Congress to "speak [in] a clear voice" to ensure an equal understanding of the terms of the agreement.

67. Id. at 694-96; see 42 U.S.C. § 2000(d) (1994) (stating within Title VI: "No person in the United States shall, on the ground of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance.").
68. Cannon, 441 U.S. at 694-96.
71. Id.; see U.S. CONST. art. 1, § 8, cl. 1 for origin of Congress's Spending Power, which provides in part: "The Congress shall have Power To lay and collect Taxes . . . to . . . provide for the . . . general Welfare of the United States."
76. Id.
77. Halderman, 451 U.S. at 17.
While the Monroe County Board of Education argued in *Davis* that school districts could not have anticipated liability stemming from student-on-student sexual harassment, the Court relied on its previous holdings in *Gebser* and *Franklin* to suggest that Title IX peer sexual harassment claims only seek to hold the school district liable for its own acts of subjecting students to sexual harassment in its educational programs. Thus, the person committing the acts of sexual harassment becomes less relevant (although not entirely irrelevant), while the reaction the school has to the student’s complaint of sexual harassment becomes key.

**B. The Davis Standard: A Heightened Standard for Student Victims**

1. **Severe, Pervasive, and Objectively Offensive**

   The issue in *Davis* was not whether sexual harassment is sex discrimination. Nor should the question have been what type of conduct constitutes sexual harassment. Rather, the issue in *Davis* was whether a recipient of federal funding for an educational program or activity may be held liable for damages under Title IX for sex discrimination in the form of student-on-student sexual harassment. The answer to this question was a no-brainer based on the Court’s previous line of Title IX decisions.

   O’Connor reaffirmed in *Davis* that the Court had “previously determined that ‘sexual harassment’ is ‘discrimination’ in the school context under Title IX” when a school acts with deliberate indifference to complaints of sexual harassment. The Court easily could have determined that the same actions sufficient to raise a hostile environment claim under Title VII, including demands for sexual favors, sexual advances, fondling, indecent exposure, and sexual assault, are equally sufficient to raise a hostile environment claim under Title IX for student-on-student sexual harassment when these...

---

79. *Id.*
80. The U.S. Supreme Court established that sexual harassment is a form of sex discrimination in *Cannon v. University of Chicago*, 441 U.S. 677 (1979).
82. 526 U.S. at 643.
83. Or at least Justice O’Connor’s opinion suggests the answer was easy. See *Gebser v. Lago Vista Indep. Sch. Dist.*, 524 U.S. 274 (1998) (rejecting agency principles, but defining standard for school reaction using the deliberate indifference standard to determine whether the school subjected the student to sexual harassment after receiving notice of teacher-student harassment); *Franklin v. Gwinnett County Pub. Schs.*, 503 U.S. 60, 76 (1992) (same). Of course, if Title IX serves to protect students from sex discrimination, it would be absurd to bar some people in the school community from engaging in such discriminatory conduct but not others.
84. 526 U.S. at 650.
acts are ignored by a school that has the authority and control to correct them.

Yet, in Davis, Justice O’Connor used the Spending Clause’s adequate notice requirement to define “discrimination” in a new light. While admitting that the Court had “elsewhere concluded that sexual harassment is a form of discrimination for Title IX purposes and that Title IX proscribes harassment with sufficient clarity to satisfy Pennhurst’s notice requirement and serve as a basis for a damages action,” O’Connor went on to conclude that Title IX’s other provisions “help give content to the term ‘discrimination’ in this context.” At that point, she started down the path of categorizing a special brand of sexual harassment specifically for students in the context of a Title IX action.

Justice O’Connor explained that “[s]tudents are not only protected from discrimination, but also specifically shielded from being ‘excluded from participation in’ or ‘denied the benefits of’ any ‘education program or activity receiving Federal financial assistance.” Her summary of Title IX suggests that it not only serves to protect students generally from sexual harassment as we have come to define it in the massive catalogs of sexual harassment cases, but also to protect students from discriminatory conduct that would exclude or obstruct their access to the same educational opportunities and benefits that all students are free to enjoy. This was likely the intent of Congress in passing Title IX.

However, O’Connor skipped over her own “not only” language to combine the two standards into one: Title IX only protects sex discrimination that serves to bar access to educational programs and activities. This simplification of standards ignores the portion of Title IX’s language that grants students the right to be free from exclusion from an educational program on the basis of sex or be subjected to discrimination under an educational program on the basis of sex. Justice O’Connor stated, “[t]he statute makes clear that, whatever else it prohibits, students must not be denied access to educational benefits and opportunities on the basis of gender,” but neglected to return to the question of what else the statute prohibits. The clear language of the statute, in allowing for several kinds of conduct to be

86. Davis, 526 U.S. at 649-50.
87. Id. at 650 (emphasis added).
88. Id. (emphasis added) (quoting 20 U.S.C. § 1681(a) (1994)).
89. Id.
90. As Senator Bayh, the sponsor of the amendment that created Title IX, stated: “[Title IX] is a strong and comprehensive measure which . . . is needed if we are to provide women with solid legal protection as they seek education and training for later careers.” 118 Cong. Rec. 5803, 5806-07 (1972) (statement of Sen. Bayh).
92. Davis, 526 U.S. at 650 (emphasis added).
violative of Title IX, demonstrates Congress's intention for Title IX to eliminate sexual harassment, not just sexual harassment that reaches such a severe level that it debilitating a student to the point of emotionally or physically barring access to his or her education.

Thus, based on this extremely narrow reading of Title IX's prohibited conduct, the Court felt the need to be “constrained to conclude that student-on-student sexual harassment, if sufficiently severe, can likewise rise to the level of discrimination actionable under the statute.”93 It is only when the sexual harassment is sufficiently severe that it can be said to bar access to educational opportunities and activities that the school must become involved to correct the acts. The reasoning is that Title IX only prohibits the school from denying educational opportunities, programs, and activities based on sex or gender. Thus, if the sexual harassment does not reach this level, it is not within the realm of Title IX protection.94

By placing the standard for the sexual harassment level of severity so high, it effectively destroys Title IX's ability to achieve Congress's goal, while simultaneously redefining what types of sexual harassment constitute sex discrimination for a certain subclass of victims. The students that Title IX is intended to protect often have few options for mobility, virtually no authority to correct disruptive behavior on their own, and are required to attend an interactive educational institution where they must face their harassers on a daily basis. However, Justice O'Connor in

93. Id. (emphasis added).
94. Id.
95. Id. at 648 (citation omitted).
97. Id. at 20.
98. Ellison v. Brady, 924 F.2d 872, 878 (9th Cir. 1991).
assessment is unlawful under Title VII, even without evidence that the harassment damaged the victim’s well-being.

_Harris_ involved an employee who left her employer after months of enduring crude remarks and propositions from the firm’s president.99 The employer attempted to argue that because Harris had no physical injury, nor any evidence of psychological damage, the conduct was not severe and pervasive enough to affect her working conditions. The Court found that the behavior the plaintiff endured would reasonably be harmful to women, whereas it probably would have been merely offensive to men and not so severe that it would affect work performance.100

In _Harris_, the Court attempted to find a compromise between making any conduct that is merely offensive actionable and requiring the conduct to cause a “tangible psychological injury.”101 The Court found that Title VII protects an employee from having to “endure sexual harassment until their psychological well-being is seriously affected to the extent that they suffer anxiety and debilitation.”102 Thus, the Court began to develop what later would be dubbed the “reasonable woman” standard, which allowed the Court to look at whether sexual conduct in the workplace reasonably could affect a woman being subjected to that type of sexual conduct. This standard did not rely on tangible injury but rather considered whether the conduct was the type that could reasonably lead a woman to suffer tangible injury.103

In _Harris_, the Court recognized the congressional purpose of Title VII was to protect employees from discrimination. However, in _Davis_, the Court forgot this important goal.104 Rather, we see schoolgirls having mental breakdowns from enduring daily threats, physical

---

99. _Harris_, 510 U.S. at 19.
100. _Id._
101. _Id._ at 21.
102. _Id._
103. This standard later became referred to as the “reasonable person” standard, but still considers whether the conduct is the type that could lead to tangible injury for a person similarly situated to the plaintiff.
104. _Davis v. Monroe County Bd. of Educ._, 526 U.S. 629, 652 (1999) (“The dissent fails to appreciate these very real limitations on a funding recipient’s liability under Title IX . . . we [do not] contemplate, much less hold, that a mere ‘decline in grades is enough to survive’ a motion to dismiss.”) (quoting _id._ at 677 (Kennedy, J., dissenting)); see Gabrielle M. _v. Park Forest-Chic. Heights, Ill. Sch. Dist._ 163, 315 F.3d 817, 828-29 (7th Cir. 2003) (Rovner, J., concurring) (noting that like in the employment context, where the Supreme Court has “firmly rejected any requirement that the victim of harassment suffer the equivalent of a nervous breakdown before she can recover under a hostile environment theory . . . a hostile environment should be actionable before it results in consequences so dramatic as hospitalization or leaving school.”).
withdraw from school, and escalating suicidal tendencies.\textsuperscript{105} This can hardly be the optimal goal Title IX was created to achieve.\textsuperscript{106}

2. \textit{Deliberate Indifference to Actual Notice}

Because the \textit{Davis} Court rejected the use of agency principles to support a claim of sex discrimination for peer sexual harassment under Title IX,\textsuperscript{107} the school cannot be held liable for the independent acts of third parties who are not in authoritative positions capable of effectuating change.\textsuperscript{108} Rather, as \textit{Gebser} established, an educational institution violates Title IX and can be liable for damages where it is “deliberately indifferent” to known acts of harassment.\textsuperscript{109}

This \textit{Davis} standard was borrowed directly from \textit{Gebser}, where the Court previously determined that a student could sue under Title IX for a school’s deliberate indifference to known acts of harassment by a teacher. While the Monroe County Board of Education tried to argue in \textit{Davis} that liability for sexual harassment by \textit{students} rather than teachers was beyond the scope of Title IX, the Court reiterated its rejection of agency principles in Title IX liability.\textsuperscript{110} The \textit{Davis} Court answered “whether the misconduct identified in \textit{Gebser}—deliberate indifference to known acts of harassment—amounts to an intentional violation of Title IX, capable of supporting a private damages action, when the harasser is a student rather than a teacher. We conclude . . . it does.”\textsuperscript{111}

The Court explained, as had the OCR previously, that the liability stems not from the fact that the sexual harassment occurred, but rather begins when the educational institution knows that the harassment is occurring and fails to respond.\textsuperscript{112} Thus, the response of the educational institution in the face of known sexual misconduct

\begin{flushright}
\textsuperscript{105} \textit{See} \textit{Davis}, 526 U.S. 629; Vance v. Spencer County Pub. Sch. Dist., 231 F.3d 253 (6th Cir. 2000); Murrell v. Sch. Dist. No. 1, 186 F.3d 1238, 1243 (10th Cir. 1999).
\textsuperscript{106} \textit{See} Cannon v. Univ. of Chi., 441 U.S. 677, 704 (1979) (Congress enacted Title IX with two principal objectives in mind: “[T]o avoid the use of federal resources to support discriminatory practices” and “to provide individual citizens effective protection against those practices.”).
\textsuperscript{107} \textit{See} \textit{Davis}, 526 U.S. at 643 (“As an initial matter, in \textit{Gebser}, we expressly rejected the use of agency principles in the Title IX context, noting the textual difference between Title IX and Title VII . . . (invoking agency principles on ground that definition of ‘employer’ in Title VII includes agents of employer),”) (citations omitted).
\textsuperscript{108} \textit{See id.} at 644 (“A recipient cannot be directly liable . . . where it lacks the authority to take remedial action.”).
\textsuperscript{109} \textit{Id.} at 643.
\textsuperscript{110} \textit{Id.} at 640-42.
\textsuperscript{111} \textit{Id.} at 643.
\textsuperscript{112} \textit{Id.} at 644-45 (stating that “[i]f a funding recipient does not engage in harassment directly, it may not be liable for damages unless its deliberate indifference ‘subject[s]’ its students to harassment. That is, the deliberate indifference must, at a minimum, ‘cause [students] to undergo’ harassment or ‘make them liable or vulnerable’ to it.”) (quoting \textit{RANDOM HOUSE DICTIONARY OF THE ENGLISH LANGUAGE} 1415 (1966)).
\end{flushright}
within the control and boundaries of the school is key in determining whether the school caused the student to be “subjected to discrimination under any education program or activity receiving Federal financial assistance.”

For the Court to reach a determination of whether an educational institution acted with deliberate indifference, a plaintiff must establish that actual notice was given to the educational institution. However, unlike in Title VII sexual harassment cases, the Davis Court did not require that the school have a formal school policy describing the method of giving actual notice (to put potential victims on notice of their rights and obligations) or describe the form of notice required. The only standard given in Davis was that “[d]eliberate indifference makes sense as a theory of direct liability under Title IX only where the funding recipient has some control over the alleged harassment. A recipient cannot be directly liable for its indifference where it lacks the authority to take remedial action.” Naturally, the federal circuit courts have used different methods to determine what this actual notice requirement means.

In Davis, LaShonda Davis, the student victim, made repeated reports to her classroom teacher and her mother. Upon inquiry, the teacher told the mother that the school principal had been informed of the incidents. No disciplinary action was taken subsequent to these reports. Following the reports, the sexual harassment continued. LaShonda reported these incidents to her physical education teacher. One week later, another sexually harassing incident occurred under the supervision of another teacher. LaShonda reported it to the supervising teacher and her mother again followed up. Eventually, with no action even attempted by the school to curb the behavior toward LaShonda, the primary harasser was charged with and pleaded guilty to sexual battery. In the end, LaShonda was subjected to five months of harassment without so much as a seat change made to move her away from her harasser. Finally, the harassment ceased, but only after LaShonda’s previously high grades had dropped and she had written a suicide note.

---

114. Davis, 526 U.S. at 644.
115. See discussion infra Part IV.B.
116. Davis, 526 U.S. at 633-34. The incidents at this point included peers attempting to touch LaShonda’s breasts and genital area and vulgar statements such as “I want to get in bed with you” and “I want to feel your boobs.” Id.
117. Id. Following the first set of reports, the sexual harassment included a student placing a doorstop in his pants followed by sexually suggestive mannerisms during physical education class. Id.
118. Id. at 635.
119. Id. at 634.
Because there were so many attempts made by both the student and the mother, the Davis Court had no difficulty determining whether there had been actual notice. Additionally, because the complaints continued over five months with the harassment ceasing only when the criminal justice system stepped in, it was clear that the school had not responded to the complaints in an effective manner. Nor did the school provide training on how to handle student complaints of sexual harassment, have a policy for students and parents to follow if they needed to report sexual harassment, or have any personnel designated to handle sexual harassment complaints. However, whether actual notice had been given becomes more difficult to determine in cases where parents fail to become involved, teachers fail to pass the reports on to authoritative personnel who have the power to make corrections, or fewer reports of harassment are made.

Once a plaintiff overcomes the hurdle of actual notice, the Davis Court established that the school must act with “deliberate indifference” to that complaint for Title IX liability to attach. However, the phrase is amorphous. While on one hand it suggests a school must be responsive in the face of a sexual harassment complaint, the Court took a step back, explaining that “courts should refrain from second-guessing the disciplinary decisions made by school administrators.” Rather, courts should declare a school deliberately indifferent “only where the recipient’s response to the harassment or lack thereof clearly unreasonable in light of the known circumstances.” This is not a reasonableness standard or a simple negligence standard. As Justice O’Connor explains, this standard does not actually require the school to “remedy” peer harassment, just to act in a manner not clearly unreasonable as a matter of law.

This student standard is so high that Justice O’Connor expressed doubt regarding whether LaShonda Davis would be able to show on remand that the school’s response to her five months of complaints was “clearly unreasonable.” O’Connor suggested that Davis may be able to show that because the school failed to respond in any way during those five months that the school acted “clearly unreasonable.” However, this twinge of doubt laid the foundation for the nearly impossible standard under which students could effectuate the broad and noble goals of Title IX.

120. Id. at 635.
121. Id. at 648.
122. Id.
123. Id. at 648-49.
124. Id. at 649 (stating that “it remains to be seen whether petitioner can show that the Board’s response to reports of G.F.’s misconduct was clearly unreasonable in light of the known circumstances”).
IV. COMPARATIVE ANALYSIS OF CIRCUIT COURT DECISIONS: DISPARITY AMONG THE CIRCUITS

Since the *Davis* ruling in 1999, the federal circuit courts have reviewed numerous peer sexual harassment cases. While not exactly the flood of litigation of which the dissent in *Davis* warned, there are enough cases to determine that the federal circuit courts of appeal are not entirely comfortable with the *Davis* standard. Although the federal circuit courts understand the general test set forth in *Davis*, they are struggling to define the vague terms within the *Davis* test. Accordingly, the courts have narrowly construed the *Davis* standard, cognizant of the unanswered questions. This conservativism has resulted in few winning student Title IX claims for either teacher-student or student-on-student sexual harassment.

The federal circuit courts, while understanding the general *Davis* standard, still vary in their presentation of the essential elements of *Davis* liability. While some circuits focus primarily on the specifically enumerated *Davis* elements—that Title IX liability requires that 1) the sexual harassment be so severe, pervasive, and objectively offensive that it could be said to deprive the plaintiff of access to the educational opportunities or benefits provided by the school; 2) the funding recipient had actual knowledge of the sexual harassment; and 3) the funding recipient was deliberately indifferent to the harassment—other circuits also include an element that the school district must have the power to exercise substantial control over both the harasser and the context in which the known harassment occurs.

However, how the federal circuits have defined each of these elements varies widely, with several circuits commenting on the lack of guidance in O’Connor’s *Davis* opinion. Consequently, the courts have been left to search in the pages of the *Davis* opinion for some guiding light. What they have been left with is a dissent in *Davis* that warns of a barrage of litigation that will drain taxpayer dollars

---

125. This includes the First, Sixth, Seventh, and Tenth Circuits. See Vance v. Spencer County Pub. Sch. Dist., 231 F.3d 253 (6th Cir. 2000); Wills v. Brown Univ., 184 F.3d 20 (1st Cir. 1999); Adusumilli v. Ill. Inst. of Tech., No. 98-3561, 1999 U.S. App. LEXIS 17954 (7th Cir. July 21, 1999); Murrell v. Sch. Dist. No. 1, 186 F.3d 1238 (10th Cir. 1999).

126. This includes the Eighth and Ninth Circuits. See P.H. v. The Sch. Dist. of Kan. City, 265 F.3d 653 (8th Cir. 2001); Reese v. Jefferson Sch. Dist. No. 14J, 208 F.3d 736 (9th Cir. 2000).

127. See, e.g., *P.H.*, 265 F.3d at 662 (“[T]he actual notice standard of *Gebser* has not yet been clearly defined.”); *Murrell*, 186 F.3d at 1252 (Anderson, J., concurring) (“The . . . majority wisely ‘decline[s] . . . to name job titles that would or would not adequately satisfy’ *Davis*’ requirement that the school have control over the harassing student[,] . . . leaving liability limited in general terms to cases involving ‘an official decision by the [Title IX] recipient not to remedy the violation.’”) (quoting *Davis*, 526 U.S. at 642); *Wills*, 184 F.3d at 31 (López, J., dissenting) (“This is a vexing case for many reasons. The facts are difficult. The applicable law is complex and evolving . . . with . . . unruly elements . . . .”).
and a majority opinion that attempts to persuade readers that it is sufficiently narrow to combat this misconception while still offering relief to victims.

A. Severe, Pervasive, and Objectively Offensive

Understanding that *Davis* requires sexual harassment to be “so severe, pervasive, and objectively offensive that it can be said to deprive the victims of access to the educational opportunities or benefits provided by the school,” the First, Sixth, Seventh, and Tenth Circuits have attempted to determine at what threshold sexual harassment becomes actionable sex discrimination under Title IX. While always a prong that must be satisfied, other circuits stopped short of this severe and pervasive inquiry upon finding other prongs of the *Davis* standard were not met. However, four circuits that have addressed the severe and pervasive prong represent the continuum on which most peer sexual harassment cases will fall: the Seventh Circuit found the conduct was not severe enough; the First Circuit found that while the one incident of sexual harassment was severe, without a second incident it could not be considered to pervade the student’s educational environment to the point of compromising her educational opportunities; and the Sixth and Tenth Circuits found the conduct to be egregious enough to be considered severe, pervasive, and objectively offensive enough to satisfy the prong.

The Seventh Circuit in *Adusumilli v. Illinois Institute of Technology* defined the severe and pervasive requirement through the use of scattered narrowing terms and phrases used in the *Davis* majority opinion. It stated that actionable conduct must be “severe and repeated . . . and must have a systemic effect.” The Seventh Circuit went on to explain that “[s]ingle incidents of student misconduct are unlikely to have such an effect . . . [s]ince in each instance the conduct ceased as soon as it occurred, and was not repeated.”

In *Adusumilli*, the student-plaintiff filed an action under Title IX for being subjected to sexual harassment on twelve separate occasions by four professors and six students. While some of these incidents were simply described by the court as “ogling” and “unwanted

---

128. *Davis*, 526 U.S. at 650.
130. See *Wills*, 184 F.3d 20.
133. *Id.*
134. *Id.*
135. *Id.* at *1.
touching," at least two incidents involved touching her breasts. Because the Seventh Circuit found the student had only reported two of the twelve incidents, the Court declined to address whether those unreported incidents were severe. The two incidents that were reported, including the touching of her shoulder by one student and the touching of her breast by another student, were analyzed under the Davis severity prong.

Because the Circuit Court refused to consider all of the incidents due to failure to give actual notice, the Court concluded that the two reported incidents were each single occurrences of isolated incidents that in themselves were not severe and did not permeate the student’s educational experience with sex discrimination. Since there were no future incidents, the effect of the harassment ceased with the conduct, and there was no action the school needed to take.

Under the Seventh Circuit’s interpretation of the Davis “severe and pervasive” requirement, after a report of sexual harassment is made, there must be repeated occurrences of sexual harassment by the same perpetrator to evidence harassment that is “severe and repeated” enough to cause a “systemic effect” coupled with a school that did nothing to keep the repeated instances from happening. This view of the severe and pervasive requirement suggests that it measures not whether the harassment in itself is objectively or subjectively severe, but rather, the effect the harassment has on the student after repeated instances.

136. Id. at *2.
137. Id. at *4.
138. Id.
139. See id.
140. Additionally, the Seventh Circuit recently went on to add that actionable conduct must be “so severe, pervasive, and objectively offensive that it has a ‘concrete, negative effect’ on the victim’s access to education.” Gabrielle M. v. Park Forest-Chicago Heights, Ill. Sch. Dist. 163, 315 F.3d 817, 821-23 (7th Cir. 2003) (emphasis added). However, the concurrence takes issue with the majority’s determination that “[a]lthough [Gabrielle M.] was diagnosed with some psychological problems, the record shows that her grades remained steady and her absenteeism from school did not increase. Nothing in the record shows that she was denied any educational opportunities by [the harassing student’s] actions.” Id. at 823. This concurrence explains that this “view of the way in which harassment can interfere with a student’s educational opportunities is too narrow . . . . Certainly at the kindergarten level, where learning social skills is at least as important as academic instruction, grades do not tell the complete story . . . .” in Title IX cases we have repeatedly rejected the notion that a victim’s ability to keep doing her job in the face of harassment will defeat her contention that the workplace was hostile.” Id. at 828 (Rovner, J., concurring). The concurrence then goes on to suggest that

]if anything, courts ought to be more flexible in assessing the harms that a child experiences as a result of harassment, given that children (especially young children) are far less able to articulate the fact and extent of their injuries and may manifest an array of different reactions to the harassment . . . . Neither she nor future victims of school place harassment should be penalized simply because they seem resilient.
The First Circuit in *Wills v. Brown University* defined the *Davis* severity requirement as having a lower threshold than that of the Seventh Circuit.\textsuperscript{141} The First Circuit attempted to define Title IX hostile environment sexual harassment in light of Title VII’s more developed standard, as the Court stated:

> The rubric ‘hostile environment’ applies where the acts of sexual harassment are sufficiently severe to interfere with the workplace or school opportunities normally available to the worker or student . . . . Broadly speaking, a hostile environment claim requires the victim to have been subjected to harassment severe enough to compromise the victim’s employment or educational opportunities . . \textsuperscript{142}

Thus, the First Circuit defined the severe and pervasive requirement as simply “compromising” a student’s educational opportunities, rather than barring or denying educational opportunities. However, although the First Circuit appeared to present a lower severe and pervasive standard for Title IX liability than *Davis* set forth, the student’s claim in *Wills* did not survive the legal analysis actually applied in her case.\textsuperscript{143}

Wills sued Brown University after her chemistry teacher inappropriately touched her.\textsuperscript{144} Wills had approached her teacher after having difficulty with her organic chemistry class.\textsuperscript{145} While purporting to pray with her in his office, the teacher pulled Wills into his lap and fondled her breasts under her shirt.\textsuperscript{146} Wills immediately met with the university official responsible for administering complaints of sexual harassment to report the incident.\textsuperscript{147} In response to the complaint, Brown University officials placed the teacher on probation and issued a written reprimand warning against another such incident.\textsuperscript{148} The university did not inflict harsher punishment because they believed it to be the teacher’s first incident.\textsuperscript{149} They were wrong.\textsuperscript{150} Not only did the teacher have a string of complaints prior to Wills’s incident, but he continued similar behavior following the rep-

\textsuperscript{141} 184 F.3d 20 (1st Cir. 1999).
\textsuperscript{142} Id. at 25-26.
\textsuperscript{143} Id. at 31.
\textsuperscript{144} Id. at 23.
\textsuperscript{145} Id.
\textsuperscript{146} Id.
\textsuperscript{147} Id.
\textsuperscript{148} Id.
\textsuperscript{149} Id.
\textsuperscript{150} Id. at 24.
rimand with no further action taken. Only two months later, Brown University renewed his contract for another year and awarded him a raise.

Wills’s Title IX claim was unsuccessful because her claim “was of a single specific harassment incident that occurred before the reprimand and the later complaints (albeit one that caused continuing damages).” While Wills attempted to show that Brown University was subjecting her and other students to sex discrimination through a general indifference to continuing sexual harassment, the First Circuit held that subsequent reports by other students of similar harassment by the same teacher were not relevant. Rather, “absent a second physical assault by [the teacher] on Wills, or some form of direct harassment, Wills had no claim for sex discrimination against Brown occurring after [the first incident].”

Yet, as the dissent in Wills pointed out, the majority failed to appreciate “[t]he proposition that the presence of a harasser can rise to the level of hostile environment sex discrimination [which] finds support in the Title VII context.” The dissent encouraged the majority to consider adopting an objective standard of whether a student in the plaintiff’s position “would find [the teacher’s] mere presence at [the school] created a hostile environment.”

While O’Connor in Davis cautioned against lower courts finding schools liable for a single instance of even sufficiently severe sexual harassment, O’Connor limited private damages to situations “having a systemic effect on educational programs or activities,” not simply on one individual student. As O’Connor states, “[e]ven the dissent suggests that Title IX liability may arise when a funding recipient remains indifferent to severe, gender-based mistreatment played out on a ‘widespread level’ among students.” There is room under Davis for the First Circuit to have determined that Brown’s general attitude toward student complaints of sexual harassment by a particular teacher was enough to establish a severe permeation of sex discrimination throughout the program. By defeating an individual

151. Id.
152. Id.
153. Id. at 26-27.
154. Id.
155. Id. at 33 (Lipez, J., dissenting) (alteration in original) (describing the position taken by the majority of the court).
156. Id. at 38 (Lipez, J., dissenting) (alteration in original) (referring to Ellison v. Brady, 924 F.2d 872, 883 (9th Cir. 1991), where an employer’s decision to allow one employee who had formerly harassed a female co-worker to transfer back into her office after a six-month “cooling off period,” created a hostile working environment).
157. Id. (creating a “reasonable female student” standard by harkening to the “reasonable woman standard” used in Harris v. Forklift Sys., Inc., 510 U.S. 17 (1993)).
159. Id.

student’s claim on the basis that her single instance of individual sexual harassment was not severe enough to result in a systemic effect on that individual, harassers can harass indefinitely, so long as they never harass the same student more than once.

The result in Wills is important because it is a Title IX teacher-student sexual harassment claim that fails. Though the First Circuit has yet to decide a Title IX peer sexual harassment case, the result will likely be similarly disheartening without facts as strong as those found in Davis. This hypothesis finds support in Davis, as O’Connor explained that:

> The fact that it was a teacher who engaged in harassment in Franklin and Gebser is relevant. The relationship between the harasser and the victim necessarily affects the extent to which the misconduct can be said to breach Title IX’s guarantee of equal access to educational benefits and to have a systemic effect on a program or activity. Peer harassment, in particular, is less likely to satisfy these requirements than is teacher-student harassment.

The only federal circuit courts to find peer sexual harassment severe and pervasive enough to satisfy the Davis standard are the Sixth and Tenth Circuits. Unlike the First and Seventh Circuits, these courts look to the facts of Davis, rather than the terminology, to define what conduct can be considered severe and pervasive. In the cases decided to date, these circuits considered the pattern of behavior, its tangible effects on the student victims, and then compared them with Davis.

In Vance v. Spencer County Public School District, a student, Alma McGowen, experienced physical and verbal sexual torment over the course of several years, including being stabbed in the hand, being held by several classmates while others tried to rip off her clothes, and being subjected to continuous verbal sexual comments, even after a detailed complaint had been filed with the school’s Title IX coordinator. No investigation resulted, and school officials continued to use the same ineffective method of merely discussing the incidents with the perpetrators. Typically, following these discussions, the sexual harassment of Alma by her classmates would escalate.

In Murrell v. School District No. 1, a developmentally and physically disabled student, Penelope Jones, was allegedly subjected to sustained sexual harassment, including sexual assault and battery,

---

160. 184 F.3d 20 (1st Cir. 1999).
162. 231 F.3d 253 (6th Cir. 2000).
163. Id. at 256. The court seems to imply the lack of reaction may have stemmed from one of the perpetrators being a school board member’s son.
164. Id.
over the course of one month.\textsuperscript{165} After being warned by the assaulting student’s mother of his sexually aggressive tendencies and after teachers became aware that he was, in fact, engaging in sexually aggressive behavior toward Penelope, the school allowed the harasser to act as a janitor’s assistant, granting him special access to unsupervised areas of the school.\textsuperscript{166} It was in this capacity that he took Penelope to a secluded area and sexually assaulted her.\textsuperscript{167} Due to the escalation of the harassment, Penelope became self-destructive and suicidal and entered a psychiatric hospital.\textsuperscript{168} Upon her release and return to school, Penelope was back only one day when she was once again battered by her former harasser and humiliated by the other students who knew about the earlier sexual attacks.\textsuperscript{169}

Similar to \textit{Davis}, \textit{Vance} and \textit{Murrell} had a pattern of escalating sexual behavior committed by the same peers each time that resulted in a serious tangible physical effect on the student victim. In \textit{Davis}, LaShonda’s grades had dropped and she had written a suicide note.\textsuperscript{170} In \textit{Vance}, Alma had to complete her studies at home after being diagnosed with depression.\textsuperscript{171} In \textit{Murrell}, Penelope suffered self-destructive and suicidal tendencies necessitating entering a psychiatric hospital.\textsuperscript{172}

These cases are clearly severe, and neither the Sixth nor Tenth Circuits had difficulty casting them as such. However, it seems odd that Title IX would require harassment to reach this level of extremity.\textsuperscript{173} Not only is the \textit{Harris} purpose left by the wayside,\textsuperscript{174} but also Title IX is likely to have little impact upon schools that think nothing this extreme could ever happen under their control. If Title IX has any purpose, it must be to eradicate sex discrimination in federally funded educational institutions, not merely to compensate student victims who have been driven over the brink. For this goal to be realized, the severe and pervasive standard must allow student victims to succeed in court without having to suffer a mental breakdown. This may be the only way to persuade schools to react after the very first complaint, before any more damage can be inflicted.

\textsuperscript{165} 186 F.3d 1238, 1243 (10th Cir. 1999) (construing the facts as true to review the lower court’s granting of a motion to dismiss).
\textsuperscript{166} Id.
\textsuperscript{167} Id.
\textsuperscript{168} Id.
\textsuperscript{169} Id. at 1244.
\textsuperscript{170} 526 U.S. 629, 634 (1999).
\textsuperscript{171} 231 F.3d 253, 257 (6th Cir. 2000).
\textsuperscript{172} 186 F.3d at 1244.
\textsuperscript{173} Id. at 1252 (Anderson, J. concurring) (“The allegations in this case are . . . egregious. . . . Whether less egregious facts will suffice in future cases remains to be seen.”).
\textsuperscript{174} Harris v. Forklift Sys., Inc., 510 U.S. 17 (1995) (holding that employees should not be made to endure sexual harassment until their psychological well-being is seriously affected to the extent that they suffer anxiety and debilitation).
B. Actual Notice

Unlike the Sixth and Tenth Circuits, where there was clear and ample actual notice to high-ranking school authorities, the question of actual notice to school officials defeated student claims in the Eighth and Ninth Circuits. Unlike other circuits, the Eighth and Ninth Circuits each focused on whether a district exercised substantial control over both the harasser and the context in which the harassment occurred as an ancillary prong simultaneously analyzed with the actual notice requirement. While most circuits have not formally added this fourth prong to the Davis test, O'Connor did suggest in Davis that “[d]eliberate indifference makes sense as a theory of direct liability under Title IX only where the funding recipient has some control over the alleged harassment. A recipient cannot be directly liable for its indifference where it lacks the authority to take remedial action.”

The Ninth Circuit, in Reese v. Jefferson School District No. 14J, determined that the students suing for Title IX sexual harassment did not give actual notice to school officials who had the power to remedy the harassment. The students, who had already graduated, sued the school for ongoing harassment during high school by a group of male peers. However, the Ninth Circuit held that although there may have been evidence that a teacher witnessed conduct that may have put the school on notice, this was not sufficient to establish that actual notice had been given to an official with the power to correct the harassment. Thus, the Ninth Circuit defined the actual notice requirement narrowly to include direct reporting to an official with the authority to effectuate change. While other circuits have questioned whether notice to a teacher, who either has limited control over the immediate conduct or can pass the information on to an official with authority, should satisfy the actual notice requirement, the Ninth Circuit appears to have determined that actual notice is both a formalistic and substantive standard.

The Eighth Circuit has taken an even more narrow view of the Davis actual notice requirement. In P.H. v. School District of Kansas City, where a student sued for sex discrimination under Title IX for a two-year sexual relationship with a teacher, the Eighth Circuit defined actual notice as an actual notice-plus standard: “a school district must have had actual notice of a teacher’s sexual harassment of a student and the school district must have made an official decision

175. 526 U.S. at 644.
176. 208 F.3d 736, 740-41 (9th Cir. 2000).
177. Id. at 738.
178. Id. at 740.
179. 265 F.3d 653 (8th Cir. 2001).
not to remedy the violation in order for liability to attach to the school district.”\textsuperscript{180} While P.H. argued that the actual notice standard had not yet been clearly defined and suggested that teachers could in some situations have the necessary control to take corrective action, the Eighth Circuit held that no constructive notice is permissible under \textit{Davis}; rather, the notice must be actual and to a school official.\textsuperscript{181}

From the current case law guidance, to be successful under a Title IX peer sexual harassment claim there must be actual notice to an official with the power to address the complaint. However, because \textit{Davis} never required schools to maintain sexual harassment policies and reporting procedures, it is questionable whether students and parents will understand either their rights or their obligations under \textit{Davis}. Additionally, it is questionable whether a student’s individual complaint to a teacher, the most logical person to whom a student would report incidents, may establish the actual notice required to seek damages for continued sexual harassment. Rather, students, regardless of age, will have to recognize that their complaints must be made to a high-ranking official of the school. Thus far, this person has been the school principal in the cases that have been successful.\textsuperscript{182} The \textit{Murrell} concurrence criticizes \textit{Davis}, as it points to Justice Kennedy’s dissent, which stated: “‘[T]he majority says not one word about the type of school employee who must know about the harassment before it is actionable.’”\textsuperscript{183} If the student is more comfortable discussing such matters with a teacher or guidance counselor and is told the situation will be corrected, but that complaint never reaches the principal, it is possible that even if the sexual harassment continues, the school will bear no responsibility for its inaction. This appears to be the standard irrespective of whether the principal has made any effort to require faculty, when a complaint is brought to their attention, to instruct students of their rights and responsibilities and encourage the student to report the incident to the principal.

\textbf{C. Deliberate Indifference}

\textit{Davis} defined deliberate indifference as the equivalent to not acting “clearly unreasonable,”\textsuperscript{184} a phrase the courts have had to struggle to define. Under the two successful circuit level student peer sexual harassment Title IX claims, the phrase “deliberate indifference”

\begin{footnotesize}
\textsuperscript{180} Id. at 661.
\textsuperscript{181} Id. at 663.
\textsuperscript{182} See \textit{Murrell} v. Sch. Dist. No. 1, 186 F.3d 1238, 1252 (10th Cir. 1999) (Anderson, J., concurring) (\textit{Davis} did not answer this question precisely, leaving liability limited in general terms to cases involving ‘an official decision by the [Title IX] recipient not to remedy the violation.’) (quoting \textit{Davis} v. Monroe County Bd. of Educ., 526 U.S. 629, 642 (1999)).
\textsuperscript{183} Id. (quoting \textit{Davis}, 526 U.S. at 679 (Kennedy, J., dissenting)).
\textsuperscript{184} 526 U.S. at 630.
\end{footnotesize}
was defined as a two-step inquiry into 1) whether a school took any steps to address the complaint of sexual harassment;\(^{185}\) and if so, 2) whether the steps taken were not clearly unreasonable steps to address the complaint of sexual harassment.\(^{186}\)

However, in at least two losing Title IX student actions, the First and Sixth circuits only looked to whether the school took any remedial steps, without considering the effectiveness or timing.\(^{187}\) In *Wills v. Brown University*,\(^{188}\) because the university took some action after Wills’s complaint, and because a second incident did not take place, the First Circuit found that the university did take action when it issued the chemistry teacher a letter of reprimand and instructed him not to do it again or be fired.\(^{189}\) However, the crux of Wills’s argument was that the only reason this incident occurred was because the university had failed to respond to previous students’ complaints about the same teacher. Had the university responded to the sexual harassment at that time, the teacher would not have continued to harass students.\(^{190}\) Additionally, Wills argued that because other students were subsequently harassed, she should be permitted to admit evidence that even after her complaint, still the university acted with deliberate indifference to the continued sexual harassment.\(^{191}\) The First Circuit, however, held that the incidents following Wills’s complaint were irrelevant as to her claim, and because no second incident occurred to her, the university’s letter of reprimand must have worked to protect Wills.\(^{192}\)

The Sixth Circuit in *Soper v. Hoben*\(^{193}\) also held for the school district when remedial steps were taken after the rape of a special education student by three of her classmates at school and on the bus.\(^{194}\)

The Sixth Circuit in the *Soper* majority opinion stated that:

> [P]laintiffs have failed to present any evidence of deliberate indifference attributable to defendants. Once they did learn of the inci-

\(^{185}\) See *Murrell*, 186 F.3d at 1248 (stating that the school “had actual knowledge . . . from almost the moment it began to occur, and not only refused to remedy the harassment but actively participated in concealing it”).

\(^{186}\) See *Vance v. Spencer County Pub. Sch. Dist.*, 231 F.3d 253, 262 (6th Cir. 2000) (“Spencer continued to use the same ineffective methods to no acknowledged avail. Although ‘talking to the offenders’ produced no results, Spencer continued to employ this ineffective method. . . . However, the harassing conduct not only continued but also increased as a result.”).


\(^{188}\) 184 F.3d 20.

\(^{189}\) Id. at 23.

\(^{190}\) Id. at 26.

\(^{191}\) Id.

\(^{192}\) Id. at 26-27.

\(^{193}\) 195 F.3d 845 (6th Cir. 1999).

\(^{194}\) Id.
dents, they quickly and effectively corrected the situation. Defendants immediately contacted the proper authorities, investigated the incidents themselves, installed windows in the doors of the special education classroom, placed an aide in Harmala’s classroom, and created student counseling sessions concerning how to function socially with the opposite sex.\textsuperscript{195}

However, as the Soper partial dissent points out, these are admirable steps for the school to take, but they can be considered immediately responsive only if the final rape is considered the lone reported incident.\textsuperscript{196} Prior to the rape, however, there were known incidents reported to teachers: earlier sexual advances on the plaintiff by one of the boys, and the victim’s mother’s requests for the two students to never be left alone together unsupervised.\textsuperscript{197} Although the victim’s mother was assured care would be taken, “no steps were actually taken to minimize or stop the harassment. The specific request that Renee not be alone in the presence of Boy A was ignored. Arguably, these actions amounted to deliberate indifference to the concerns about harassment brought to Renee’s teachers by her mother.”\textsuperscript{198}

The two-step process used in Murrell and Vance is superior for defining the Davis deliberate indifference standard. Courts must ask whether any response was made following the initial complaint of sexual harassment, and whether that response was effective in deterring continued sexual harassment. Otherwise, useless remedial efforts, or efforts that come too late to protect a student from seriously debilitating acts of sexual harassment, become a loophole through which schools may escape Title IX liability.

V. CONCLUSION: LINGERING QUESTIONS AFTER DAVIS

Justice O’Connor’s majority opinion in Davis sets forth a standard under which students have had difficulty winning their Title IX peer sexual harassment claims.\textsuperscript{199} While reiterating the Court’s rejection of the use of agency principles, O’Connor concluded that for federally funded educational institutions to be liable for peer sexual harassment, they must have actual notice of, and act deliberately indifferent to, sexual harassment that is so severe, pervasive and objectively offensive that it can be said to deprive the victims of access to the educational opportunities or benefits provided by the school.\textsuperscript{200} O’Connor stated that the notice requirement “cabins the range of

\textsuperscript{195} Id. at 855.
\textsuperscript{196} Id. at 857 (Moore, J., dissenting in part).
\textsuperscript{197} Id. at 848.
\textsuperscript{198} Id.
\textsuperscript{199} Davis v. Monroe County Bd. of Educ., 526 U.S. 629, 650 (1999).
\textsuperscript{200} Id. at 650.
misconduct that the statute proscribes”201 and that all of the required factors under this new standard “combine to limit a recipient’s damages liability to circumstances wherein the recipient exercises substantial control over both the harasser and the context in which the known harassment occurs.”202

While the symbolic significance of granting students the ability to be awarded monetary damages for being subjected to sexual harassment is great, the practical reality is that the federal circuit courts have been careful to construe O’Connor’s standard narrowly, constructing a high hurdle for students to overcome. This has led to the dismissal of many students’ Title IX claims even though the student has been subjected to what should amount to actionable sexual harassment.

Distillations of the Davis standard vary from circuit to circuit, and questions remain unanswered as to the form of the actual notice necessary, to whom the notice must be given, and the level of control required by the school official. Additionally, courts will continue to struggle with determining whether one incident, even if a severe sexually violent act, can ever be sufficient under Davis, and whether the severity requirement should be viewed through a reasonable student perspective, the eyes of the school official accepting the complaint, or an objective standard based on tangible injury to the harassed student. How extreme post-notice harassment must be to show that an educational institution has acted with deliberate indifference remains a disturbingly high threshold, and whether the Davis standard is dependent on educational level or age remains unclear.203

201. Id. at 644.
202. Id. at 645.
203. In Gabriele M. v. Park Forest-Chicago Heights, Ill. Sch. Dist. 163, 315 F.3d 817, 821-23 (7th Cir. 2003), the court stated that “[t]here is a threshold question, altogether reasonable and rational, of whether a five or six year old kindergartener can ever engage in conduct constituting ‘sexual harassment’ or ‘gender discrimination under Title IX. Common sense, at least, would reject any such extension of Title IX.” While the Seventh Circuit determined it need not answer this question and assumed arguendo that the conduct alleged was sexual harassment, the concurrence took issue with the majority’s comment. While Davis did acknowledge simple acts of teasing and name calling, and actionable ages of the children involved and the likelihood they were unaware of the sexual nature of their behavior, with whether the harassment had a “concrete, negative effect” on the victim’s education. The majority concluded that because the children “were” not engaging in knowingly sexual acts . . . (at a minimum) [this] detracts from the severity and offensiveness of their actions.” Id. The concurrence compellingly attacks the majority’s proposition on three grounds. First, “[i]t is the school district, not [the harasser], that is charged with liability . . . whatever the children’s comprehension may have been, the adults charged with their care and education had the ability to appreciate the inappropriate and potentially harmful nature of the conduct.” Id. at 826-27 (Rovner, J., concurring). This appreciation that a student is harassing another student is what triggers the school district’s liability, not the harasser’s intent or full awareness of the inappropriateness of the behavior. Second, while children may not fully appreciate the sexual nature of the conduct, “harassing conduct need not be motivated by sexual desire, nor must it be overtly
Until Congress defines the true nature of Title IX, Davis will continue to spawn inconsistency and subjectivity in the federal circuit courts and will continue to serve as more of an obstacle than a tool for student victims of sexual harassment in federally funded educational institutions.