ARTICLES

FIXING A HOLE: HOW THE CRIMINAL LAW CAN BOLSTER REPARATIONS THEORY

Eric L. Muller

[pages 659–704]

Abstract: High-profile popular-press authors recently have challenged the mainstream consensus that certain historical events should be condemned as injustices. These authors argue that such condemnation unfairly imposes modern standards on historical actors. Until now, the redress movement has largely ignored these partisan revisionists who have sought to justify the harmful decisions made by past generations. Such revisionism, however, threatens the very foundation of reparations theory by persuading the public that redress is unnecessary because historical figures actually committed no injustice by merely acting appropriately, given the historical context in which they lived. This Article seeks to initiate a dialogue regarding how to approach the task of defining a historical injustice. The Article draws an analogy to the criminal law’s “cultural defense,” proposing a framework by which legal scholars may fairly judge the wrongdoing of historical actors. Although the analogy between foreign cultures and historical eras is imperfect, it presents a useful starting point to stimulate critical discussion about how best to address the growing structural weakness in the foundation of reparations theory.

MISSING THE FOREST FOR A TREE: UNPUBLISHED OPINIONS AND NEW FEDERAL RULE OF APPELLATE PROCEDURE 32.1

Scott E. Gant

[pages 705–736]

Abstract: On December 1, 2006, Federal Rule of Appellate Procedure 32.1 will take effect, allowing citation to all opinions issued on or after January 1, 2007 that have been designated “unpublished” or “non-precedential.” The new Rule, under consideration by the Judicial Confer-
ence of the United States since the 1990s, seemingly puts an end to the long and sometimes contentious debate over whether citation to unpublished opinions should be permitted. But the Rule does not address a more important issue: whether the federal courts of appeals should designate some of their opinions as nonprecedential. This Article argues the notion that judges can and should determine an opinion’s precedential value at the time they issue it is based upon a flawed and outdated view of how the law develops. Whether an opinion has made “new law” or is otherwise significant is a judgment best made with the benefit of time, and with input from lawyers, litigants, and other judges.

NOTES

THE COMMERCE CLAUSE MEETS ENVIRONMENTAL PROTECTION: THE COMPENSATORY TAX DOCTRINE AS A DEFENSE OF POTENTIAL REGIONAL CARBON DIOXIDE REGULATION

Heddy Bolster

[pages 737–772]

Abstract: On December 20, 2005, seven northeastern states signed the Regional Greenhouse Gas Initiative (“RGGI”), an agreement aimed at reducing greenhouse gas pollution from power plants. Once enacted by each state’s legislature or rulemaking agencies, this agreement will establish a “cap-and-trade” program to cap greenhouse gas emissions within the region and allow power plants to trade emissions allocations. This program faces a significant challenge, however. Electricity suppliers within the region may import power from outside the regulated region to avoid the constraints of the emissions cap, resulting in little or no net decrease in overall emissions—a problem known as “leakage.” Because limiting emissions imports would inevitably place burdens on the interstate trade of electricity, the regulatory approaches available to RGGI states to limit imports of power from unregulated regions may be subject to attack as violations of the Commerce Clause of the U.S. Constitution. This Note explores the possibility of applying the concepts embodied in the compensatory tax doctrine to defend a regulatory scheme that the RGGI states might employ to combat leakage.
REDEFINING PROPERTY UNDER THE DUE PROCESS CLAUSE: TOWN OF CASTLE ROCK v. GONZALES AND THE DEMISE OF THE POSITIVE LAW APPROACH

Joel Hugenberger

[pages 773–814]

Abstract: Since Board of Regents of State Colleges v. Roth, the U.S. Supreme Court has defined property for due process purposes as a legitimate claim of entitlement rooted in a source of law independent from the Constitution and has recognized a broad variety of property interests. In 2005, however, the Court in Town of Castle Rock v. Gonzales reined in its due process property jurisprudence in determining that a court-ordered restraining order did not create property because its language was discretionary. The Court also suggested that no police protection statute, however worded, could ever constitute property because its enforcement lacks ascertainable monetary value and only indirectly benefits the protected person. This Note argues that the Court should refine the definition of property as a benefit rooted in a source of law independent from the Constitution that is conferred on a specific class subject to specific conditions and terminable only under specific conditions.

LIFE-SUSTAINING TREATMENT LAW: A MODEL FOR BALANCING A WOMAN’S REPRODUCTIVE RIGHTS WITH A PHARMACIST’S CONSCIENTIOUS OBJECTION

Natalie Langlois

[pages 815–852]

Abstract: In recent years, there have been several incidents where pharmacists refused to dispense prescriptions for emergency contraception, as well as other types of contraceptives, because of their ethical, moral, or religious beliefs. State law has attempted to address this problem in various ways, but frequently fails to balance adequately the rights of a woman to access lawful contraceptive prescriptions against a pharmacist’s right to conscientiously object. This Note argues that pharmacist refusal laws should seek guidance from a similar conflict in the life-sustaining treatment context. Life-sustaining treatment law permits a health care provider to refuse to comply with a patient’s decision regarding life-sustaining treatment, but imposes additional duties on the health care provider who does so. These additional duties—requiring the health care provider to notify the patient of its policy in advance
and transfer the patient to another health care facility—prevent both the patient’s and health care provider’s rights from being compromised. The Note concludes that analogous transfer and notice requirements should be placed on pharmacists who conscientiously object to dispensing contraceptive prescriptions.

A HOSTILE ENVIRONMENT: HOW THE “SEVERE OR PERVERSIVE” REQUIREMENT AND THE EMPLOYER’S AFFIRMATIVE DEFENSE TRAP SEXUAL HARASSMENT PLAINTIFFS IN A CATCH-22

Evan D. H. White

[pages 853–890]

Abstract: This Note argues that the combination of the “severe or pervasive” requirement and the employer’s affirmative defense, as applied in lower federal courts, makes it very difficult for a hostile work environment sexual harassment plaintiff to prevail against his or her employer. Courts require actionable harassment to consist of one extremely severe incident or to continue long enough to become cumulatively severe or pervasive. But once the harassment has gone on long enough to become severe or pervasive, the employer’s affirmative defense is increasingly likely to bar the plaintiff’s prima facie case. The result is that as the plaintiff’s prima facie case grows stronger, the probability that the employer will prevail on its affirmative defense also increases. This Note argues that such a contradictory approach is unfair to plaintiffs and proposes a more equitable standard that would place plaintiffs and employers on equal footing.
FIXING A HOLE: HOW THE CRIMINAL LAW CAN BOLSTER REPARATIONS THEORY

ERIC L. MULLER*

Abstract: High-profile popular-press authors recently have challenged the mainstream consensus that certain historical events should be condemned as injustices. These authors argue that such condemnation unfairly imposes modern standards on historical actors. Until now, the redress movement has largely ignored these partisan revisionists who have sought to justify the harmful decisions made by past generations. Such revisionism, however, threatens the very foundation of reparations theory by persuading the public that redress is unnecessary because historical figures actually committed no injustice by merely acting appropriately, given the historical context in which they lived. This Article seeks to initiate a dialogue regarding how to approach the task of defining a historical injustice. The Article draws an analogy to the criminal law’s “cultural defense,” proposing a framework by which legal scholars may fairly judge the wrongdoing of historical actors. Although the analogy between foreign cultures and historical eras is imperfect, it presents a useful starting point to stimulate critical discussion about how best to address the growing structural weakness in the foundation of reparations theory.

The past, they say, is another country.
—Rhys Isaac

INTRODUCTION

Over approximately the past twenty years, scholars have built an elaborate theoretical structure of reparations for historical injustices. These scholars have paid careful attention to issues of causation, methods of compensation and restitution, the problem of privity between wrongdoers and victims, the necessity of apology, and the viabil-

* © 2006, Eric L. Muller, George R. Ward Professor, University of North Carolina School of Law. I had helpful conversations about many of the ideas in this Article with Bill Marshall, Kay Levine, Deborah Weissman, Joe Kennedy, Melissa Jacoby, Maxine Eichner, Scott Baker, Andrew Chin, Richard Myers, Leslie Branden-Muller, Abby Muller, and Nina Muller.

ity of reconciliation as an alternative or supplement to reparation.\(^2\) Naturally, disagreements about the design of the structure remain, and considerable construction work remains to be done. But there is no doubt that a sizable scholarly structure now stands in what was once an empty field.

That structure, however, has an unseen hole in its foundation which threatens the integrity of the entire structure that sits atop it. The hole has been opened by partisan historical revisionists who deliberately manipulate and misrepresent historical evidence for political purposes.\(^3\) The most notorious example of this sort of revisionism is the continuing efforts of some to undermine settled accounts of the existence, nature, and scope of Nazi atrocities in Europe during World War II.\(^4\)

Recently, high-profile partisan revisionism has crossed the Atlantic and begun to poke at the foundations of American historiography. In the last several years, no fewer than three works of partisan revisionism have climbed high on the New York Times best-seller list: Ann Coulter’s 2003 book, *Treason*,\(^5\) which attempts, among other things, to justify the excesses of McCarthyism; Michelle Malkin’s 2004 book, *In Defense of Internment*,\(^6\) which seeks to justify the racial incarceration of some 50,000 Japanese aliens and some 70,000 U.S. citizens of Japanese ancestry during World War II; and Thomas E. Woods’s 2005 book, *The Politically Incorrect Guide to American History*,\(^7\) which tries, among many other things, to establish the virtue of Southern secessionism and to defend the power of states to permit slavery.\(^8\) In each instance, the author has a clear partisan purpose for his or her retelling of the episode in question: Coulter seeks to establish the Democ-

---


3 Partisan historical revisionism must be distinguished from the healthy process of revising historical understanding to take into account newly discovered evidence. Responsible and dispassionate consideration of new evidence is the essence of good historical practice; it ensures that the study of history remains fresh and vibrant.

4 David Irving is probably the best-known representative of this movement. See Deborah Lipstadt, *Denying the Holocaust: The Growing Assault on Truth and Memory* 8 (1994).


8 See id. at 43–75.
ratic Party as a party of traitors to the United States, Malkin seeks to advance an agenda of aggressive racial and religious profiling against Arabs and Muslims, and Woods seeks to further the efforts of radical libertarians to undermine all federal power.

Scholars might prefer to ignore these popular-press books and their authors, but they do so at their peril. Not only do these revisionist works draw more public attention and a larger readership than any work of mainstream scholarship could hope to attain, but, in the name of current political goals, these authors also bash away at the foundations for recognizing three of the most important episodes of historic American injustice: slavery, the Japanese American internment, and McCarthyism.

For scholars who focus on reparations theory, these works could not be more threatening; they tear open holes in the foundation of reparations theory by seeking to persuade millions of Americans that what we commonly take to be episodes of historical injustice were not, in fact, unjust. All of the carefully wrought theories of restitution, recovery, causation, reconciliation, and apology that scholars have developed in recent decades will surely be for naught if Americans come to believe that American history includes no episodes of true injustice.

This Article calls for repair of this critical hole in the foundation of reparations theory. It seeks to initiate discussion about how best to repair the hole by proposing an approach to defining historical injustice derived from principles of American criminal law. The method proceeds by analogy: it compares the problem of judging the wrongdoing of historical actors with the familiar criminal law problem known as the “cultural defense.” The cultural defense is a legal strategy in which a defendant presents evidence of his or her cultural background to attempt to avoid criminal liability. In a typical cultural defense case, a recent immigrant from a foreign culture argues that it is unfair to judge her here for an act that was permissible, or at least tolerated, under her foreign culture’s legal, ethical, or social norms.

The cultural defense might first appear to have little to do with the problem of defining historical injustice. But the problem of defining historical injustice typically arises as a debate about whether it is fair to judge the actions of a historical figure by current legal and ethical

---


10 For a full definition of the term “cultural defense,” see Volpp, supra note 9, at 57.
standards, rather than the standards of her earlier era.\textsuperscript{11} Seen this way, the analogy emerges: just as we might consider excusing a person because of \textit{where} she is from, we might also consider excusing a person because of \textit{when} she is from. It is an analogy between the culture of a place and the culture of an era.

The fact is, however, that American criminal law very rarely excuses a person because of culture.\textsuperscript{12} Indeed, no American jurisdiction has recognized a complete, free-standing cultural defense to criminal liability.\textsuperscript{13} By and large, the cultural defense has failed as an innovation in American law, and immigrants are usually judged by the same standards as non-immigrants.\textsuperscript{14} Only in the rare circumstance where the overwhelming influence of an immigrant’s culture completely vitiates his mens rea does the law relent and excuse the immigrant’s act.\textsuperscript{15}

The criminal law’s near-total repudiation of the cultural defense suggests that we ought to be similarly wary of its temporal cousin. That is, just as we have good reasons to decline to excuse a person simply because of his cultural background, we have similarly good reasons to decline to excuse a person simply because of his historical background. We should not shy away from labeling slave ownership, racial internment, or McCarthyism as historical injustices simply because their wrongfulness is clearer to us today than it may have been to the perpetrators at the time. Instead, as with the cultural defense, we should excuse a historical actor and find no historical injustice

\textsuperscript{11} See generally, e.g., Malkin, supra note 6.


\textsuperscript{13} See Dierdre Evans-Pritchard & Alison Dundes Renteln, \textit{The Interpretation and Distortion of Culture: A Hmong “Marriage by Capture” Case in Fresno, California}, 4 S. CAL. INTERDISC. L.J. 1, 19 (1994); see also Volpp, supra note 9, at 57.

\textsuperscript{14} See Alison Dundes Renteln, \textit{The Cultural Defense} 6 (2004) (“[J]udges often exclude evidence about cultural background on the ground that it is ‘irrelevant.’”).

only where the attitudes and urgencies of an actor’s historical era vitiate his mens rea.

Although it is admittedly odd to think of historical actors having or not having a mens rea, this Article identifies several factors relating to the *chosenness* of the actor’s conduct that can help us distinguish past acts that should be wholly or partly excused from those that should not. The inquiry will focus on the political and social climate, moral norms, and behavioral expectations of the actor’s day. Were they truly monolithic? Did the actor’s society know a rival tradition? Did the actor have access to that rival tradition? In other words, do we look back on the actor’s behavior and see it as a *choice*, an act of independent moral agency, rather than an inevitability? Only to the extent that we see behavior as inevitable should we think about labeling it as something other than a historical injustice.

This Article does not maintain that the analogy between foreign cultures and earlier eras is perfect, or that it single-handedly solves the problem of defining historical injustice. Rather, it is merely an initial effort to fix a hole in the foundation of reparations theory that scholars have ignored until now, which is growing larger and more dangerous with the ascension of each new partisan revisionist tome up the New York Times best-seller list. Other methods for fixing the hole should—and hopefully will—come from other areas of the law and from disciplines outside of law entirely, most notably history and philosophy. This Article will have more than served its task if it draws attention to this growing structural weakness in the foundation of reparations theory and stimulates critical discussion about how best to fix it—whether by analogy to the criminal law’s cultural defense, or by some other method.

Part I of this Article demonstrates that leading works on reparations theory have not adequately defined what constitutes a historical injustice. Part II uses the example of the Japanese American internment during World War II to illustrate the problems that can result from the failure to define a historical injustice. Part III introduces an analogy between the criminal law’s “cultural defense” and a “temporal defense,” suggesting a framework by which reparations scholars may begin to address the challenge of assigning culpability to historical actors. This Part then considers several possible flaws in

---

16 See infra notes 23–52 and accompanying text.
17 See infra notes 53–86 and accompanying text.
18 See infra notes 87–91 and accompanying text.
the analogy and examines how such challenges may be addressed. Part IV examines the scholarly debate regarding the dangers and advantages of allowing cultural evidence to excuse the actions of criminal defendants, and discusses how these debates can provide guidance for assigning culpability to historical actors. Part V discusses how cultural evidence can negate the mens rea element of a criminal act, and draws upon historical examples to recommend a similar “lack of mens rea” defense for historical actors. Finally, Part VI applies this framework to examine the culpability of historical actors responsible for the Japanese American internment to determine whether this episode should be regarded as a historical injustice meriting redress.

I. THE DANGEROUS HOLE IN THE FOUNDATION OF REPARATIONS THEORY

It is difficult to point out something that is absent. But something crucial—indeed, foundational—is missing from the sizable literature on reparations. What is missing is a theory of how to define a potentially reparable historic injustice.

Consider several leading works on reparations. In Politics and the Past: On Repairing Historical Injustices, edited by sociologist John Torpey, an array of scholars in law, sociology, history, and anthropology discuss and debate a variety of problems in the theory and practice of reparations, restitution, and apology. Yet all have a common point of departure: an assumption that what needs discussing is how best to respond to “claims for mending past wrongs that are . . . extremely varied, running the gamut from specific human rights abuses against individuals such as unjust imprisonment and torture to such diverse social systems as plantation slavery, apartheid, and colonialism.” That is, the existence of these “past wrongs” is the shared premise of the entire project. Each scholar simply assumes that instances of torture, unjust imprisonment, slavery, apartheid, and colonialism are “wrongs,” and the more interesting question is how scholars with a focus on transitional justice, reparations, apology, and reconciliation

---

19 See infra notes 92–96 and accompanying text.
20 See infra notes 97–129 and accompanying text.
21 See infra notes 130–97 and accompanying text.
22 See infra notes 198–234 and accompanying text.
24 Id. at 5–6.
respond to these “wrongs.” Nowhere in the volume does any scholar attempt to explain what makes these historical episodes the sorts of “wrongs” that potentially merit redress.

Martha Minow’s excellent volume, *Between Vengeance and Forgiveness: Facing History After Genocide and Mass Violence*, proceeds from the same premise. “This book explores how some nations have searched for a formal response to atrocity, some national or international re-framing of the events,” explains Minow in the book’s introduction. The book’s very subtitle makes her premise clear: its two factual predicates—”genocide” and “mass violence”—leave no space for disagreement over whether a wrong occurred. Injustice is simply (and, in the cases of genocide and mass violence, undoubtedly correctly) assumed; Minow’s question is how to steer a path “between vengeance and forgiveness” in reconciliation efforts in injustice’s wake.

The same premise supports Jeremy Waldron’s influential and challenging essay, *Historic Injustice: Its Remembrance and Supersession*. Unlike many other scholars who write about reparations, Waldron is not particularly supportive of non-symbolic compensatory reparations for historic injustices. But Waldron’s reason for opposing reparations is not that he questions the existence or the definition of underlying historic injustices. Quite the opposite is true. “Historic injustice” is not merely the premise but the *title* of his essay, and he begins

---

25 One slight exception to this statement is Alan Cairns’s contribution to this volume. See Alan Cairns, *Coming to Terms with the Past*, in *Politics and the Past*, supra note 23, at 63–90. Cairns writes broadly of the need to “come to terms with the past,” which entails “seeing the behavior of our predecessors, and sometimes of our earlier selves, in terms of its consequences for contemporary generations.” Id. at 66. He also emphasizes the importance of “the scholarly recovery and interpretation of historical facts, including the identification of who did what to whom”—”the prosaic, yet noble, goal of finding out what happened.” Id. at 70. Cairns does not, however, suggest any sort of method for such an inquiry, nor any way of distinguishing historical wrong from right. See id. at 63–90.

26 See generally *Politics and the Past*, supra note 23.


28 Id. at 2.


30 Waldron does, however, support efforts at redress that he deems symbolic rather than compensatory. The federal government’s token redress payments of $20,000 to surviving Japanese American internees is an example of the sort of symbolic reparations that Waldron supports. See id. at 143.

31 Waldron argues that changes in the circumstances of the descendants of the wrong-doer, the victim, and third parties result in the supersession of the underlying wrong, and that compensatory reparations are therefore unjust. See id. at 139–66.
by writing: “In reviewing our history, we come across events *that can only be described as injustices.*”

This is not an essay that entertains doubt about what constitutes a potentially compensable wrong. “People, or whole peoples, were attacked, defrauded, and expropriated,” Waldron reports; “their lands were stolen and their lives were ruined.”

Waldron’s question is not whether these were in fact instances of injustice; instead he asks, “As we become aware of these injustices, what are we to do about them?”

Mari Matsuda’s seminal treatment of reparations, *Looking to the Bottom: Critical Legal Studies and Reparations* similarly slights the question of how to define a compensable historical injustice. Matsuda contemplates reparations for “racist acts,” but does not define them further. She also writes aspirationally of “physical liberty from constraint of the person”; “freedom from life-threatening abuse as well as the right to seek the nurturance and livelihood human beings need for survival”; rights to “personhood and participation—the recognition of one’s existence as a human being, free and equal, with power and control over the political processes that govern one’s life”; “freedom from public and private racism[,] freedom from inequalities of wealth distribution[,] and freedom from domination by dynasties.”

Nowhere, however, does Matsuda establish that the intentional or unintentional denial of these privileges and freedoms at any particular point in American history was (or will be) an episode that would cause a redressable injury to accrue.

To its credit, the most comprehensive edited volume in the legal literature on reparations recognizes that a historic injustice cannot be assumed, but must be proven. Roy L. Brooks, the volume’s editor, identifies five requirements for “a meritorious redress claim,” the first

---

32 Id. at 139 (emphasis added).
33 Id.
34 Waldron, *supra* note 29, at 139.
36 Id. at 389.
37 Id. at 389–90.
38 Writing in the mid-1980s, as the movement for Japanese American redress was gathering steam, Matsuda may have assumed a growing consensus about the injustice of at least some episodes in American history, such as the Japanese American internment. Sadly, however, Michelle Malkin’s popular book justifying the internment shows that any such optimism was misplaced. See generally MALKIN, *supra* note 6.
two of which are that “a human injustice must have been committed” and that “it must be well-documented.”

Brooks does not leave matters there; he defines his terms. But perplexingly, Brooks anchors his definition of a historical “human injustice” in the present day. Brooks begins with Article 55(c) of the Charter of the United Nations (the “U.N.”), which commits the U.N. to the promotion of “universal respect for, and observance of, human rights and fundamental freedoms for all without distinction as to race, sex, language, or religion.” He adds to that the “various multilateral and bilateral conventions, covenants, resolutions, and treaties” that “more sharply define the rights of all humans,” as well as the prohibitions of customary international law. In sum, he states that a historical human injustice is simply “the violation or suppression of human rights or fundamental freedoms recognized by international law.”

Brooks deserves credit for attending to the problem of defining historical injustice, something other scholars ignore. But surely the definitional problem cannot be as focused in the present as Brooks makes it out to be. Debates about reparations are typically about past, not present, wrongdoing. This is certainly true in the American context. Our great reparations debates are, and will continue to be, about conduct that is generations old: slavery, Jim Crow, the eviction of American Indians from their native lands, and the decimation of the American Indian population and culture. Various devastating effects of these old American policies linger to the present day, but the primary human choices that created them and nourished them in their

---

40 Id. (citing Matsuda, supra note 35, at 362–97).
41 Id.
42 Id. (quoting U.N. Charter art. 55, para. c).
43 Id.
44 See Brooks, Introduction to When Sorry Isn’t Enough, supra note 39, at 3–11. In a slightly more recent work, Professor Brooks modified his approach, contending that reparations “apply only to certain types of wrongs, to wit, gross violations of fundamental international human rights, such as slavery, genocide, and Apartheid.” Roy L. Brooks, Getting Reparations for Slavery Right—A Response to Posner and Vermeule, 80 NOTRE DAME L. REV. 251, 255 (2004) (emphasis added) [hereinafter Brooks, Getting Reparations for Slavery Right]. Claims for reparations, Brooks argued, “characteristically arise only in the context of an atrocity.” Id. The leading American example of reparations, however, is the Japanese American internment—a deep injustice and a civil liberties disaster, to be sure, but not what the word “atrocity” typically connotes.
45 To be sure, some wrongdoing—the genocide in Rwanda, for example—is sufficiently recent to be controlled by current treaties, human rights law, and modern customary international law. See Minow, supra note 27, at 122–31 (discussing modern use and shortcomings of international war crimes tribunals and truth commissions).
46 This list is illustrative, not exhaustive.
infancy are decades, even centuries, old. It is, to say the least, deeply contestable to maintain that the international human rights law of the year 2006 should govern our inquiry into the extent to which, for example, seventeenth-century slave ownership or nineteenth-century American Indian policy was an injustice in its own time.

Naturally, we might choose to create systems of redress for these episodes without regard for whether they were unjust in their own time. We might create such systems for what Jeremy Waldron calls the “symbolic function of embodying . . . remembrance,” or we might do so for a variety of essentially present-focused or forward-looking reasons. We might choose to redress past suffering to make a statement about what our present values are, regardless of what our predecessors’ values were, or to foster political reconciliation. But to the extent that our reasons for creating such systems are essentially backward-looking—that is, to the extent that they are compensatory and predicated on a debt or obligation that accrued in the past and survives to this day—we must confront the question of whether a perpetrator’s conduct was wrong in its day.

For this task, current-day human rights law will often be beside the point. For example, Brooks includes the following on his list of potentially redressable historic injustices:

- genocide;
- slavery;
- extrajudicial killings;
- torture and other cruel or degrading treatment;
- arbitrary detention;
- rape;
- the denial of due process of law;
- forced refugee movements;
- the deprivation of a means of subsistence;
- the denial of universal suffrage;
- and discrimination, distinction, exclusion, or preference based on race, sex, descent, religion, or other identifying factor with the purpose or effect of impairing the recognition, enjoyment, or exercise, on an equal footing, of human rights and fundamental freedoms in the political, social, economic, cultural, or any other field of public life.

It is troubling enough that a few of the listed behaviors, especially those toward the bottom of the list, are not condemned under current

47 See Waldron, supra note 29, at 143.
48 Professor Brooks helpfully distinguishes among various backward- and forward-looking approaches to redress in Getting Reparations for Slavery Right. See Brooks, Getting Reparations for Slavery Right, supra note 44, at 260–86.
49 Brooks, Introduction to When Sorry Isn’t Enough, supra note 39, at 7.
But many more of the items on the list would not have raised concerns of injustice in countless societies through much of human history. And at least one important item on the list—“the denial of due process of law”—has meant vastly different things at different moments in American history. A “denial of due process of law” defies understanding as a historical injustice except by reference to the historical era in which it occurred.

Of course, the mere fact that earlier generations may have endorsed or condoned certain practices should not by itself prevent us from seeing them as historical injustices. It is precisely the goal of this Article to stimulate thinking and debate about which historical practices, although condoned in their day, ought nonetheless to qualify as potentially redressable injustices. But careful analysis and debate are necessary—rather than merely invoking modern legal standards—if reparations theory is to avoid the risks of presentism and hindsight judgment that can skew our assessment of the acts of earlier generations.

II. The Danger Illustrated: The Example of the Japanese American Internment

The absence of a definition of historical injustice is not a mere conceptual flaw in reparations theory. It is a structural weakness that threatens the integrity of the entire project of compensatory reparations. No episode of American historical injustice illustrates the danger as vividly as the incarceration of Japanese Americans during World War II and the successful—but newly questioned—movement to obtain reparations for surviving internees.

The underlying story of historical injustice is well known. In early 1942, after intense lobbying by West Coast racial nativists, economic
competitors of Japanese farmers, hysterical newspaper columnists, and elected officials, the federal government evicted all people of Japanese ancestry from their homes in a wide strip along the Pacific Coast and forced them into detention behind barbed wire.\textsuperscript{55} Nearly 120,000 people were displaced, two-thirds of whom were U.S. citizens.\textsuperscript{56} No legal process of any sort accompanied this mass exclusion or the multi-year incarceration that followed.\textsuperscript{57} The government’s stated theory was that each and every one of these people—citizen and alien, child and adult, the orphaned, the disabled, and the terminally ill—were members of an “enemy race” and, therefore, potential spies and saboteurs.\textsuperscript{58}

It was not until the mid-1970s—three decades after the last of the internees emerged from behind barbed wire—that survivors and their children began to press for a government inquiry into how the oppressive wartime policy came about.\textsuperscript{59} That pressure resulted in the establishment in 1980 of the Commission on Wartime Relocation and Internment of Civilians (the “Commission”), a blue-ribbon and bipartisan panel of political leaders, administrators, judges, community leaders, and scholars that was to investigate the government’s wartime policies and report its findings to Congress.\textsuperscript{60} The Commission held twenty days of public hearings and delved deeply into archival holdings during a comprehensive investigation which lasted almost two years.\textsuperscript{61} In February of 1983, the Commission issued its unanimous report to Congress, concluding that the policy of excluding Japanese Americans from the West Coast “was not justified by military necessity” and that the long-term mass detention that followed was “not

\textsuperscript{55} Excellent accounts of the decision-making process that led to the exclusion and incarceration of Japanese aliens and American citizens of Japanese ancestry include ROGER DANIELS, CONCENTRATION CAMPS USA: JAPANESE AMERICANS AND WORLD WAR II (1972); PETER IRONS, JUSTICE AT WAR: THE STORY OF THE JAPANESE AMERICAN INTERMENT CASES (1983); TETSUDEN KASHIMA, JUDGMENT WITHOUT TRIAL: JAPANESE AMERICAN IMPRISONMENT DURING WORLD WAR II (2003); GREG ROBINSON, BY ORDER OF THE PRESIDENT: FDR AND THE INTERMENT OF JAPANESE AMERICANS (2001).

\textsuperscript{56} See KASHIMA, supra note 55, at 4.

\textsuperscript{57} See id. at 48–66.


\textsuperscript{59} Roger Daniels has summarized the background of the redress movement. DANIELS, supra note 54, at 331–41.


\textsuperscript{61} Id. at xvii.
driven by analysis of military conditions.”[^62] “The broad historical causes which shaped these decisions,” said the Commission, “were race prejudice, war hysteria and a failure of political leadership.”[^63]

The Commission also recommended a number of remedial actions to Congress “as an act of national apology.”[^64] These included legislation “recogniz[ing] that a grave injustice was done and offer[ing] the apologies of the nation for the acts of exclusion, removal and detention”; presidential pardons for Japanese Americans convicted of resisting exclusion and internment; liberal administrative review of Japanese American claims for restitution for lost positions, status, or entitlements; the creation of an educational and humanitarian foundation; and a “one-time per capita compensatory payment of $20,000” to each surviving internee.[^65] Five years later, after extensive lobbying and debate, the Commission’s five recommendations became law when President Reagan signed the Civil Liberties Act of 1988.[^66] In 1990, President Bush issued a formal apology and the government sent out the first $20,000 redress checks.[^67]

The success of the Japanese American redress movement made it a sort of “poster child” of American reparations theory—a “monumental,” even “unique” political achievement[^68] that had, and continues to have, the potential to serve as a model for the redress claims of other victims of historical injustices.[^69] Yet that success has also placed the movement for Japanese American redress in the crosshairs of the partisan historical revisionists. For a time, the revisionist efforts sat at the fringes of public discourse, appearing mostly in vanity-press books[^70] and Internet discussions. In 2004, however, Fox News commentator Michelle Malkin grabbed these fringe critiques and thrust them before the eyes and ears of millions of readers, television view-

[^62]: Id. at xvii; Personal Justice Denied, supra note 60, at 18.
[^63]: Personal Justice Denied, supra note 60, at 18.
[^64]: Id. at 462.
[^65]: See id. at 462–66. Commissioner Daniel Lungren, a U.S. Representative from California, dissented from the last of these recommendations. See id.
[^67]: See Kashima, supra note 55, at 221.

Those who oppose reparations for historical injustices typically have a number of arguments at their disposal: the perpetrators are long dead, or the victims are long dead, or no non-arbitrary class of victims can be designated, or too many decades or centuries have passed to allow a rational calculation of compensation, or the wrongs were insufficiently documented.\(^{\text{72}}\) In the case of the Japanese American internment, these usual arguments fail: some perpetrators and victims were (and are) still alive, redress payments went only to those personally affected by the unjust policy, and the injuries are well documented and remain in the memories of people still living.\(^{\text{73}}\) The revisionist strategy must therefore be more direct: it must challenge the premise that a redressable injury ever occurred. It must maintain that the Japanese American internment was justified.

That is precisely the argument that Malkin’s book makes. Its self-described mission is to “debunk[] the great myth of the ‘Japanese American internment’ as ‘racist’ and ‘unjustified.’”\(^{\text{74}}\) Its “central thesis . . . is that the national security measures taken during World War II were justifiable, given what was known and not known at the time.”\(^{\text{75}}\) These measures were “not based primarily on racism and wartime hysteria,”\(^{\text{76}}\) the book maintains, but rather were military reactions to a handful of top-secret decoded Japanese diplomatic messages that suggested Japanese efforts to recruit Japanese American spies on the West Coast.\(^{\text{77}}\) They were reasoned decisions by “serious men, aware of the gravity of both their actions and inaction”—men who “did not have the luxury of a rearview mirror.”\(^{\text{78}}\)

This last point is the book’s springboard from justifying internment to assaulting redress. According to Malkin, the Commission on Wartime Relocation and Internment of Civilians issued “a false account” of the historical episode.\(^{\text{79}}\) It did so because of intimidation by

---

\(^{\text{71}}\) See generally Malkin, * supra* note 6.

\(^{\text{72}}\) See Brophy, * supra* note 2, at 501–25.

\(^{\text{73}}\) See Daniels, * supra* note 54, at 340 (explaining that the Japanese American redress program provided a one-time per capita compensatory payment to each of the survivors of the relocation).

\(^{\text{74}}\) Malkin, * supra* note 6, at xii.

\(^{\text{75}}\) Id. at xxxiii.

\(^{\text{76}}\) Id. at 80.

\(^{\text{77}}\) Id. at 37–51, 78–80.

\(^{\text{78}}\) Id. at xxxiv.

\(^{\text{79}}\) Malkin, * supra* note 6, at 115.
the “America-bashing” Japanese American redress movement, a group of “civil rights absolutists and ethnic lobbyists . . . with an intellectual arrogance that only 20/20 hindsight can instill.” According to Malkin, the redress movement fell victim to “the danger of judging the wartime measures after the fact and out of historical context,” and ignored the “risks of sabotage and espionage” that “reasonable men” of that day “vividly perceived.” For Malkin, the nub of the problem is hindsight: the exclusion and internment of 120,000 people were the justified judgments of their era’s reasonable minds. Such judgments reflected wisdom, not injustice—and wisdom needs no apology.

It is not the concern of this Article to establish the incompleteness, the illogic, the deceptiveness, and the outright falsehood of Malkin’s historical account. What is important is the danger that Malkin’s method poses for reparations theory and practice. Reparations theory sits atop an assumption that American history includes episodes of injustice—policies and behaviors that were wrong in their time, and whose traumatizing, stigmatizing, and impoverishing effects linger to the present day. Actual reparations movements depend upon something approaching political consensus that such episodes occurred. Even for such horrific chapters as slave ownership, the suppression of indigenous peoples and cultures, and mass racial internment, the consensus that these were unambiguous instances of injustice is, in the context of American history, recent. Partisan historical revisionism of the sort preached in Malkin’s In Defense of Internment tears a hole in that consensus, and accordingly, the structure built upon the consensus weakens.

It is time for scholars to begin to fill that hole.

80 Id. at 116.
81 Id. at xxxiv.
82 Id. at 118.
83 Id. (quoting James J. Kilpatrick, Op-Ed., $1.2-billion Worth of Hindsight, St. Petersburg Times, Mar. 8, 1988, at 19A.)
84 See Malkin, supra note 6, at xxxii–xxxv.
85 See id.
86 That work has been amply done elsewhere. See Postings of Eric L. Muller and Greg Robinson to The Volokh Conspiracy and IsThatLegal? blogs, http://www.isthatlegal.org/Muller_and_Robinson_on_Malkin.html (last visited Aug. 24, 2006).
III. TRYING OUT AN ANALOGY: THE CULTURAL DEFENSE AND THE "TEMPORAL DEFENSE"

In which discipline should the hole-filling project occur? Thus far, the problem of defining historical injustice seems to have fallen through the gaps between disciplines. Legal scholars have been content either to assume consensus about historical injustice or to use presentist definitions that ignore the considerable problem of hindsight. Historians are typically more concerned with explaining why people acted as they did than with condemning particular policies and choices. And philosophers have paid scant attention to the problem.

Because the problem of defining historical injustice sits at the intersection of law, history, and philosophy, the project will best proceed as a cross-disciplinary effort. This Article, however, seeks to begin the dialogue by positing an analogy between the problem of defining historical injustice and the well-known criminal law problem of judging the wrongdoing of immigrants from foreign cultures. The analogy is admittedly imperfect, and does not single-handedly resolve the difficulties of defining historical injustice. It is, however, a provocative and helpful start.

Consider the following story: John and his lover Sara are from a different culture from our own. John learns that Sara has had sex with another man. In John and Sara’s culture, a woman’s sexual infidelity is a grave insult to a man’s honor. Their culture also tolerates corporal punishment for such betrayals. John therefore gives Sara ten lashes across the back with a whip. The trouble for John, however, is that he and Sara no longer live in their culture of origin; they live in the United States. Under twenty-first-century American law, John’s actions constitute the felony of aggravated battery.

John’s case starkly presents the problem of the “cultural defense” in the criminal law. Everyone called upon to assess John’s culpability—from the police officer who arrests and charges him, to the prosecutor who negotiates with his lawyer for a possible guilty plea, to the jurors who determine his guilt or innocence, to the judge who ultimately sentences him—will have to decide whether and to what

87 See supra notes 23–52 and accompanying text.
88 See Trevor Burnard, Mastery, Tyranny, & Desire: Thomas Thistlewood and His Slaves in the Anglo-Jamaican World 31 (2004) ("As historians, it is not our responsibility to attribute retrospective blame.").
extent the norms of John’s society of origin should extinguish or dimin­ish John’s culpability for whipping Sara. Do we treat John fairly if we assess his conduct by reference to our cultural standards rather than his? If, in fairness to John, we use his cultural standards rather than ours, do we compromise our efforts to deter crime and thereby place potential victims—especially powerless and vulnerable victims—at greater risk?

Now let us change one aspect of John’s case: John’s native society is not a foreign country, but rather eighteenth-century America. He is a plantation owner, and Sara is not just his lover, but also his slave. Their culture is the culture of that day. What has brought John and Sara to us is not a ship or an airplane, but rather the passage of time. We are historians, looking back on John and Sara from our early-twenty-first-century vantage point and trying to determine John’s culpabil­ity for whipping Sara for her sexual infidelity 200 years ago.

Intriguing questions very similar to those raised by the criminal law’s cultural defense now arise. As we assess the conduct and prac­tices of an earlier generation, to what extent should we be bound by that earlier society’s norms? To be fair to John, must we credit what we might call his “temporal defense”—that is, must we assess his behavior by reference to the standards of his day? Can we avoid judging John’s acts according to our own standards without undermining our efforts to deter the type of misdeed that John committed? In sparing John from judgment, do we make current and future generations of potential victims more vulnerable to a rein­ infliction of the same type of harm that John caused long ago?

Note how powerfully these questions resonate with the defensive strategies of the partisan historical revisionists. Michelle Malkin’s case for Franklin Roosevelt and his top military advisers is an instance of what this Article terms a “temporal defense.” She maintains that we commit an injustice against our World War II leaders when we assess their policy of Japanese American exclusion and detention from our modern vantage point.90 We must assess it from theirs, and when we do so, she maintains, it emerges as a reasonable military judgment rather than the product of racism and hysteria.91

The analogy between cultural and temporal defense is striking, and this Article maintains that it can teach us a great deal about how to think about historical injustice. It bears emphasis, however, that the

90 See Malkin, supra note 6, at xxxiii–xxxv.
91 See id.
analogy of culture to time is far from perfect. Three difficulties quickly present themselves, and they are all apparent in the story about John and Sara. First, Immigrant John has presumably chosen to join our culture and might therefore be more fairly saddled with judgment by our culture’s norms; Historical John made no such choice. Second, Immigrant John, by virtue of his actual physical presence among us, has access to the behavioral and moral norms of our society, can learn and follow them if he chooses, and is fairly condemned if he does not do so; Historical John is stuck in his moment in the past and cannot know or follow the norms of today’s society. Finally, a message of forgiveness to Immigrant John (and his cultural community in the United States) will pose clear and direct risks to people living today—most immediately Sara, who will be left at John’s mercy if his conduct is excused. But to speak of “sending a message” to Historical John and members of his generation is incoherent; they are dead and gone. Any risk that excusing their misconduct poses to current or future generations is therefore remote.

These objections based on choice, access, and risk are considerable, but they should not doom this Article’s goal of launching a discussion about defining historical injustice. On careful examination, each of the flaws in the analogy between culture and time is a bit less problematic than it initially appears. Consider first the problem of choice. Although it is true that many immigrants freely choose to come to the United States, immigration is not an unfettered choice for all immigrants. Some immigrants come here because danger effectively forces them to quit their native lands. Many others, particularly dependent spouses, children, and elderly parents come here because a more powerful family decisionmaker has decided to do so. On the other hand, although people from earlier times do not really choose to “appear” at this (or any) particular moment in history, they also really have no choice but to do so. Time marches on. People understand that later generations will scrutinize the choices they make and the actions they take. Thus, especially if Immigrant John has fled to the United States to save his life, or if he has been brought here by someone else, his position with respect to the norms of today’s dominant culture is more like Historical John’s than it may first appear.

92 The Hmong, who are the source of many cultural defense cases, are a good example of this: by virtue of their support for the United States and South Vietnam during the conflict with North Vietnam, it was dangerous for them to stay in Vietnam after the South fell. See Tim Pfaff, Hmong in America: Journey from a Secret War 49–63 (1995).
Consider next the objection based on cultural access. For many immigrants from cultures with practices that diverge significantly from American behavioral norms, the American majority’s culture is often not especially accessible or intelligible. Many first-generation immigrants (and in some communities, even later-generation immigrants) live culturally insular lives. Economic, religious, linguistic, and social barriers and preferences can keep them well apart from the mainstream culture. By contrast, historical figures from the previous one or two generations whose actions are later challenged typically stand well within the cultural mainstream because the mainstream changes slowly over time and usually conforms with at least faintly visible trends. As between, say, a Vietnamese immigrant who has lived entirely within San Francisco’s Chinatown since arriving three months ago and an American political figure from the World War II era, it is not at all clear that the Chinese immigrant is able to develop a clearer sense of current American cultural norms than the World War II leader might have been able to extrapolate.

Consider finally the objection based on risk. On careful examination, the risk of excusing conduct that is approved or condoned in a different culture is similar to the risk of excusing conduct that was approved or condoned at an earlier time. In both cases, the excuse endangers those who might again be victimized by the sort of behavior that first caused injury. In the case of the cultural defense, there are two potential groups of perpetrators and victims, one narrow and one broad. The narrow group is the accused person himself and the person or people he allegedly victimized. The broader group includes other potential perpetrators, who might be encouraged to act if they learn that American law condones a violent or harmful cultural practice, as well as the larger class of potential victims of that cultural practice. Similarly, the temporal defense also involves both a narrow and a broad class of perpetrators and victims. The narrow class of perpetrators and victims has of course passed away. But the broader class remains—and this is the very heart of George Santayana’s famous warn-

ing that those who cannot remember the past are condemned to repeat it. 94 To say that some past victimization was understandable, or even perhaps justified under the circumstances, is to invite that sort of victimization to recur if and when those types of circumstances reappear. Why else were Arab American groups among the first and loudest to complain 95 when U.S. Representative from North Carolina Howard Coble said, after September 11, 2001, that he thought the incarceration of Japanese Americans during World War II was justified? 96 Undoubtedly, the risks the cultural defense poses to potential victims are more immediate and palpable than the risks of the temporal defense. The difference, however, is one of degree, not kind.

To be sure, the analogy between the cultural defense and its temporal cousin is imperfect; judgment in a court of law is not the same as judgment in the court of history. But the analogy does not need to be perfect. It is not the goal of this Article to maintain that the standards of the criminal law should strictly govern our reflection on history. Rather, this Article’s purpose is to initiate conversation about the undertheorized problem of defining historical injustice. Its claim is simply that the criminal law’s cultural defense offers tools for working through a set of problems similar to those we encounter when we attempt to assess the wrongdoing of those who went before us. Let us therefore examine the cultural defense and the progress the criminal law has made toward solving those problems.

IV. THE RISKS AND BENEFITS OF CULTURAL AND TEMPORAL EXCUSES

The cultural defense in the criminal law is the subject of a fairly rich literature, most of it less than twenty years old. 97 That literature

94 See George Santayana, Reason in Common Sense (1905), reprinted in The Life of Reason: Or the Phases of Human Progress 284 (2d ed. 1936).
is, perhaps not surprisingly, marked by disagreement. Something approaching agreement has emerged, however, on a couple of points. First, most scholars now agree that a new, full-blown defense to criminal liability grounded in cultural difference is a bad idea.\textsuperscript{98} And second, most scholars now agree that evidence of cultural practices ought to be admitted, on a case-by-case basis, to support certain existing criminal law defenses.\textsuperscript{99}

The boundaries of the debate over the criminal law’s cultural defense were marked in two law review pieces published in 1986. One, a student note published in the \textit{Harvard Law Review}, presented an unabashed plea for the recognition of a formal cultural defense that would completely exonerate people who, through their cultural practices, commit criminal acts.\textsuperscript{100} The note’s author contended that a defense based in cultural difference is a necessary and sensible way of honoring the nation’s commitments to cultural pluralism and to individualized justice.\textsuperscript{101} “A new immigrant,” the author explained, “has not been given the same opportunity [as a long-term resident] to ab-

\textsuperscript{98} See generally Coleman, \textit{supra} note 97; Evans-Pritchard & Renteln, \textit{supra} note 13; Kim, \textit{supra} note 9; Maguigan, \textit{supra} note 15; Renteln, \textit{supra} note 97; Sacks, \textit{supra} note 97; Volpp, \textit{supra} note 9; Goldstein, \textit{supra} note 97; Sikora, \textit{supra} note 15; Wu, \textit{supra} note 97.

\textsuperscript{99} See generally Coleman, \textit{supra} note 97; Evans-Pritchard & Renteln, \textit{supra} note 13; Kim, \textit{supra} note 9; Maguigan, \textit{supra} note 15; Renteln, \textit{supra} note 97; Sacks, \textit{supra} note 97; Volpp, \textit{supra} note 9; Goldstein, \textit{supra} note 97; Sikora, \textit{supra} note 15; Wu, \textit{supra} note 97.

\textsuperscript{100} See \textit{Note, The Cultural Defense in the Criminal Law, supra} note 97, at 1311.

\textsuperscript{101} See id. at 1298–307.
sorb—through exposure to important socializing institutions—the norms underlying this nation’s criminal laws.” The author conceded that a cultural defense might make it harder for the law to articulate a coherent social order and to deter crime, but countered that the effects might well be the opposite if immigrant groups came to recognize that the American legal system is sensitive to their cultural traditions.

The other 1986 piece, a student comment by Julia P. Sams in the Georgia Journal of International and Comparative Law, presented a full-bore assault on the notion of a cultural defense. Sams argued that a complete cultural defense to criminal liability would be difficult to administer because the legal system would have to decide which immigrant groups qualify for it and which do not, as well as who among the qualifying groups deserve it and who do not. She further contended that a cultural defense would undermine the deterrent effect of the law by removing an important incentive for immigrant groups to learn American laws, and would undermine the principle of legality by placing “newcomers’ . . . opinions and ideas about the laws . . . above the laws as declared by the officials.” She also maintained that a cultural defense would be unfair to members of the cultural majority, to whom the defense would not be available. Although she recognized that “immigrant groups . . . embellish society with their diverse customs and beliefs,” she concluded that those advantages would be “nullified if United States residents are threatened by the crimes of newcomers who are excused from punishment.”

Since 1986, much has been written about the cultural defense. Although the terrain remains contested, a middle ground has opened up between the two poles fixed by those early pieces. That middle ground does not include a new, free-standing complete defense for

---

102 Id. at 1299.
103 See id. at 1304–07.
104 See generally Sams, supra note 97.
105 See id. at 345–48.
106 See id. at 348–50.
107 Id. at 352.
108 See id. at 350–51.
109 See Sams, supra note 97, at 353.
110 See generally Coleman, supra note 97; Evans-Pritchard & Renteln, supra note 13; Kim, supra note 9; Maguigan, supra note 15; Neff, supra note 97; Renteln, supra note 97; Sacks, supra note 97; Volpp, supra note 9; Wanderer & Connors, supra note 12; Fischer, supra note 97; Goldstein, supra note 97; Sikora, supra note 15; Sing, supra note 97; Tomao, supra note 97; Wu, supra note 97.
Most scholars agree that an independent complete cultural defense is both unnecessary and unwise. It is unnecessary because existing law already takes important cultural determinants of a defendant’s mental state into account. And it is unwise for a number of reasons: the defense would, for example, undermine the maxim that ignorance of the law is no excuse, and would plunge the legal system into a messy and potentially offensive set of decisions about what counts as culture and who counts as a true cultural practitioner.

But by far, the most important reason scholars have advanced for rejecting a free-standing cultural defense is that the defense would be perilous for crime victims, especially women and children. A large percentage of cases involving the cultural defense are domestic abuse cases—prosecutions of adult men who batter or kill women with whom they are in relationships, and prosecutions of adult men and women who abuse or kill their children. To exonerate abusers on the basis of a claim that their own culture does not condemn their violence would make that violence harder to deter.

This poignant insight was apparent in the quoted reaction of one battered Chinese woman to People v. Chen, a case in which a male Chinese immigrant successfully invoked alleged Chinese cultural practices to win a sentence of probation for killing his adulterous wife.

---

111 See generally Coleman, supra note 97; Evans-Pritchard & Renteln, supra note 13; Kim, supra note 9; Maguigan, supra note 15; Renteln, supra note 97; Volpp, supra note 9; Sikora, supra note 15.

112 The leading exception is Alison Dundes Renteln, who argues for what appears to be a free-standing cultural defense in her 2004 book, The Cultural Defense. Renteln, supra note 14, at 200–01. I say that she argues for “what appears to be” a free-standing cultural defense because in actuality she believes that cultural evidence should be, and sometimes is, admitted in support of existing criminal law defenses. Indeed, her main reason for insisting on a free-standing cultural defense is what she views as the inappropriate reluctance of judges to accept cultural evidence in support of existing defenses. See infra notes 147–49 and accompanying text.

113 See Wanderer & Connors, supra note 12, at 873; Sing, supra note 97, at 1863–70; Tomao, supra note 97, at 243–56.

114 See Kim, supra note 9, at 110–15; Maguigan, supra note 15, at 44–45; Wanderer & Connors, supra note 12, at 873; Tomao, supra note 97, at 243–56.


117 See Coleman, supra note 97, at 1136–44; Rimonte, supra note 115, at 1326; Sacks, supra note 97, at 541–42; Fischer, supra note 97, at 690–91; Sikora, supra note 15, at 1709–11.

118 See Kim, supra note 9, at 119–21 (referencing People v. Chen, No. 87-7774 (N.Y. Sup. Ct. Mar. 21, 1989)).
“Even thinking about that case makes me afraid,” the battered woman explained. “My husband has told me: ‘If this is the kind of sentence you get for killing your wife, I could do anything to you. I have the money for a good attorney.’”119 This abused woman and many vulnerable women and children like her would be the ones to bear the brunt of a complete defense to criminal liability grounded in culture. Not surprisingly, most scholars agree that further disempowering those who are already vulnerable to violent attack is too high a price to expect them to pay for a legal doctrine that would exonerate their abusers.120

But just as the weight of scholarly opinion has rejected a complete cultural defense, it has also rejected the notion that cultural influences on a defendant’s behavior ought to be irrelevant to guilt or innocence. Scholars have recognized that in some situations, evidence of cultural practices and beliefs might be relevant not to a new, complete cultural defense but to old and well-established ones, especially mistake of fact, lack of intent, and duress.121 Consider, as an example, the case of a Vietnamese immigrant parent charged with child abuse for engaging in cao gio, or “coining”—the practice of massaging an ill child’s back with a medicated oil, and then rubbing the back with the edge of a serrated coin.122 If the child abuse statute requires proof of a specific intent to harm the child, then evidence that the parent engaged in cao gio in order to heal the child would be relevant—indeed, probably crucial—to the issue of intent. Similarly, in a rape prosecution, evidence of cultural practices surrounding sexual intercourse might tend to support a defendant’s claim that he made a mistake about the fact of the victim’s consent.123 And a Yoruban immigrant from Nigeria who makes tribal markings with a razor blade on her

120 See Coleman, supra note 97, at 1136–44; Rimonte, supra note 115, at 1326; Sacks, supra note 97, at 541–42; Fischer, supra note 97, at 690–91; Sikora, supra note 15, at 1709–11.
121 Kay L. Levine’s excellent article Negotiating the Boundaries of Crime and Culture is especially useful on this point. See Kay L. Levine, Negotiating the Boundaries of Crime and Culture: A Sociolegal Perspective on Cultural Defense Strategies, 28 Law & Soc. Inquiry 39, 49–66 (2003). Holly Maguigan also has written an excellent article. See generally Maguigan, supra note 15.
123 This fact pattern alludes to the California case of People v. Moua. See generally Evans-Pritchard & Renteln, supra note 13 (discussing and analyzing at length People v. Moua, No. 315972-0 (Cal. Super. Ct. Feb. 17, 1985)).
son’s face might be able to offer a defense of necessity or duress if she can offer evidence that her culture predicted far greater harm to the child if he was not so marked.\textsuperscript{124}

To be sure, authors have emphasized that it is important for prosecutors to make evidence available to the fact finder that counters or challenges defendants’ cultural claims, in order to reduce the risk that fact finders will be tricked into mistaking willfully misogynist or child-abusive conduct for innocent cultural practice.\textsuperscript{125} But the weight of scholarly opinion has rejected the claim that all evidence of cultural practice should be completely barred.\textsuperscript{126} Instead, it has settled around the idea that cultural evidence ought to be admissible to support established legal defenses based on the defendant’s mental state.\textsuperscript{127}

These insights from the literature on the cultural defense offer us guidance on the problem of defining historical injustices. On the one hand, we benefit from the literature’s firm recognition that cultural context sometimes ought to temper judgment. That recognition reassures us that our concern for the unfairness of judging with the benefit of hindsight stands on something more substantial than a taboo against speaking ill of the dead or a worry about how we will ourselves someday be judged.\textsuperscript{128} The criminal law’s recognition of the relevance of cultural context also highlights the importance of an actor’s mental state at the time she acted. What that mental state might be for historical actors (as opposed to living criminal defendants) remains to be seen. For now, it is enough to note that the cultural defense offers support for the excusing instinct that is at the heart of its temporal cousin.

That support, however, is drastically limited by the near-universal insight in the criminal law literature that a freely available cultural defense imperils the powerless by undermining the deterrent effect of

\textsuperscript{124} See Renteln, supra note 122, at 29–30.

\textsuperscript{125} See Maguigan, supra note 15, at 90–92.

\textsuperscript{126} Perhaps the most hostile to any use of cultural evidence is Doriane Lambert Coleman. See Coleman, supra note 97, at 1103–04 nn.47–48 (discussing inherent flaws in the argument that custom could impair a defendant’s ability to think rationally, and thus negate the mens rea requirement, but acknowledging that her discussion and analysis apply only to situations where cultural evidence is used to exonerate an “otherwise criminal defendant”).

\textsuperscript{127} See Maguigan, supra note 15, at 87–88; Sacks, supra note 97, at 547–50; Wu, supra note 97, at 1020–22.

\textsuperscript{128} See Thomas Babington Macaulay, Sir James Mackintosh (1835), reprinted in CRITICAL AND HISTORICAL ESSAYS: THOMAS BABINGTON, LORD MACAULAY 163 (Hugh Trevor-Roper ed., 1965) (“As we would have our descendants judge us, so ought we to judge our fathers.”).
A similar danger lurks in the temporal context. In every generation there are ruling and ruled classes, and the distribution of these groups across race, gender, and time in the American experience has not been random. The privileged and powerful of today are, on balance, a good deal likelier to be the progeny of yesterday’s elites than of yesterday’s oppressed. This is, in fact, precisely what makes the task of judging the actions of earlier American generations’ leaders so difficult. Even though we are not related to them by blood, they are nonetheless in some sense our ancestors by virtue of the leadership positions they occupied, and our judgment is compromised by the allegiance we feel we owe them. But when we too quickly excuse them for their impositions on the powerless of their day, we deprive the powerless of today, and their advocates, of the important example of their experience. We undermine the deterrent effect of history, and in so doing, create conditions that are more conducive to renewed victimization. The criminal law’s cultural defense thus reminds us that suspending judgment of the wrongdoing of prior generations is not harmless generosity to the departed. It carries a cost, and that cost likely will be born by the powerless.

V. The Mens Rea of Historical Actors

The preceding section might lead a reader to think that the literature on the criminal law’s cultural defense has reached consensus on the relevance and admissibility of cultural evidence. This is not so. Although the more recent literature does not swing wildly from arguments for wholesale exculpation to arguments for categorical inadmissibility as it did twenty years ago, scholars now pursue a narrower set of disagreements about the precise role that cultural evidence might permissibly play at trial and in sentencing. Those disagreements, however, all work from the shared premise that the central legitimate reason to admit evidence of culture is to shed light on the mental

---

129 See Coleman, supra note 97, at 1136–44; Kim, supra note 9, at 111–12; Rimonte, supra note 115, at 1326; Sacks, supra note 97, at 541–42; Fischer, supra note 97, at 690–91; Sikora, supra note 15, at 1709–11.

130 Compare Kim, supra note 9, at 133–38 (proposing a framework for judges to determine whether to admit cultural evidence at criminal trials), with Neff, supra note 97, at 468–75 (advocating use of cultural evidence as a mitigating factor during sentencing), and Sikora, supra note 15, at 1714–24 (arguing judges only should take cultural evidence into account in the sentencing phase of a criminal trial).
state of the accused. This focus on the mental state of the actor from a different culture will help us to develop a better understanding of what we mean when we speak of the culpability of an actor from an earlier day.

In a recent article, Kay Levine presents a compelling explanation for the focus on mental state in the debate over the cultural defense. She explains that the cultural defense raises a fundamental question about the relationship between culture and action. At one extreme, “soft” or “external” theorists of that relationship contend that people choose their actions freely, without cultural influence, and use culture merely to explain or justify what they have done. At the other end of the continuum, “hard” or “internal” theorists maintain that culture “program[s]” actors to behave as they do, “eliminating their sense of agency or responsibility for their actions.” An intermediate position suggests a sort of feedback loop between an actor and culture: an actor initially chooses his actions, but upon recognizing that certain actions fit within a larger cultural schema, he internalizes or appropriates that schema. The schema assumes an ever-more-powerful role in determining the actor’s conduct, and earlier individual motivations recede. It becomes intuitive, natural, and ultimately coercive.

Levine argues that the cultural defense brings a claim of cultural coercion into conflict with the baseline assumption of American criminal law that people freely choose their actions and are individually responsible for everything they do. At the practical level, this conflict between culture and autonomy plays itself out in disputes about a criminal defendant’s mental state. Inquiry into the mens rea—the actor’s allegedly culpable state of mind—is the logical place for a finder of fact to determine the extent to which larger forces out-

---

131 See Maguigan, supra note 15, at 87–88; Sacks, supra note 97, at 547–50; Wu, supra note 97, at 1020–22.
132 See generally Levine, supra note 121.
133 See id. at 42–43.
134 Id. at 43.
135 Id. at 44.
136 Id. at 45–46.
138 See Levine, supra note 121, at 46.
139 Id. at 47.
side the defendant’s own will and awareness shaped or dictated the harmful act. Levine concludes that evidence of culture ought to be admissible where it undermines the defendant’s criminal intent by “provid[ing] a . . . noncriminal explanation for the defendant’s actions or where cultural demands place the defendant under extraordinary stress and rob him of meaningful agency.” Levine contends that no free-standing cultural defense to criminal liability is necessary for this purpose because our well-established criminal law mens rea defenses are adequate to the task. On the other hand, Levine argues that cultural evidence should not be admissible where it does nothing to disprove the defendant’s intent to harm the victim, but instead merely reveals that the defendant’s culture tends to tolerate such harm more readily than does American culture.

Whereas Levine opposes a free-standing cultural defense to criminal liability, Alison Dundes Renteln supports it. In Renteln’s view, an independent, free-standing culture-based defense is necessary to overcome the extreme reluctance of most judges to admit evidence of cultural context in support of an existing criminal law defense. In her recent book, The Cultural Defense, Renteln notes that a person’s cultural context often supplies that person’s motive for engaging in an act that harms another, even if it does not negate the person’s intent to do harm. That culturally-based motive, she contends, is at least partially exculpating, even if it does not negate criminal intent and thereby supply a fully exculpating excuse. What Renteln has in mind is thus a partial excuse, a sort of generic lesser-included-offense for all crimes that would be available to a defendant who acted from a culturally-influenced motive.

What is notable about the disagreement between Levine, an opponent of a free-standing cultural defense, and Renteln, a supporter, is a hidden point of agreement. Both believe that to the extent that culture mitigates or cancels criminal liability, it does so because culture can undermine the culpability of a mental state. Renteln writes clearly that “the rationale behind [a free-standing cultural defense]

---

140 See id.
141 See id. at 80.
142 See id. at 77 n.60. This is also a major thrust of Holly Maguigan’s work. See Maguigan, supra note 15, at 87–90.
143 See Levine, supra note 121, at 75.
144 See Renteln, supra note 14, at 200–01.
145 See id. at 189–201.
146 See id. at 191–92.
... is that an individual’s behavior is influenced to such a large extent by his culture that either (1) the individual simply did not believe that his actions contravened any laws, or (2) the individual felt compelled to act the way he did.”

Renteln wishes to be certain that the criminal law appropriately mitigates the liability of persons whose “cultural conditioning predispose[d] [them] to act in certain ways,” and who therefore acted with a “beneficent motive.”

Levine makes more or less the same point in her work, but with different words: the defendants entitled to mitigation are those whose “cultural demands place [them] under extraordinary stress and rob [them] of meaningful agency.” As with Renteln, Levine’s focus is on cultural compulsion. She notes that a successful defense based on culture in a criminal case requires proof both that the defendant honestly acted for the claimed cultural reason at the time of the harmful act and that the defendant’s “reliance on culture (rather than on American legal standards) was reasonable.”

The inquiry into reasonableness focuses on “the actor’s (in)ability or (un)willingness to disregard cultural prescriptions—to resist cultural schemas—in response to the victim’s behavior or mainstream social pressures.” Cultural schemas are not, after all, invariably and irreversibly coercive: “A cultural frame that has been taken into the self can be taken out again—when others fail to react in expected ways, for example, or when circumstances change, or simply when a person matures.”

Thus, Levine argues, the reasonableness of a defendant’s reliance on culture will turn on the degree of its chosenness. If the fact finder “believes that the defendant was able to operate outside of the schema but simply chose not to do so, it will likely find her reliance on culture unreasonable.”

If scholars agree that harmful acts are excusable to the extent that a defendant’s cultural framework compelled her to act as she did, then perhaps something similar might be true for the cultural defense’s temporal cousin. The analogy is admittedly not easy to draw because historians, unlike lawyers, do not have doctrines (or even

---

147 Id. at 187.
148 Id. at 13.
149 Renteln, supra note 14, at 201.
150 Levine, supra note 121, at 80.
151 See id. at 48.
152 Id.
153 See id. (quoting Ortner, supra note 137, at 89).
154 Levine, supra note 121, at 48.
really much of a vocabulary) of culpability. Historians are simply not accustomed to examining the mens rea of historical actors. But the literature of the cultural defense suggests fairly clearly that *compulsion and choice* are important factors in assessing the culpability of the actions of past generations. To what extent, we should ask, did the political and social climate, the moral norms, and the behavioral expectations of the actor’s day compel her to do what she did? To what extent was it possible for the person to act differently and more consistently with what would emerge as the standards of a later age? To what extent can we look back on the actor’s behavior and see it as a *choice* of consequence, an act of independent moral agency?

A number of factors will inform this inquiry into the historical figure’s agency. At a basic level, we must ask whether the circumstances of the actor’s time allowed any access at all to standards different from his own. Consider, for example, a hypothetical Virginia planter in the year 1690 who made a decision to switch his crop from wheat to tobacco. A modern anti-smoking activist, who knows that tobacco is an addictive agent of illness and death for millions of smokers, might condemn this decision. But that activist surely cannot condemn the planter’s decision in its own historical moment. Tobacco’s lethal properties were not known in the late seventeenth century—“the relationship of tobacco to health lay deep in the shadow of ignorance.” Attacks on smoking before the mid-nineteenth century “were couched largely in moral, xenophobic, and economic terms,” not in terms of health. Indeed, seventeenth-century Englishmen saw smoking as a defense against the plague. Even as late as the 1850s, the leading English-language medical journal opined that smoking

---

156 Historian Gordon A. Craig beautifully expressed this point about the choices confronting historical figures in his essay entitled *History as a Humanistic Discipline*.

To forget that the present is the result of many developments that might have taken a different course and of decisions that might not have been made, or not at the same time or in the same way, is seriously to foreshorten our historical perspective and to indulge in linear thinking of the most restricted kind. The duty of the historian . . . is to restore to the past the options it once had.

158 *Id.* at 14–15.
159 *Id.* at 15.
could “be indulged in with moderation, without manifest injurious effect on the health for the time being.” Just as the law might be willing to entertain a claim in mitigation on behalf of a rainforest tribesman suddenly plopped down in the middle of early-twenty-first-century America, on the basis that the tribesman lacked any meaningful access to a tradition other than his own, so should we be open to such a claim on behalf of a temporally-remote ancestor.

Now contrast this Virginia planter’s farming decision with the decision of his hypothetical great-great-grandson to purchase slaves in 1840. A plea to suspend judgment of the nineteenth-century slave-holding planter in deference to his historical moment is surely weaker than the plea on behalf of his tobacco-harvesting forebear. Whereas the late-seventeenth-century planter lived in a time of ignorant consensus on questions of tobacco and health, the nineteenth-century slave owner lived in a time of intense debate on the propriety of owning other human beings. Change was afoot, and had been for some time. The American and French revolutions of the late eighteenth century upended the earlier understanding of the world as a divinely ordained hierarchy that assigned everyone—masters and slaves alike—fixed positions of social and political control. Writing in 1782, no less a figure than Thomas Jefferson had noted that “the whole commerce between master and slave is . . . the most unmitting despotism on the one part, and degrading submissions on the other.” By 1807, the slave trade had been abolished in England, and a slave revolt had won independence for what would become the Republic of Haiti. In the decades that followed, an abolition movement took root and flourished in the United States even while Southern jurisdictions passed laws designed to protect the institution of slavery and to make voluntary emancipation more difficult. Slavery

160 Id. at 16.
161 See BURNARD, supra note 88, at 104–06.
162 See ISAAC, supra note 1, at 46–47, 105, 180–83.
165 See George F. Tyson, Jr., Toussaint L’Ouverture 25 (1973).
itself (and not merely the slave trade) was abolished throughout the British Empire in 1834. Slave acquisition and ownership in the 1840s were inevitably acts of conscious moral choice in a way that they had not been in an earlier time. Thus, just as a fact finder would be interested in knowing whether a recent Hmong immigrant came to the United States from a place of genuinely monolithic culture (his remote home village in the Laotian mountains) or from a place of broad and mixed influences (years of interim residence in an Asian metropolis), we should be interested in knowing whether our ancestor lived in a time of stasis on the practice in question or in a time of flux.

Also relevant to what I am calling the chosenness of past wrongdoing is the position of the questioned practice in the cultural spectrum of its day. We know, for example, that eighteenth-century Jamaican slavery was a brutal institution—more so than the contemporaneous slavery of the southeastern United States. Yet even within that violent framework, we can identify sadistic sociopaths. Jamaican slave owner Thomas Thistlewood did not confine himself to the flogging and sexual harassment that were then common methods of slave control. He devised a punishment he called “Derby’s Dose,” in which he whipped a misbehaving slave, rubbed “salt pickle, lime juice and bird pepper” in the open wounds, forced another slave to defecate in his mouth, and then gagged him for four or five hours. Thistlewood also forced slaves to urinate into the eyes and mouths of others, and sometimes rubbed slaves with molasses and left them naked outside overnight to be devoured by mosquitoes. Just as the law would insist on knowing whether a particular immigrant’s violent behavior is truly an artifact of his culture or a deviant choice of his own, we should ask the same question about the harms that our ancestors inflicted. “Derby’s Dose” was an outrage even in its time, and we should not allow arguments about hindsight and historical context to dull our appreciation of that.

Just as the position of the wrongdoing in the spectrum of an era’s conduct is an important facet of the chosenness of past wrongdoing, so is the position of the wrongdoer. Not every member of a generation shares equally in the maintenance of its social and cultural practices. Those who seek or inherit positions of influence and prominence

---

167 See Walvin, supra note 164, at 264.
168 See Burnard, supra note 88, at 32–33, 150, 177–79.
169 See id. at 149–50.
170 See id. at 104.
171 See id.
bear special responsibility for those practices because they set an example for others and are in a position to effect change. In this sense, Thomas Thistlewood contrasts usefully with his contemporary, Landon Carter. Both men were slave owners—Thistlewood in Jamaica and Carter in Virginia. Thistlewood was notably more brutal with his slaves than was Carter, but was also a “nobody” in the society of his day—unknown outside his small Jamaican parish, and even there only for his beautiful garden. His influence—terrorizing as it was—spread no further than the boundaries of his 160-acre farm.

Carter, by contrast, was an heir to one of Virginia’s great families—the son of Robert “King” Carter, who owned more than 300,000 acres and more than 700 slaves and was, in the words of Rhys Isaac, “a grandee among the grandees.” Landon himself, even after sharing his father’s estate with siblings, was one of Virginia’s twelve richest men and the owner of more than 400 slaves. He was also a member of Virginia’s House of Burgesses, the presiding judge of his county court, the chair in 1774 of his county’s committee to boycott British goods, a friend of the powerful Lee family, and, although not a trained physician, a respected medical practitioner. He was, in short, a man of influence in his county and his colony’s social and political life.

Some such men chose to free some or all of their slaves. Landon Carter’s nephew, Robert Carter III, an even wealthier man than his uncle, began emancipating his 452 slaves in 1791, not much more than a decade after his uncle’s death, and continued the process through the rest of his life. Fellow Virginian George Wythe and

172 See id. at 7.
173 Isaac, supra note 1, at xvii.
174 See Burnard, supra note 88, at 9.
175 See id. at 9–10.
176 Isaac, supra note 1, at xvii.
177 Id. at 60.
178 Id. at 123–61.
179 Id. at 241.
180 Id. at xviii–xix.
181 See Isaac, supra note 1, at 237, 350 n.40.
182 See id. at 105–20.
183 See Andrew Levy, The First Emancipator: The Forgotten Story of Robert Carter, the Founding Father Who Freed His Slaves, at xi (2005). Although wealthier than his uncle Landon, Robert Carter III was less politically influential. See id. at xi–xii.
Pennsylvanian John Dickinson\textsuperscript{185} made similar, if less sweeping, decisions for emancipation. Other men, such as Virginian George Washington,\textsuperscript{186} held on to their slaves during their own lifetimes but emancipated them at death.

By contrast, Landon Carter—like fellow Virginians Thomas Jefferson\textsuperscript{187} and George Mason\textsuperscript{188}—emancipated no one either during his life or upon his death. Indeed, unlike Jefferson and Mason, he never spoke publicly against slavery or even registered hesitations about the institution in his private diary. Landon Carter had a greater opportunity than most men of his day to exert influence on his society’s continued endorsement of the ownership of human beings. With that opportunity came responsibility. It is entirely appropriate for us to weigh his failure to shoulder that responsibility as we assess his participation in the slaveholding culture of his time.

To some, it might seem viscerally unfair to condemn centuries-old acts that were not unambiguously immoral in their own time but have become so in ours. The question that this Article explores, however, is whether such a declaration of historical injustice is any \textit{more} unfair than the decision our legal system makes to condemn a recent immigrant from a foreign culture for an act that is not unambiguously immoral in his own society but is in ours. There is no powerful reason to think that it is.

Indeed, judging a historical event may well be \textit{less} unfair than judging an immigrant’s conduct. A person charged with a crime risks jail time and a monetary fine if convicted. His cultural defense is thus grounded in due process concerns; he claims that he should not have to surrender liberty and property for conduct that was not actually culpable. In other words, the consequences of conviction are severe enough to the defendant that he is in some sense \texti{owed} consideration of his individuating cultural circumstances. By contrast, when we reflect on the conduct of a person from the past, nothing we say or do can have any impact on his liberty or property. He is gone. If we judge


\textsuperscript{188} See Mason Mystery, http://mason.gmu.edu/~rmellen/masonmystery.htm (last visited Aug. 24, 2006) (stating that Mason “was one of the first Americans and one of the very first southern plantation owners to denounce slavery, yet he owned slaves until the day he died”).
his conduct by the standards of his time rather than ours, we do so not because we actually owe him anything or because he might suffer the physical, emotional, or financial consequences of our judgment. Perhaps we feel that we owe his memory a certain sort of consideration. But surely living people who risk hard time in jail have a greater claim to careful and contextually sensitive judgment than do people who have passed on.

It bears emphasis that this Article does not contend that prior generations are unentitled to contextually-sensitive judgment. Indeed, even as to actors from prior generations who can be said to have chosen to act in accordance with the wisdom of their times, rather than embrace a mainstream challenge to it, this Article does not argue against contextually-sensitive judgment. The Article maintains simply that those historical actors do not deserve to be wholly excused for the choices they made. It will usually turn out, however, that they do not deserve to be wholly condemned either.

Again, the criminal law helpfully models the point. Where evidence of cultural influence on behavior does not substantiate a traditional defense to criminal liability such as duress or necessity, that evidence is not simply eliminated from consideration. Rather, it reappears at the defendant’s sentencing hearing, where the defendant’s lawyer offers it anew to place the defendant’s behavior in context and show how it is less culpable than similar behavior by a person from the dominant culture.189 The purpose of a sentencing hearing is to develop as complete an account as possible of the offender, her background, and the circumstances of her offense, so that the judge may tailor the offender’s punishment to her culpability.190 The inquiry is “broad in scope, largely unlimited either as to the kind of information [the sentencer] may consider, or the source from which it may come.”191 Thus, the defendant who fails in his effort at outright acquittal on grounds of culture nonetheless has the opportunity to argue for a lesser sentence on the same grounds. Moreover, such an opportunity should often succeed because a person who harms another in a way common to his foreign culture is, in fact, less culpable

189 See Maguigan, supra note 15, at 62–69 (discussing the use of cultural background information in plea bargaining and sentencing proceedings); Neff, supra note 97, at 445 (advocating consideration of cultural background information in sentencing). See generally Sikora, supra note 15 (arguing a defendant’s cultural circumstances should be allowed to serve as a mitigating factor in sentencing).

190 See United States v. Lynch, 934 F.2d 1226, 1235 (11th Cir. 1991).

than an otherwise similarly-situated American who causes the same harm with no cultural support.

Something quite similar obtains when we consider the case of a historical actor who chose to embrace rather than challenge the harmful mores of his day. Even if we decide that a monolithic morality in his era did not completely prevent him from choosing different actions, we still owe him careful consideration of the circumstances of his time. We must ask how much debate on the moral question his society actually knew, and whether he had access to it. We must ask what costs the historical actor would have faced for making a different choice, and whether he was able to shoulder such a burden. The answers to these questions will, and should, temper our judgment of the choice he made. But they should not direct us to suspend our judgment entirely. Such considerations should not lead to the whitewashing position of the partisan revisionist who dismisses every claimed injustice as a mere product of its time.

A final example will illustrate the point. Consider a hypothetical Virginia planter who purchased slaves in 1690. We already have seen that slave acquisition in the mid-nineteenth century was an inevitably contestable choice.\textsuperscript{192} This was markedly less true in 1690.\textsuperscript{193} A religious condemnation of slavery was certainly emerging. It was, however, new; only in 1688 did an American religious movement publicly articulate a case against slavery.\textsuperscript{194} And that movement was the Quakers—one centered in the mid-Atlantic colonies, with which our hypothetical Virginia planter would have been, at best, barely familiar. Other anti-slavery voices, most notably that of British slavery abolitionist Thomas Tryon, were also starting to be heard late in the seven-

\textsuperscript{192} See supra notes 161–67 and accompanying text.

\textsuperscript{193} In the late-seventeenth century, the practice of slavery in the American colonies was still in the process of formation. See Dwight Lowell Dumond, Antislavery: The Crusade for Freedom in America 5–12 (1961) (discussing the evolving legal status of the institution of slavery in American colonies during the late-seventeenth to early-eighteenth centuries). Although Philippe Rosenberg argues powerfully that “the development of anti-slavery opinion in Britain was coextensive with the institutionalization of slavery itself,” Philippe Rosenberg, Thomas Tryon and the Seventeenth-Century Dimensions of Antislavery, 61 Wm. & Mary Q. 3d 609, 640 (2004), it is nonetheless clear that late-seventeenth-century antislavery sentiment was neither broad-based nor loud.

\textsuperscript{194} In 1688, a group of Pennsylvania Quakers issued the Germantown Petition, which maintained that slavery was inconsistent with Christian principles. See David Brion Davis, The Problem of Slavery in Western Culture 308 & n.25 (1966); Thomas E. Drake, Quakers and Slavery in America 11–14 (1950).
teenth century, but they were relatively few and had not yet begun to draw significant attention in the colonies. The development of an organized abolitionist movement in the colonies was still many decades off. Slave ownership in the Virginia of the 1690s therefore stood in a different position from the slave ownership that followed 150 years later. In assessing the wrongfulness of a planter’s slave ownership in 1690s Virginia, we must temper our judgment to take account of that different position.

VI. Fixing the Hole: A Return to the Example of the Japanese American Internment

To begin filling the hole in reparations theory torn open by partisan historical revisionists, this Article has suggested an analogy between temporal and cultural excuse as a provocative and helpful starting place for a theory of historic injustice. The Article cited recent revisionist justifications of the Japanese American internment during World War II as the leading example of the revisionist threat. It will be helpful now to return to that example, to see how the lessons of the analogy to the criminal law’s cultural defense help bolster the

---

195 See Rosenberg, supra note 193, at 612. But see Burnard, supra note 88, at 105–06 (“Until 1750, antislavery sentiment was close to nonexistent.”).
196 See Dillon, supra note 166, at 87–111 (discussing growing abolitionist sentiment in early-nineteenth-century United States).
197 A difficult, perhaps intractable, problem remains. Must we say that slavery in a historical era in which slave owners had no access whatsoever to an antislavery discourse was not wrong in its time? Or has the entire human drama been played before a “natural law” backdrop that would condemn some practices as wrong regardless of the context in which they occurred?

This question is beyond this Article’s scope. Two observations are, however, in order. First, it is probably inaccurate to claim there has ever been a slave owner at any time in history who had truly no access whatsoever to an antislavery discourse. Surely, there was always such a discourse among his slaves—a discourse from which the slave owner no doubt insulated himself.

Second, a very careful inquiry into the true diversity of opinion and argument in a given historical period is essential before resorting to natural law. It is tempting to view an era’s prevalent discourse as more monolithic than it actually was. Trevor Burnard, for example, asserts that “until 1750, antislavery sentiment was close to nonexistent.” Burnard, supra note 88, at 105–06. Yet Philippe Rosenberg, a scholar whose work focuses on the precise issue of pre-1750 antislavery discourse, asserts otherwise and identifies a significant number of published critiques of slavery in the British world in the late-seventeenth and early-eighteenth centuries. See Rosenberg, supra note 193, at 626, 640. If Rosenberg is correct, and there never really was a period of institutionalized slavery in the British Atlantic without antislavery argument or agitation, then it may not be necessary to resort entirely to natural law to assess the wrongfulness of late-seventeenth-century slaveholding in its time.

198 See supra notes 70–85 and accompanying text.
conclusion that the Roosevelt Administration’s policies were an injustice even in their own time.

The plea on behalf of those who designed and approved the program of Japanese American exclusion and detention is that, in the context of the vicious Japanese surprise attack on Pearl Harbor and the military threat that Japanese forces posed to the United States mainland in the following months, the mass prophylactic detention of all people of Japanese ancestry along the West Coast was necessary and understandable. The argument is that the American people were justifiably shocked and frightened by Pearl Harbor and believed this preventive measure was needed. As Alan Simpson stated on the floor of the U.S. Senate during debate on what would become the Civil Liberties Act of 1988, “at that time, in most every structure of our citizenry, our Government and our bureaucracy, it seemed the very right thing to do.” 199 If this observation is accurate, then it would be very difficult to fault the architects of the government’s program for setting up a structure that was universally considered wise. To fault them, it seems, would just be, as Michelle Malkin argues, an unfair condemnation of “serious men . . . who did not have the luxury of a rearview mirror.”

The trouble with this view is that the historical record does not support a story of monolithic support and approval of the government’s program. Not every American believed in the military necessity of evicting and jailing the West Coast’s Japanese American population. Indeed, very powerful and articulate people in the American mainstream opposed the program. And Franklin Roosevelt, who gave the order approving the War Department’s eviction and detention of Japanese Americans, had easy access to these substantial opposing views.

When Roosevelt signed Executive Order 9066, he knew that the top officials in his Justice Department had vigorously opposed it. 201 In debates with War Department officials, James Rowe and Edward Ennis, two top deputies to Attorney General Francis Biddle, argued that the forcible relocation of American citizens of Japanese ancestry (as opposed to Japanese aliens) would be illegal. 202 Rowe, 203

---


200 Malkin, supra note 6, at xxxiv.

201 Peter Irons’s book Justice at War is the most detailed account of the interdepartmental wrangling that preceded Roosevelt’s decision to approve of the War Department’s proposal to evict Japanese Americans from the West Coast. See Irons, supra note 55, at 25–74.

202 See id. at 55, 62.

203 See id. at 44.
nis, Biddle, and Federal Bureau of Investigation ("FBI") Director J. Edgar Hoover all maintained that a program of mass removal and relocation was entirely unnecessary. Biddle made this position clear to Roosevelt in a meeting on February 7, 1942, just four days before Roosevelt orally approved the War Department’s plan. In short, it is simply not true that eviction and incarceration seemed “the right thing to do” to everyone who had lived through the trauma of Pearl Harbor. It seemed the wrong thing to Roosevelt’s own Attorney General and to the Director of the FBI.

It also seemed the wrong thing to the two men who had preceded Francis Biddle in the Attorney General’s Office in the Roosevelt Administration—Frank Murphy (1939–1940) and Robert H. Jackson (1940–1941)—both of whom had become Associate Justices of the U.S. Supreme Court by the time of the Pearl Harbor attack. Reluctantly concurring in the Court’s 1943 decision upholding the constitutionality of a dusk-to-dawn curfew imposed on Japanese Americans, Justice Murphy described the race-based curfew as bearing “a melancholy resemblance to the treatment accorded to members of the Jewish race in Germany and in other parts of Europe.” The curfew, Murphy said, went “to the very brink of constitutional power.”

A year later, Justice Murphy dissented from the Court’s 6–3 decision upholding the Administration’s program evicting Japanese Americans from their homes and indefinitely excluding them from the West Coast. Murphy’s condemnation of the program was stark. Citing his words of a year earlier, Murphy charged that the Administration’s program went “over ‘the very brink of constitutional power’ . . . into the ugly abyss of racism.” It stemmed from “an erroneous assumption of racial guilt rather than bona fide military necessity,” and the justifications offered by the government were nothing but “an accumulation of much of the misinformation, half-truths and insinuations that for years have been directed against Japanese Americans by people with racial and economic prejudices.”

204 See id.
205 See id. at 53.
207 See id. at 53.
208 Hirabayashi v. United States, 320 U.S. 81, 111 (1943) (Murphy, J., concurring).
209 Id.
210 Korematsu v. United States, 323 U.S. 214, 233–42 (1944) (Murphy, J., dissenting).
211 Id. at 233.
212 Id. at 235–36.
213 Id. at 239.
the Administration’s program a “legalization of racism, . . . unattractive in any setting but . . . utterly revolting among a free people who have embraced the principles set forth in the Constitution of the United States.”

Justice Jackson, not just a former U.S. Attorney General but also a close friend and confidante of President Roosevelt, also condemned the eviction and exclusion of Japanese Americans on the basis of the military’s flimsy claim of necessity. In Korematsu v. United States in 1944, Jackson emphasized that the only thing that made Fred Korematsu’s continued presence in California a crime was the ancestry he inherited from his parents. His conviction for violating the military’s exclusion order therefore violated the “fundamental assumption underlying our system . . . that guilt is personal and not inheritable.” To “approve [what] the military . . . deem[ed] expedient,” said Justice Jackson, the Court had to “distort the Constitution” and to “validate the principle of racial discrimination in criminal procedure and of transplanting American citizens.”

Neither is it the case that the American people as a whole believed in the military necessity of incarcerating the West Coast’s entire Japanese American population. Well-respected Americans publicly and articulately opposed the plan. Consider, for example, a letter that a group of prominent American educators, politicians, businesspersons, and artists sent to President Roosevelt on April 30, 1942. Although they said they “recognize[d] fully the difficulties of the situation” involving Japanese Americans on the West Coast, they maintained that they had seen “no adequate evidence . . . that an order giving complete power to the Secretary of War or to the commander of each military area to exclude from designated areas all citizens, or to restrict

---

214 Id. at 242.
216 Korematsu, 323 U.S. at 243 (Jackson, J., dissenting).
217 Id.
218 Id. at 244–45.
219 Id. at 246. Justice Roberts also dissented in the Korematsu case, condemning the program of eviction and exclusion as a clear violation of constitutional rights. See id. at 225–26 (Roberts, J., dissenting) (stating that the case was one of “punishment of an individual for not submitting to imprisonment in a concentration camp, based on his ancestry, and solely because of his ancestry, without evidence or inquiry concerning his loyalty and good disposition towards the United States”).
their actions in any way he sees fit, is either constitutional or democratic.” Singling out “the Japanese alone” for these burdens, they argued, “approximates the totalitarian theory of justice practiced by the Nazis in their treatment of the Jews.” Signatories to this blunt and contemporaneous condemnation of the government’s evacuation policy included Alfred M. Bingham, the editor of the progressive journal Common Sense; George S. Counts, the well-known sociologist and educational reformer; Countee Cullen, the African-American poet; John Dewey, the founder of Pragmatist philosophy and educational reformer; Sherwood Eddy, the national secretary of the YMCA; clergyman Harry Emerson Fosdick; University of North Carolina President Frank Porter Graham; prominent Unitarians Mary W. Hillyer and John Haynes Holmes; James Wood Johnson, the co-founder and ex-president of the Johnson & Johnson Company; Christian intellectual Reinhold Neibuhr; Clarence E. Pickett, the executive secretary of the American Friends Service Committee; pastor and presidential candidate Norman Thomas; Mount Holyoke ex-president Mary E. Woolley; and others. These were no fringe figures in American intellectual and political life, and they made their dissent known to the President at the very moment that the Administration was laying its plans for long-term exclusion and detention.

Consider also an editorial by the highly regarded Washington Post editorialist Merlo Pusey entitled “War vs. Civil Rights,” published in May of 1944. Pusey maintained that when the war was over, the nation would be “very much ashamed” of the “mistreatment of loyal American citizens of Japanese origin.” Pusey pointed out that the nation had seen “no sabotage by a Japanese American in this war,” and that “[a]n overwhelming majority of those who were ousted from the Western States have never committed a crime of any sort and have always remained loyal to the United States.” Pusey demanded to know why “the Administration . . . continues to punish loyal citizens solely because of their racial origin.” Predicting that the courts would “compel the Administration to retreat from what many of its own officials recognize to be an indefensible position,” the columnist called the exclusion and internment policies “black patches on an
otherwise very creditable record in the protection of civil rights.” 227

These are not the words of the lunatic fringe; they are the words of a well-respected, nationally syndicated political columnist.

Merlo Pusey and the signatories to the April 1942 letter to the President were hardly alone in their criticism of the government’s policies. A March 1943 article in *The Reader’s Digest* criticized the policies as a costly mistake, noting that the supposedly dangerous Japanese American population had been left more or less untouched in Hawaii, that German Americans and Italian Americans on the East Coast posed similar supposed dangers but also went untouched, and that individualized loyalty inquiries would have been possible for a fraction of the cost. 228 In June of 1942, Charles Iglehart wrote in the pages of *The Nation* that “even as a war measure evacuation was unnecessary.” 229 “The slumbering embers of public antagonism to the alien group,” he wrote, “were . . . deliberately fanned by interested persons and organizations until a conflagration was threatened, but at any time it could have been quenched if the authorities had shown the proper firmness.” 230 Writing in *Commonweal* in October of 1943, Harold J. Filsicker argued that the military situation along the West Coast was “well in hand” before the first evacuation orders were issued, and that any potential danger to the Coast was “practically over” before the process of exclusion was complete. 231 Articles critical of the Japanese American program also appeared in *The New Republic*, *Business Week*, *Christian Century*, *Asia*, *Collier’s*, *Time*, *Newsweek*, *Harper’s*, *Fortune*, and *The Saturday Evening Post* magazines, among others. 232

Well-respected figures quite near the centers of American political power and public life vocally opposed the wholesale exclusion and incarceration of American citizens of Japanese ancestry during World War II. These prestigious opposing voices surely reached the ears of the men who designed and implemented the program, as well as of the President who authorized it and allowed it to operate for several years. The program was not remotely foreordained; it was rather their choice. In Kay Levine’s words, these were men who were “able to oper-

---

227 Id.
230 Id.
ate outside of the schema [of their time] but simply chose not to do so.”

Their choice was a wrong one that shattered the lives of tens of thousands of innocent Americans. The historical moment in which they lived was not monolithic, and it does not excuse or minimize the injustice of the choices they made. Neither does it undermine the extraordinary apology and redress that former internees and their children and allies managed to secure in the 1980s and 1990s.

**Conclusion**

In June of 2005, Wachovia, the nation’s fourth largest bank, made a stunning announcement. Research had revealed that two of its predecessor corporations had been involved in the market for

---

233 Levine, supra note 121, at 48.

234 Neither, it should go without saying, does their historical moment justify the choice they made. In the criminal law, a defense of justification differs from a defense of excuse in an important moral sense. Conduct that is excusable is conduct that is morally wrong but contextually understandable. Joshua Dressler, UNDERSTANDING CRIMINAL LAW 203 (3d ed. 2001). Conduct that is justifiable is conduct that is morally correct—the right thing to do in the circumstances. See Mitchell N. Berman, Justification and Excuse, Law and Morality, 53 DUKE L.J. 1, 7–9 (2003).

The criminal law’s cultural defense is a defense of excuse: the defendant’s claim is not that he acted rightly, but that his cultural context relieves him of responsibility for the wrong choice he made. See Gordon, supra note 97, at 1810. Yet at times, the most enthusiastic defenders of the wrongdoing of prior generations speak the language of justification rather than excuse. When U.S. Representative Howard Coble recently defended the government’s wartime treatment of Japanese Americans, he did not contend that the mass incarceration of tens of thousands of American citizens was a contextually understandable judgment. He maintained that it was a correct judgment, justified both to protect Japanese Americans from vigilante violence and to protect the United States from those Japanese Americans who meant the country harm. See Associated Press, N.C. Rep.: Internment Camps Were Meant to Help, FoxNews.com, Feb. 5, 2003, http://www.foxnews.com/story/0,2933,77677,00.html.

This sort of enthusiasm for the wrongdoing of the past is doubly dangerous. Not only does it carry all of the risk of a recurrence of the tragedy that this Article has described, but it also entails a disturbing conclusion about people who resisted the action that caused them harm. The criminal law shows this clearly: a person who resists a justifiable aggressive act acts culpably, whereas a person who resists an excusable aggressive act does not. See Joshua Dressler, New Thoughts About the Concept of Justification in the Criminal Law: A Critique of Fletcher’s Thinking and Rethinking, 32 UCLA L. REV. 61, 61–62 n.2 (1984). A person may lawfully resist an insane person who threatens deadly force, but may not lawfully resist a person who acts in self-defense. See id.

The consequence of justifying (as opposed to excusing) the harmful acts of a prior generation is therefore to condemn those who resisted them. See James W. Loewen, Lies Across America 28 (1999) (”[I]t is hard for residents of Edgefield to honor Americans who fought against the Vietnam War so long as their downtown monument credits those who fought in the war for being right.”). If the eviction and incarceration of Japanese Americans was a justified program, then Fred Korematsu acted wrongly when he decided not to report for detention as ordered. See generally Korematsu, 323 U.S. 214.
slaves in the nineteenth century: The Georgia Railroad and Banking Company had owned 162 slaves between 1836 and 1842, and the Bank of Charleston had accepted 529 slaves as collateral on loans and mortgages between 1841 and 1860. “We are deeply saddened by these findings,” Wachovia said. “[We] apologize to all Americans, and especially to African-Americans and people of African descent.” As compensation, the bank committed to donate $11 million to charities emphasizing African-American educational causes. Wachovia was, moreover, not the only American bank to make such an announcement in 2005; JP Morgan Chase, Bank of America, and Lehman Brothers each made similar announcements about predecessor entities and committed to similar compensatory schemes during 2005.

This did not sit well with Jeff Jacoby, a syndicated columnist for the Boston Globe. In a column entitled The Slavery Shakedown, Jacoby rehearsed the common arguments against reparations for slavery: “Living white Americans bear no culpability for slavery,” he argued, “and living black Americans never suffered from it.” Jacoby added a new twist, however: not only was Wachovia’s trade in slaves too remote to permit reparations, but “[t]he slaves for which [Wachovia] was so apologetic were owned decades before the Civil War, when slavery was

238 See Binyamin Appelbaum, Bank Ups Giving to Black Causes, CHARLOTTE OBSERVER, July 29, 2005, at 1D.
239 Id.
242 All of the corporate soul-searching research was prompted by a Chicago ordinance that requires all banks doing business with the city to disclose their past ties to slavery. See City of Chicago Reparation Ordinance Gets Support from Mayor Richard M. Daley, U.S. Conference of Mayors, Oct. 21, 2002, http://www.usmayors.org/uscma/us_mayor_newspaper/documents/10_21_02/chicago.asp.
still lawful throughout the South.”244 That last clause bears repeating: “when slavery was still lawful throughout the South.” The mind struggles to grasp the relevance of the fact that slavery remained technically lawful in the South in the two or three decades before the Civil War. Could it be that the syndicated columnist was subtly offering a substantive defense of the ownership of slaves?

That is precisely what he was doing. A line or two later in the column, Jacoby summarized his criticism of Wachovia’s slavery apology: it was contrition “for something Wachovia didn’t do, in an era when it didn’t exist, under laws it didn’t break.”245 This is not simply an argument that Wachovia’s predecessors’ wrongs were too remote in time to redress. It is an argument that Wachovia’s predecessors broke no law—that is, that there is no wrong to redress.246 According to Jacoby, the argument for reparations is not just impractical or logically flawed. It makes “a mockery of historical truth”247—the historical “truth” that these banks’ ownership of human beings between 1836 and 1860 was not an injustice, because it was lawful.248

The supposed concern for “historical truth” and the insistence that slave owners be judged under the legal code of their time should look familiar. These are the strategies of partisan historical revisionism—the germ of Michelle Malkin’s justification of racial internment, now spread to American chattel slavery.

Reparations theory is woefully unprepared to meet the challenge of partisan revisionism. It has a rich language for debating and specifying the remedies for historical injustice, but it lacks even a basic language for specifying what ought to count as a historical injustice. This Article has argued that the criminal law’s cultural defense, focusing, as it does, on the extent to which a perpetrator consciously chose to inflict harm, offers an analogy from which a language of historical injustice might develop. It is not a perfect analogy. By itself, it cannot plug the growing hole in the foundation of reparations theory. That important repair job will require more than simply a single provocative analogy from the criminal law—it will require a conversation

244 Id.
245 Id. (emphasis added).
246 At other points in the column, Jacoby spoke of slavery as “wrongdoing,” which is inconsistent with his insistence that the banks had broken no laws. See id.
247 Id. (emphasis added).
248 The National Legal and Policy Center, probably the leading lobbyist against slavery reparations, makes this argument as well. See Peter Flaherty & John Carlisle, The Case Against Slave Reparations 5 (2004), available at http://www.nlpc.org/pdfs/Final_NLPC_Reparations.pdf (“As tragic as slavery was, it was legal in the South between 1789 and 1865.”).
among historians, philosophers, and legal scholars. But time is of the essence. Partisan revisionists are hard at work weakening the foundations of redress and of American historical understanding, and they are on the television, radio, and the best-seller lists. The conversation must begin promptly.
MISSING THE FOREST FOR A TREE:
UNPUBLISHED OPINIONS AND
NEW FEDERAL RULE OF
APPELLATE PROCEDURE 32.1

SCOTT E. GANT*

Abstract: On December 1, 2006, Federal Rule of Appellate Procedure 32.1 will take effect, allowing citation to all opinions issued on or after January 1, 2007 that have been designated “unpublished” or “non-precedential.” The new Rule, under consideration by the Judicial Conference of the United States since the 1990s, seemingly puts an end to the long and sometimes contentious debate over whether citation to unpublished opinions should be permitted. But the Rule does not address a more important issue: whether the federal courts of appeals should designate some of their opinions as nonprecedential. This Article argues the notion that judges can and should determine an opinion’s precedential value at the time they issue it is based upon a flawed and outdated view of how the law develops. Whether an opinion has made “new law” or is otherwise significant is a judgment best made with the benefit of time, and with input from lawyers, litigants, and other judges.

Introduction

On December 1, 2006, new Federal Rule of Appellate Procedure 32.1 will take effect, displacing the array of rules in the individual federal appeals courts governing citation to “unpublished” opinions and other case dispositions designated as nonprecedential, and imposing in their place a uniform rule allowing citation to such decisions.1 The Rule’s enactment (in which both Chief Justice Roberts and Justice Alito played significant roles) follows several years of vigorous debate not ordinarily associated with consideration of a

rule of procedure, let alone the typically quiet rulemaking process for the federal courts of appeals.\(^2\)

Unfortunately, the furor over the new citation rule has overshadowed a more important and divisive issue—whether it is appropriate for appeals courts to designate some (in fact, most) of their decisions as nonprecedential.\(^3\) The sponsors of Rule 32.1 were aware of the controversy that has emerged during the past decade regarding the wisdom and constitutionality of this practice, but they steered clear of it, emphasizing early on that the Rule is “extremely limited” and avoids taking any position on this highly-charged question.\(^4\)

Supporters of unpublished opinions principally contend that the opinions enable appeals courts to conserve and sensibly allocate scarce judicial resources, and promote clarity, uniformity, and cohesiveness of the law. Critics of unpublished opinions charge that they permit appeals courts to suppress precedent, lead to the dedication of insufficient attention to unpublished dispositions, and are inconsistent with principles of judicial accountability.\(^5\) These arguments

---

\(^2\) See infra notes 67–100 and accompanying text.

\(^3\) District courts are not faced with the challenge of deciding which, if any, of their opinions should be designated as unpublished because, as a formal matter, no district court opinions are “precedential.” See Colby v. J.C. Penney Co., 811 F.2d 1119, 1124 (7th Cir. 1987) (“[D]istrict judges in this circuit must not treat decisions by other district judges, in this and a fortiori in other circuits, as controlling, unless of course the doctrine of res judicata or of collateral estoppel applies. . . . [T]he responsibility for maintaining the law’s uniformity is a responsibility of appellate rather than trial judges . . . .”); Nat’l Union Fire Ins. Co. v. Allfirst Bank, 282 F. Supp. 2d 339, 351 (D. Md. 2003) (“Of course, no decision of a district court judge is technically binding on another district court judge, even within the same district.”) (citation omitted); In re Oxford Health Plans, Inc., 191 F.R.D. 369, 377 (S.D.N.Y. 2000) (“Principles of stare decisis do not require this Court to give any deference to decisions of another district judge.”) (citation omitted).

\(^4\) Fed. R. App. P. 32.1 (proposed) advisory committee’s note (“Rule 32.1 is extremely limited. It takes no position on whether refusing to treat an ‘unpublished’ opinion as binding precedent is constitutional. It does not require any court to issue an ‘unpublished’ opinion or forbid any court from doing so. It does not dictate the circumstances under which a court may choose to designate an opinion as ‘unpublished’ or specify the procedure that a court must follow in making that decision. It says nothing about what effect a court must give to one of its ‘unpublished’ opinions or to the ‘unpublished’ opinions of another court.”) (citations omitted); Letter from Seth P. Waxman, U.S. Solicitor General, to Judge Will Garwood, U.S. Court of Appeals for the Fifth Circuit 1 (Jan. 16, 2001) (on file with author) (“Although this is a sensitive topic, I believe that [proposed Rule 32.1] is narrowly framed and focused solely on citation rules that, by their nature, are an appropriate topic for national rule-making.”).

\(^5\) Most of the arguments for and against unpublished opinions are well covered in other articles and will not be repeated here. See generally Richard S. Arnold, Unpublished Opinions: A Comment, 1 J. App. Prac. & Process 219 (1999); Stephen R. Barnett, From
were featured in the wrangling over whether to enact a uniform rule governing citation to unpublished opinions, which culminated in adoption of the new Rule.\(^6\)

What has been missing from the debate over unpublished opinions, however, is an examination of its conceptual foundation. Underpinning the practice of designating certain opinions “unpublished” is the notion that appeals court judges themselves can and should determine an opinion’s precedential authority at the time they issue the opinion, based on their view about whether that opinion has made “new law” or is otherwise significant.\(^7\) That idea is predicated upon a flawed and outdated view of judicial decisionmaking and development of the law, and should be abandoned.\(^8\) Stripped of its anachronistic foundation, it becomes difficult to justify the existing

---


\(^7\) See *infra* notes 101–26 and accompanying text.

\(^8\) See *infra* notes 127–30 and accompanying text.
system of unpublished opinions, which is desperately in need of reconsideration.

I. UNPUBLISHED OPINIONS AND PUBLICATION RULES

A. The History of Unpublished Opinions

Today the designation of an opinion as “unpublished” refers to its status as nonprecedential. But when the rules regarding unpublished opinions emerged several decades ago, the designation of a decision as published or unpublished actually bore some relationship to its availability to lawyers and the public, as the label would suggest.

Commentators expressed concerns about the growing number of judicial opinions as early as the 1800s. The movement towards formal limitations on the publication of, and citation to, appellate rulings did not emerge until 1964, however, when the Judicial Conference of the United States (the “Judicial Conference”) resolved that courts of appeals should publish “only those opinions which are of general precedential value.” Seven years later, in 1971, the Federal Judicial Center (the “FJC”) observed in its annual report there was “widespread consensus that too many opinions are being printed or published or otherwise disseminated.” The following

9 See Robert J. Martineau, Restrictions on Publication and Citation of Judicial Opinions: A Reassessment, 28 U. Mich. J.L. Reform 119, 121 & n.3 (1994). Some American commentators voiced concerns about the proliferation of judicial decisions as early as the first half of the nineteenth century, and British commentators voiced concerns as early as the late 1700s. See David Greenwald & Frederick A. O. Schwarz, Jr., The Censorial Judiciary, 35 U.C. Davis L. Rev. 1133, 1144–45 (2002).

10 The Judicial Conference was created by Congress to make policy with regard to the administration of the U.S. courts. See 28 U.S.C. § 331 (2000 & Supp. III 2003). It is comprised of the chief judge of each circuit, the chief judge of the Court of International Trade, and a district judge from each judicial circuit chosen by appellate and district judges from that circuit, with the Chief Justice of the United States as its presiding officer. See id.


year, the FJC’s Board recommended that the Judicial Conference instruct the courts of appeals to adopt procedures for publishing only some of their opinions and adopt rules limiting citation to unpublished opinions.\footnote{14 See Judicial Conference of the U.S., Report of the Proceedings of the Judicial Conference of the United States 33 (1972); see also William L. Reynolds & William M. Richman, The Non-Precedential Precedent—Limited Publication and No-Citation Rules in the United States Courts of Appeals, 78 Colum. L. Rev. 1167, 1170 (1978) (discussing the FJC Board’s recommendation).} Heeding that recommendation, in October 1972 the Judicial Conference directed the courts of appeals to develop their own plans for selective publication of opinions.\footnote{15 See Judicial Conference of the U.S., Report of the Proceedings of the Judicial Conference of the United States 33 (1972). In 1973, the Advisory Council for Appellate Justice issued a report urging appellate courts to adopt publication rules to reduce the number of published opinions. See Comm. on Use of Appellate Court Energies of the Advisory Council on Appellate Justice, Standards for Publication of Judicial Opinions 5 (1973). The Council was jointly sponsored by the Federal Judicial Center and the National Center for State Courts. See Martineau, supra note 9, at 122 n.5.} By 1974 each court had developed its own plan, and the courts implemented them over the next several years.\footnote{16 See Greenwald & Schwarz, supra note 9, at 1142; Penelope Pether, Inequitable Injunctions: The Scandal of Private Judging in the U.S. Courts, 56 Stan. L. Rev. 1435, 1443–44 (2004); Reynolds & Richman, supra note 14, at 1171; see also Fed. R. App. P. 47(a) (authorizing courts of appeals to enact local rules).}

Initially, opinions designated as unpublished were available to anyone who wished to pay a visit to the clerk’s office at each court of appeals, but were not otherwise disseminated to the public or to legal publishers. Over time, however, more and more “unpublished” opinions became widely available, primarily through private publishers such as West and Lexis.\footnote{17 See Robert C. Berring, Legal Research and Legal Concepts: Where Form Molds Substance, 75 Cal. L. Rev. 15, 17–20 (1987) (discussing the history of reporting on judicial decisions); William R. Mills, The Shape of the Universe: The Impact of Unpublished Opinions on the Process of Legal Research, 46 N.Y.L. Sch. L. Rev. 429, 440 (2002–03) (noting that “[c]ourts’ limited publication regimes have never completely prevented researchers from finding unpublished opinions, any more than their no-cite rules have prevented lawyers and judges from using them”). See generally Francine Biscardi, The Historical Development of the Law Concerning Judicial Report Publication, 85 Law Libr. J. 531 (1993) (discussing the history of judicial opinion compilation and publication).} Today, most unpublished opinions (which comprise approximately 80% of all appeals court dispositions)\footnote{18 Admin. Office of the U.S. Courts, Judicial Business of the U.S. Courts 42 tbl.S-3 (2005), available at http://www.uscourts.gov/judbus2005/tables/s3.pdf (showing that approximately 82% of dispositions on the merits during the 12-month period ending September 30, 2005 were “unpublished”). The Administrative Office of the U.S. Courts compiles annual statistics on the operation of the federal courts, including data...
West’s *Federal Appendix*, established in 2001, which includes every “unpublished” decision sent to it by courts of appeals.\(^{19}\) Moreover, the E-Government Act of 2002 requires that all opinions, published and unpublished, be posted on every federal court’s own website.\(^ {20}\)

### B. Publication and Citation Rules in the U.S. Courts of Appeals

For the past three decades, each federal appeals court has maintained, and occasionally revised, its own rules governing publication and citation.\(^ {21}\) Among them, only the U.S. Court of Appeals for the D.C. Circuit has dispensed with the practice of designating some of its dispositions as nonprecedential.\(^ {22}\) Although that court retains the distinction between “published” and “unpublished” opinions, both forms of opinions rendered after January 1, 2002 may be cited “as precedent.”\(^ {23}\) Each of the other courts of appeals currently has rules under which it relegates certain dispositions to nonprecedential status.\(^ {24}\)

---

19 See 1 *West’s Federal Appendix*, at iii (2001).


21 See *supra* notes 15–16 and accompanying text.

22 See D.C. Cir. R. 28(c)(1)(B).

23 *Id.* But see D.C. Cir. R. 36(c)(2) (“While unpublished orders and judgments may be cited to the court in accordance with Circuit Rule 28(c)(1)(B), a panel’s decision to issue an unpublished disposition means that the panel sees no precedential value in that disposition.”).

24 See, e.g., 1st Cir. R. 36(a) (“[W]here opinions are likely not to break new legal ground or contribute otherwise to legal development, they are issued in unpublished form.”); 1st Cir. R. 36(b)(1) (“In general, the court thinks it desirable that opinions be published and thus be available for citation. The policy may be overcome in some situations where an opinion does not articulate a new rule of law, modify an established rule, apply an established rule to novel facts or serve otherwise as a significant guide to future litigants.”); 2d Cir. R. § 0.23 (“[W]hen no jurisprudential purpose would be served by a written opinion, disposition will be made in open court or by summary order . . . . Where disposition is by summary order, the court may append a brief written statement to that order. . . . [But] these statements do not constitute formal opinions of the court . . . .”); 4th Cir. R. 36(a) (“Opinions delivered by the Court will be published only if the opinion satisfies one or more of the standards for publication: (i) It establishes, alters, modifies, clarifies, or explains a rule of law within this Circuit; or (ii) It involves a legal issue of continuing public interest; or (iii) It criticizes existing law; or (iv) It contains a historical review of a legal rule that is not duplicative; or (v) It resolves a conflict between panels of this Court, or creates a conflict with a decision in another circuit.”); 5th
Cir. R. 47.5.1 ("The publication of opinions that merely decide particular cases on the basis of well-settled principles of law imposes needless expense on the public and burdens on the legal profession. However, opinions that may in any way interest persons other than the parties to a case should be published. Therefore, an opinion is published if it: (a) Establishes a new rule of law, alters, or modifies an existing rule of law, or calls attention to an existing rule of law that appears to have been generally overlooked; (b) Applies an established rule of law to facts significantly different from those in previous published opinions applying the rule; (c) Explains, criticizes, or reviews the history of existing decisional or enacted law; (d) Creates or resolves a conflict of authority either within the circuit or between this circuit and another; (e) Concerns or discusses a factual or legal issue of significant public interest; or (f) Is rendered in a case that has been reviewed previously and its merits addressed by an opinion of the United States Supreme Court. An opinion may also be published if it: Is accompanied by a concurring or dissenting opinion; or reverses the decision below or affirms it upon different grounds."); 6th Cir. R. 206(a) ("The following criteria shall be considered by panels in determining whether a decision will be designated for publication . . . : (1) whether it establishes a new rule of law, or alters or modifies an existing rule of law, or applies an established rule to a novel fact situation; (2) whether it creates or resolves a conflict or [sic] authority either within the circuit or between this circuit and another; (3) whether it discusses a legal or factual issue of continuing public interest; (4) whether it is accompanied by a concurring or dissenting opinion; (5) whether it reverses the decision below . . . ; (6) whether it addresses a lower court or administrative agency decision that has been published; or (7) whether it is a decision that has been reviewed by the United States Supreme Court."); 7th Cir. R. 53(c)(1) ("A published opinion will be filed when the decision (i) establishes a new, or changes an existing rule of law; (ii) involves an issue of continuing public interest; (iii) criticizes or questions existing law; (iv) constitutes a significant and non-duplicative contribution to legal literature (A) by a historical review of law, (B) by describing legislative history, or (C) by resolving or creating a conflict in the law; (v) reverses a judgment or denies enforcement of an order when the lower court or agency has published an opinion supporting the judgment or order; or (vi) is pursuant to an order of remand from the Supreme Court and is not rendered merely in ministerial obedience to specific directions of that Court."); 7th Cir. R. 53(c)(2) ("When the decision does not satisfy the criteria for publication [as set out in Seventh Circuit Rule 53(c)(1)] it will be filed as an unpublished order."); 9th Cir. R. 36-2 ("A written, reasoned disposition shall be designated as an OPINION only if it: (a) Establishes, alters, modifies or clarifies a rule of law, or (b) Calls attention to a rule of law which appears to have been generally overlooked, or (c) Criticizes existing law, or (d) Involves a legal or factual issue of unique interest or substantial public importance, or (e) Is a disposition of a case in which there is a published opinion by a lower court or administrative agency, unless the panel determines that publication is unnecessary for clarifying the panel’s disposition of the case, or (f) Is a disposition of a case following a reversal or remand by the United States Supreme Court, or (g) Is accompanied by a separate concurring or dissenting expression, and the author of such separate expression requests publication of the disposition of the Court and the separate expression."); 11th Cir. R. 36-3, I.O.P. 6 ("Opinions that the panel believes to have no precedential value are not published."); D.C. Cir. R. 36(a)(2) ("An opinion, memorandum, or other statement explaining the basis for the court’s action in issuing an order or judgment will be published if it meets one or more of the following criteria: (A) with regard to a substantial issue it resolves, it is a case of first impression or the first case to present the issue in this court; (B) it alters, modifies, or significantly clarifies a rule of law previously announced by the court; (C) it calls attention to an existing rule of law that appears to have been generally overlooked; (D) it criticizes or questions existing law; (E) it resolves an apparent conflict in decisions..."
Some courts have dispensed with the term “unpublished” and simply refer to decisions as precedential or nonprecedential. Others specifically address the precedential or binding effect of an unpublished opinion. The Eighth and Tenth Circuits have no criteria

within the circuit or creates a conflict with another circuit; (F) it reverses a published agency or district court decision, or affirms a decision of the district court upon grounds different from those set forth in the district court’s published opinion; (G) it warrants publication in light of other factors that give it general public interest.”). See generally Stephen L. Wasby, Publication (or Not) of Appellate Rulings: An Evaluation of Guidelines, 2 SETON HALL CIR. REV. 41 (2005) (discussing nonpublication rules and norms, and how judges enforce them).

25 See 3d CIR. INTERNAL OPERATING P. 5.2 (“An opinion, whether signed or per curiam, is designated as precedential . . . when it has precedential or institutional value.”); 3d CIR. INTERNAL OPERATING P. 5.3 (“An opinion, whether signed or per curiam, that appears to have value only to the trial court or the parties is designated as not precedential . . . . A not precedential opinion may be issued without regard to whether the panel’s decision is unanimous and without regard to whether the panel affirms, reverses, or grants other relief.”); Fed. CIR. R. 47.6(b) (“An opinion or order which is designated as not to be cited as precedent is one determined by the panel issuing it as not adding significantly to the body of law.”); Fed. CIR. INTERNAL OPERATING P. 10.3 (“Disposition by nonprecedential opinion or order [means] that a precedential opinion would not add significantly to the body of law . . . .”); Fed. CIR. INTERNAL OPERATING P. 10.4 (“The court’s policy is to limit precedent to dispositions meeting one or more of these criteria: (a) The case is a test case. (b) An issue of first impression is treated. (c) A new rule of law is established. (d) An existing rule of law is criticized, clarified, altered, or modified. (e) An existing rule of law is applied to facts significantly different from those to which that rule has previously been applied. (f) An actual or apparent conflict in or with past holdings of this court or other courts is created, resolved, or continued. (g) A legal issue of substantial public interest, which the court has not sufficiently treated recently, is resolved. (h) A significantly new factual situation, likely to be of interest to a wide spectrum of persons other than the parties to a case, is set forth. (i) A new interpretation of a Supreme Court decision, or of a statute, is set forth. (j) A new constitutional or statutory issue is treated. (k) A previously overlooked rule of law is treated. (l) Procedural errors, or errors in the conduct of the judicial process, are corrected, whether by remand with instructions or otherwise. (m) The case has been returned by the U.S. Supreme Court for disposition by action of this court other than ministerial obedience to directions of the Court. (n) A panel desires to adopt as precedent in this court an opinion of a lower tribunal, in whole or in part.”).

26 See 1ST CIR. R. 36(c) (“[A] panel’s decision to issue an unpublished opinion means that the panel sees no precedential value in that opinion.”); 5TH CIR. R. 47.5.4 (“All unpublished opinions issued on or after January 1, 1996 are not precedent, except under the doctrine of res judicata, collateral estoppel or law of the case . . . .”); 6TH CIR. R. 206(c) (“Published Opinions Binding. Reported panel opinions are binding on subsequent panels.”); 8TH CIR. R. 28A(i) (“[Unpublished opinions] are not precedent . . . .”); 9TH CIR. R. 36-3(a) (“Unpublished dispositions and orders of this Court are not binding precedent, except when relevant under the doctrine of law of the case, res judicata, and collateral estoppel.”); 10TH CIR. R. 36.3(A) (“Unpublished orders and judgments of this court are not binding precedents, except under the doctrines of law of the case, res judicata, and collateral estoppel.”); 11TH CIR. R. 36-2 (“Unpublished opinions are not considered binding precedent . . . .”); D.C. CIR. R. 28(c)(1)(B) (“All unpub-
for publication or assignment of precedential value, even though they designate some opinions as unpublished and do not treat those opinions as precedential.\textsuperscript{27}

Many of the courts' citation rules address only citation of unpublished decisions to \textit{that court},\textsuperscript{28} or do not specify the reach of their rule.\textsuperscript{29} Some courts also address citation of their unpublished decisions to \textit{other courts},\textsuperscript{30} and/or \textit{by other courts}.\textsuperscript{31} A few courts

\textsuperscript{27} \textit{See} 8th Cir. \textit{Internal Operating P.} IV.B; 10th Cir. R. 36.1; 10th Cir. R. 36.2.

\textsuperscript{28} E.g., 1st Cir. R. 32.3(a) (“An unpublished opinion of this court may be cited in this court only in the following circumstances . . . .”).

\textsuperscript{29} \textit{See} 3d Cir. R. 28.3 (“Citations to federal decisions that have not been formally reported shall identify the court, docket number and date, and refer to the electronically transmitted decision.”); 5th Cir. R. 47.5.4 (“An unpublished opinion may . . . be persuasive [and can be cited].”); 8th Cir. R. 28A(i) (“[Unpublished opinions] are not precedent and parties generally should not cite them. When relevant to establishing the doctrines of res judicata, collateral estoppel, or the law of the case, however, the parties may cite any unpublished opinion. Parties may also cite an unpublished opinion of this court if the opinion has persuasive value on a material issue and no published opinion of this or another court would serve as well.”); 10th Cir. R. 36.3(B) (“Citation of an unpublished decision is disfavored. But an unpublished decision may be cited if: (1) it has persuasive value with respect to a material issue that has not been addressed in a published opinion; and (2) it would assist the court in its disposition.”); 11th Cir. R. 36-2 (“Unpublished opinions . . . may be cited as persuasive authority.”); D.C. Cir. R. 28(c)(1)(B) (“All unpublished orders or judgments of this court . . . entered on or after January 1, 2002, may be cited as precedent.”); Fed. Cir. R. 47.6(b) (“Any opinion or order [designated as nonprecedential] must not be employed or cited as precedent. [But t]his rule does not preclude assertion of claim preclusion, issue preclusion, judicial estoppel, law of the case, or the like . . . .”).

\textsuperscript{30} \textit{See} 2d Cir. R. § 0.23 (“Where disposition is by summary order, the court may append a brief written statement to that order. Since these statements do not constitute formal opinions of the court and are unreported or not uniformly available to all parties, they shall not be cited or otherwise used in unrelated cases before \textit{this or any other court.}” (emphasis added); 4th Cir. R. 36(c) (“Citation of this Court’s unpublished dispositions in briefs and oral arguments \textit{in this Court and in the district courts within this Circuit} is disfavored, except for the purpose of establishing res judicata, estoppel, or the law of the case. If counsel believes, nevertheless, that an unpublished disposition of this Court has precedential value in relation to a material issue in a case and that there is no published opinion that would serve as well, such disposition may be cited if counsel serves a copy thereof on all other parties in the case and on the Court.”) (emphasis added); 6th Cir. R. 28(g) (“Citation of unpublished decisions in briefs and oral arguments \textit{in this Court and in the district courts within this Circuit} is disfavored, except for the purpose of establishing res judicata, estoppel, or the law of the case.”) (emphasis added); 7th Cir. R. 53(b)(2) (“Unpublished orders: . . . (iv) Except to support a claim of res judicata, collateral estoppel or law of the case, shall not be cited or used as precedent (A) in any federal court within the circuit in any written document or in oral argument . . . .”). The authority of a court to prohibit citation of one of its decisions to a court in another circuit is unclear.

\textsuperscript{31} \textit{See} 7th Cir. R. 53(b)(2) (“Unpublished orders: . . . (iv) Except to support a claim of res judicata, collateral estoppel or law of the case, shall not be cited or used as prece-
also have rules governing citation of other courts’ unpublished or nonprecedential opinions.\textsuperscript{32} Others have rules or internal operating procedures describing their own citation of unpublished opinions.\textsuperscript{33} Most describe their procedures for determining the publication status of an opinion.\textsuperscript{34}
II. Anastasoff v. United States: A Shot Across the Bow

Much of the current debate over unpublished opinions can be traced to an Eighth Circuit decision issued in August 2000.\(^{35}\) In Anastasoff v. United States, the court considered an appeal by a taxpayer whose claim for a refund of overpaid taxes had been deemed untimely by the district court.\(^{36}\) Ms. Anastasoff argued to the Eighth Circuit panel that it was not bound by one of the court’s prior unpublished decisions, in which it had considered and rejected the “same legal argument” advanced in her case.\(^{37}\) In response, the panel issued a (published) opinion, authored by Judge Richard Arnold, which deemed unconstitutional the portion of the Eighth Circuit’s rule declaring unpublished opinions nonprecedential.\(^{38}\) Focusing on Article III of the Constitution—in particular, its reference to the “judicial power”\(^{39}\)—the panel concluded that the court’s rule “expands the judicial power beyond the limits set by Article III by allowing us complete discretion to determine which judicial decisions will bind us and which will not. Insofar as it limits the precedential effect of our prior decisions, the Rule is therefore unconstitutional.”\(^{40}\) The panel accordingly viewed itself bound by the court’s prior unpublished opinion on the issue, ruling against Ms. Anastasoff by finding her refund claim untimely.\(^{41}\)

While the Eighth Circuit was deciding Anastasoff, the Second Circuit issued an opinion reaching the opposite conclusion about the underlying legal question.\(^{42}\) In light of this conflict, as well as her view that the panel’s constitutional ruling regarding unpublished opinions acting and when so published may be used for any purpose for which an opinion may be used.”); 11th Cir. R. 36-2 (“An opinion shall be unpublished unless a majority of the panel decides to publish it.”); D.C. Cir. R. 36(c)(2) (“[A] panel’s decision to issue an unpublished disposition means that the panel sees no precedential value in that disposition.”); Fed. Cir. Internal Operating P. 10.5 (“The election to [issue a judgment without opinion] shall be unanimous among the judges of the panel.”).

\(^{35}\) See, e.g., Barnett, supra note 5, at 1, 8; Lee & Lehnhof, supra note 5, at 135–38; Robel, supra note 5, at 409–11.

\(^{36}\) 223 F.3d 898, 899 (8th Cir.), vacated as moot on reh’g, 235 F.3d 1054 (8th Cir. 2000) (en banc).

\(^{37}\) Id.

\(^{38}\) Id.; see 8th Cir. R. 28A(i).

\(^{39}\) U.S. Const. art. III, § 1, cl. 1.

\(^{40}\) Anastasoff, 223 F.3d at 905.

\(^{41}\) See id. at 899, 905. One year before Anastasoff, Judge Arnold had posed the same question presented by the case in an article. See Arnold, supra note 5, at 226 (asking, but not examining, whether “the assertion that unpublished opinions are not precedent and cannot be cited [is] a violation of Article III”).

\(^{42}\) See Weisbart v. U.S. Dep’t of Treasury, 222 F.3d 93, 96–97 (2d Cir. 2000).
lished opinions was erroneous, Ms. Anastasoff petitioned the Eighth Circuit for rehearing en banc.\footnote{Anastasoff v. United States, 235 F.3d 1054, 1055 (8th Cir. 2000) (en banc).} The government responded to the petition for rehearing en banc by informing the court that it intended to pay Ms. Anastasoff in full, and contending that the appeal should be dismissed as moot.\footnote{Id.} After learning that Ms. Anastasoff’s claim had been paid, in December 2000 the Eighth Circuit issued an en banc decision finding that the case was moot.\footnote{Id. at 1056.} The court’s decision also explained that, in accordance with “the appropriate and customary treatment” when a case becomes moot, the court would vacate the panel’s opinion and judgment in the case.\footnote{See id.} As for its rule on unpublished opinions, the en banc court observed: “The constitutionality of that portion of [our rule] which says that unpublished opinions have no precedential effect remains an open question in this Circuit.”\footnote{Id. The court continues to leave undecided whether its rule prohibiting the use of unpublished opinions for precedential value is constitutional. See Lederman v. Cragun’s Pine Beach Resort, 247 F.3d 812, 816 n.3 (8th Cir. 2001).}

Although the Eighth Circuit pulled back its ruling on the constitutionality of unpublished opinions, its effects have lingered, and the panel’s decision prompted responses from two other circuit courts. In 2001, the Ninth Circuit ordered an attorney to show cause why he should not be sanctioned for violating the court’s rule barring citation of unpublished opinions.\footnote{Hart v. Massanari, 266 F.3d 1155, 1159 (9th Cir. 2001); see 9th Cir. R. 36-3.} Although the court ultimately exercised its discretion not to sanction the lawyer, it extensively refuted the Anastasoff panel’s conclusion that declaring some decisions nonprecedential violates Article III.\footnote{Hart, 266 F.3d at 1175 (“The question raised by Anastasoff is whether one particular aspect of the binding authority principle—the decision of which rulings of an appellate court are binding—is a matter of judicial policy or constitutional imperative. We believe Anastasoff erred in holding that, as a constitutional matter, courts of appeals may not decide which of their opinions will be deemed binding on themselves and the courts below them.”).}

Writing for the court, Judge Alex Kozinski, one of the most ardent defenders of unpublished opinions, explained:

[W]e are unable to find within Article III of the Constitution a requirement that all case dispositions and orders issued by appellate courts be binding authority. On the con-
trary, we believe that an inherent aspect of our function as Article III judges is managing precedent to develop a coherent body of circuit law to govern litigation in our court and the other courts of this circuit.\footnote{Id. at 1180. The Ninth Circuit also affirmed a district court decision dismissing a lawsuit brought by a practicing attorney who contended that the Ninth Circuit’s rules prohibiting citation to unpublished opinions violated his constitutional rights, on the grounds that the attorney lacked standing. See Schmier v. U.S. Court of Appeals for the Ninth Circuit, 279 F.3d 817, 820–21 (9th Cir. 2002). Because all of the Ninth Circuit judges recused themselves from the case, the panel was comprised of judges from other courts. Id. at 819 n.**. Although the court did not reach the merits of the constitutional claim presented, it did note that “[g]iven the wide range of interest shown in the debate about unpublished opinions . . . , we think it is only a matter of time before the theoretical questions raised by Schmier’s complaint are all properly presented and resolved.” Id. at 825.}

A few months later the Federal Circuit echoed the Ninth Circuit’s view when it rejected a party’s contention that the court was bound by two of its prior nonprecedential opinions under the reasoning of Anastasoff.\footnote{Symbol Techs., Inc. v. Lemelson Med., 277 F.3d 1361, 1367 (Fed. Cir. 2002) (“Article III of the Constitution does not contain an express prohibition on issuing non-precedential opinions, nor can we discern one from the existence of the state of the law when the Framers drafted it.”).}

Although several academic articles published since Anastasoff have analyzed constitutional objections to unpublished opinions,\footnote{See, e.g., Jessie Allen, Just Words? The Effects of No-Citation Rules in Federal Courts of Appeals, 29 VT. L. REV. 555, 574–91 (2005) (no-citation rules violate litigants’ due process rights); Greenwald & Schwarz, supra note 9, at 1161–66 (no-citation rules violate the First Amendment guarantees of free speech and the right to petition); Hoffman, supra note 5, at 347–52 (no-citation rules violate Article III); Katsh & Chachkes, supra note 5, at 315–23 (no-citation rules violate separation of powers because they are not within courts’ Article III powers); Strongman, supra note 5, at 211–22 (no-citation rules violate procedural due process and equal protection under the Fifth Amendment); Marla Brooke Tusk, Note, No-Citations Rules as a Prior Restraint on Attorney Speech, 103 COLUM. L. REV. 1202, 1221–34 (2003) (no-citation rules violate the First Amendment’s rule against prior restraints); Lance A. Wade, Note, Honda Meets Anastasoff: The Procedural Due Process Argument Against Rules Prohibiting Citation to Unpublished Judicial Decisions, 42 B.C. L. REV. 695, 722–31 (2001) (no-citation rules violate procedural due process); see also Amy E. Sloan, A Government of Laws and Not Men: Prohibiting Non-Precendential Opinions by Statute or Procedural Rule, 79 IND. L.J. 711, 745–65 (2004) (arguing that a national procedural rule or federal statute prohibiting the federal appellate courts from prospectively designating selected opinions as nonprecedential would be constitutional).} to date no other court of appeals has weighed in on the constitutionality of unpublished opinions.\footnote{Several judges in the Fifth Circuit did, however, urge that court to “revisit the questionable practice of denying precedential status to unpublished opinions.” See Wil-} Nonetheless, Anastasoff sparked
considerable debate over the wisdom and propriety of courts declaring certain opinions nonprecedential, and about the rules regarding citation to such opinions. 54

Williams v. Dallas Area Rapid Transit, 256 F.3d 260, 260 (5th Cir. 2001) (Smith, J., dissenting from denial of rehearing en banc).

The Supreme Court has been presented with several petitions for certiorari raising questions about the propriety of unpublished opinions and/or citation to them. See, e.g., Petition for Writ of Certiorari at i, Lewin v. Cooke, 537 U.S. 881 (2002) (No. 02-49) (“When an appellant asserts that the circuit court’s unpublished opinion in his case disobeyes this Court’s controlling precedent, and that publication would help to deter the disobedience, is there a due process right to publication?”); Petition for Writ of Certiorari at i, Culp v. Hood, 519 U.S. 1042 (1996) (No. 96-696) (“Whether rules and customs regarding the issuance by United States Courts of Appeal of unpublished decisions are out of date and subject to abuse where computer retrieval of Courts of Appeal decisions has alleviated the problems that the rules and customs regarding unpublished opinions sought to correct.”); see also Motion for Leave to File Petition for Writs of Mandamus and Prohibition at 2, Do-Right Auto Sales v. U.S. Court of Appeals for the Seventh Circuit, 429 U.S. 917 (1976) (No. 75-1404) (“Whether Seventh Circuit Rule 28, which prohibits the publication of written ‘Orders’ which set forth reasons for judgments, and further prohibits a litigant from citing as precedent and relying upon such orders, denies due process of law and violates First Amendment rights.”). The Court has denied these petitions. See Lewin, 537 U.S. 881; Culp, 519 U.S. 1042; Do-Right Auto Sales, 429 U.S. 917; see also Browder v. Director, Dep’t of Corr. of Ill., 434 U.S. 257, 259 n.1 (1978) (“leav[ing for] . . . another day” petitioner’s question about the validity of the Seventh Circuit’s “unpublished opinion” rule). Of course, one cannot infer anything about the Supreme Court’s views on unpublished opinions from its denial of certiorari in these cases. Nevertheless, that a court of appeals disposition was “published” has been cited by the Supreme Court as a factor in determining whether the disposition involved a “case” in the courts of appeals, subject to review under 28 U.S.C. § 1254(1). See Hohn v. United States, 524 U.S. 256, 242 (1998).

54 Although I take no position here regarding the constitutionality of unpublished opinions, it is worth observing that the courts’ practice of sorting out which of their opinions will have precedential effect and which will not is somewhat incongruous with (and tangential to) their constitutional assignment: the resolution of discrete cases and controversies. See U.S. Const. art. III, § 2. When judges decide cases, they consider the present and the past—they consider the case before them and how it fits with the trends in the law and pertinent court decisions. When judges select decisions as nonprecedential they do more—they look forward as well as backward. They predict and shape the course of the law. This is a task distinct from those tasks that judges traditionally perform. I do not mean to adopt a formalistic view of the judges’ responsibilities. Quite clearly, they look forward when deciding cases in the here and now. In writing opinions, they attempt to anticipate how the fortuities unleashed by the passage of time will push and pull their opinions. But there is a qualitative difference between this look forward and the gaze ahead judges must undertake to differentiate between precedential and nonprecedential opinions.
III. Citation Rules and the Origins of Federal Rule of Appellate Procedure 32.1

The Judicial Conference, amongst other things, prescribes rules of practice, procedure and evidence for the federal courts, subject to the ultimate right of Congress to reject, modify, or defer any of the rules.\textsuperscript{55} The Judicial Conference’s rule-prescribing responsibilities are coordinated by its Committee on Rules of Practice and Procedure, commonly referred to as the “Standing Committee.”\textsuperscript{56} The Judicial Conference has five advisory committees to assist the Standing Committee, including one addressing appellate issues—the Judicial Conference Advisory Committee on Appellate Rules, often referred to as the “Appellate Rules Committee.”\textsuperscript{57} The Standing Committee reviews and coordinates the recommendations of the five advisory committees, and proposes rule changes to the Judicial Conference “as may be necessary to maintain consistency and otherwise promote the interest of justice.”\textsuperscript{58} Given its role in establishing federal court rules and in judicial administration, the Judicial Conference, too, has struggled with the issue of unpublished opinions from time to time and found itself a central player in the recent debate.\textsuperscript{59}

The controversy within the Judicial Conference traces back to 1990, when the Federal Courts Study Committee of the Judicial Conference recommended that the Judicial Conference appoint an ad hoc committee to develop uniform guidelines regarding the courts of appeals’ practice of designating certain opinions as unpublished.\textsuperscript{60} The Judicial Conference considered the issue in 1991 but elected to take no action at that time.\textsuperscript{61}

\textsuperscript{55} See 28 U.S.C. §§ 2071–2074 (2000). Amendments to federal rules are adopted by the Supreme Court after approval by the Judicial Conference. See id. §§ 2072–2074. Congress may act upon any rules adopted by the Supreme Court during a period prescribed by statute. See id. § 2074. If Congress does not enact legislation to reject or modify any proposed rules adopted by the Court, the rules take effect. See id.; see also Amchem Prods., Inc. v. Windsor, 521 U.S. 591, 620 (1997) (“Federal Rules take effect after an extensive deliberative process involving many reviewers: a Rules Advisory Committee, public commenters, the Judicial Conference, this Court, the Congress.”) (citation omitted).

\textsuperscript{56} See 28 U.S.C. § 2073(b).

\textsuperscript{57} Id. § 2074(a) (2), (b).

\textsuperscript{58} Id. § 2073(b).

\textsuperscript{59} See infra notes 60–98 and accompanying text.

The topic then rested dormant on the “study agenda” of the Appellate Rules Committee (the “Committee”) for six years until September 1997, when the Committee debated the issue and concluded that it should be “retained on the study agenda with high priority.” At the Committee’s next meeting in April 1998 (the first attended by then-Judge Alito as a Committee member), however, its Chair, Judge Will Garwood of the Fifth Circuit, reported that he had communicated with the chief judges of all of the circuits and that they were “adamant that they did not want national rulemaking on the topic of unpublished decisions.” Because it was clear to him based on those discussions that “rules regarding unpublished decisions have no chance of clearing the Judicial Conference in the foreseeable future,” Judge Garwood recommended, and the Committee unanimously agreed, to remove the issue from the Committee’s study agenda without prejudice to any specific proposals that might be made in the future. Although the Committee backed away from the issue at that meeting, U.S. Deputy Assistant Attorney General Stephen W. Preston, attending on behalf of U.S. Solicitor General Seth Waxman, asked the Committee whether, “notwithstanding the strong reaction of the chief judges, it might still be worthwhile to pursue rulemaking on the isolated question of the citation of unpublished opinions.” He also indicated that the Solicitor General would support a rule allowing the citation of unpublished opinions.

Neither the Solicitor General nor the Committee took any further action on the issue until the Eighth Circuit panel decided Anastasoff v. United States in August 2000. Perhaps prompted by the attention the Eighth Circuit had brought to unpublished opinions, on

---

62 Id. at 12.
64 Id. at 26, 29.
65 Id. at 26.
66 Id.
67 See 223 F.3d 898, 899–905 (8th Cir.), vacated as moot on reh’g, 235 F.3d 1054 (8th Cir. 2000) (en banc); supra notes 36–41 and accompanying text.
January 16, 2001, in the waning days of the Clinton Administration, Solicitor General Waxman, on behalf of the U.S. Department of Justice (the “Justice Department”), proposed that a new rule, 32.1, be added to the Federal Rules of Appellate Procedure that would allow citation to unpublished opinions in all federal courts of appeals.\footnote{68}{Letter from Seth P. Waxman to Judge Will Garwood, supra note 4, at 1.} The Committee took up the issue at its next meeting in April 2001, by which point now-Chief Justice Roberts had joined the Committee as one of its members from private practice.\footnote{69}{See Minutes of the April 11, 2001 Meeting of the Advisory Committee on Rules of Appellate Procedure 64–65 (Apr. 11, 2001), available at http://www.uscourts.gov/rules/Minutes/app0401.pdf.} Although members recalled the chief judges’ vehement resistance to the idea when Judge Garwood polled them in 1998, the Committee briefly discussed whether attitudes may have changed in the intervening three years given the turnover of chief judges and the controversy brewing over Anastasoff.\footnote{70}{Id.} By consensus, the Committee agreed to postpone further discussion of the proposed Rule for another meeting.\footnote{71}{Id. at 65.}

By the time the Committee next met in April 2002, then-Judge Alito had become the Committee’s new Chair.\footnote{72}{See Minutes of the April 22, 2002 Meeting of the Advisory Committee on Rules of Appellate Procedure 1 (Apr. 22, 2002), available at http://www.uscourts.gov/rules/Minutes/app0402.pdf.} At the meeting, he reported that he had surveyed the chief judges about the Justice Department’s proposal that the Federal Rules of Appellate Procedure be amended to permit citation to unpublished opinions, and that the response was mixed, with three expressing support, five opposing, and the others expressing varied views or not conveying a view.\footnote{73}{See id. at 23. The Minutes of the meeting specify each chief judge’s response. See id.} The Committee debated at length whether it should propose a national rule governing the citation of unpublished opinions, ultimately voting 6–3 in support of the idea and 6–3 in favor of the Justice Department’s proposed Rule 32.1, with minor revisions.\footnote{74}{See id. at 24–27. One of the revisions was to change any references to “unpublished” decisions to “non-precedential,” as one committee member noted that the word “unpublished” had become a “misnomer” and observed that many “unpublished” opinions were being published in the Federal Appendix. Id. at 26–27.}
The Committee next met in November 2002. At that session, the members discussed three alternative versions of proposed Rule 32.1, all of which had been drafted by the Committee’s Reporter, Professor Patrick Schiltz (now a federal district court judge in Minnesota). “Alternative A”—the broadest version of the proposed Rule—both authorized courts to issue nonprecedential opinions and allowed unqualified citation to those opinions. “Alternative B” was silent about whether courts can or should issue nonprecedential opinions, but permitted unlimited citation to any such opinions. “Alternative C” was the narrowest version, merely authorizing citation to nonprecedential opinions in limited circumstances.

The Committee quickly rejected Alternative A by consensus, with members expressing concern about adopting a procedural rule that appeared to take a side in the debate over the constitutionality of nonprecedential opinions. Members of the Committee were unanimous in wanting to limit any new rule to the issue of citation. After extended discussion, the Committee approved Alternative B

---

75 See Minutes of the November 18, 2002 Meeting of the Advisory Committee on Rules of Appellate Procedure 1 (Nov. 18, 2002) [hereinafter Minutes of the November 2002 Meeting of the Appellate Rules Advisory Committee], available at http://www.uscourts.gov/rules/Minutes/app1102.pdf. In the midst of this process, in June 2002 the Subcommittee on Courts, the Internet, and Intellectual Property of the U.S. House of Representatives Committee on the Judiciary conducted hearings on the topic of “unpublished judicial opinions.” See Unpublished Judicial Opinions: Hearing Before the Subcomm. on Courts, the Internet, and Intellectual Property of the H. Comm. on the Judiciary, 107th Cong. 1 (2002) [hereinafter Hearing]. The Subcommittee heard testimony from several witnesses: Judge (now Justice) Alito, then a member of the Third Circuit and Chair of the Advisory Committee on Appellate Rules, id. at 5–9, 55–65; Judge Alex Kozinski of the Ninth Circuit, id. at 9–20, 55–65; Kenneth Schmier, Chairman of a group called Committee for the Rule of Law, and brother of Michael Schmier, who challenged the Ninth Circuit’s rules against citation to unpublished opinions on constitutional grounds, see id. at 20–41, 55–65; supra note 50; and Professor Arthur Hellman of the University of Pittsburgh School of Law, Hearing, supra, at 42–55, 55–65. In 1996, the Subcommittee on Administrative Oversight and the Courts of the U.S. Senate Committee on the Judiciary wrote a letter to federal appeals court and district court judges soliciting their views about unpublished opinions. See generally Tatel, supra note 5 (noting the Subcommittee’s letter and stating his views in response).

76 Minutes of the November 2002 Meeting of the Appellate Rules Advisory Committee, supra note 75, at 22–35.

77 See id. at 23.

78 See id. at 28.

79 See id. at 31–32.

80 See id. at 35.

81 See Minutes of the November 2002 Meeting of the Appellate Rules Advisory Committee, supra note 75, at 35.

82 During the discussion, Douglas Letter, representing U.S. Solicitor General Ted Olson, said that although the Justice Department had originally asked the Committee to approve a citation rule and continued to favor one, based on discussions with Judge Alex
with a number of relatively minor changes by a vote of 7–1, with one abstention, and asked the Reporter to revise the draft Rule for the next meeting.\[^{83}\]

At that meeting in May 2003, the Committee (on which now-Chief Justice Roberts remained a member after his May 8, 2003 confirmation by the Senate as a judge on the U.S. Court of Appeals for the D.C. Circuit) reviewed the draft Rule 32.1 and approved a slightly modified version by a 7–1 vote, with one abstention.\[^{84}\] The approved version of the Rule provided:

Rule 32.1. Citation of Judicial Dispositions

(a) Citation Permitted. No prohibition or restriction may be imposed upon the citation of judicial opinions, orders, judgments, or other written dispositions that have been designated as “unpublished,” “not for publication,” “non-precedential,” “not precedent,” or the like, unless that prohibition or restriction is generally imposed upon the citation of all sources.\[^{85}\]

After Judge Alito, as Chair, reported the proposed Rule to the Standing Committee,\[^{86}\] the Judicial Conference solicited public comment on the proposed Rule.\[^{87}\] It received more than 500 sub-

\[^{83}\] See id. at 35–39. One revision the Committee made was to change the title of the proposal from “Citation of Non-Precedential Opinions” to “Citation of Opinions Designated As Non-Precedential”—wanting to stay as clear as possible from implying a view about the jurisprudential impact of “nonprecedential” opinions. Id. at 39.


\[^{85}\] Id. at 11, 17. Subsection (b) of the proposed Rule required that copies of written dispositions “not available in a publicly accessible electronic database” be filed with any brief or other paper in which it is cited. Id. at 11.

\[^{86}\] See Memorandum from Judge Samuel A. Alito, Jr., Chair, Advisory Committee on Appellate Rules, to Judge Anthony J. Scirica, Chair, Standing Committee on Rules of Practice and Procedure 27–36 (May 22, 2003), available at http://www.uscourts.gov/rules/app0803.pdf. By the time Judge Alito reported Rule 32.1 to the Standing Committee, the end of subsection (a) had been amended slightly to read “unless that prohibition is generally imposed upon the citation of all judicial opinions, orders, judgments, or other written dispositions.” Id. at 29.

missions in late 2003 and early 2004 (the second highest number of submissions in the history of federal rulemaking and an unprecedented number for a proposed appellate rule), including dozens from federal appeals court judges, individually or in groups.\textsuperscript{88} The Judicial Conference also conducted public hearings on April 13, 2004, during which fifteen witnesses testified.\textsuperscript{89} The following day, the Appellate Rules Committee voted in favor of proposed Rule 32.1.\textsuperscript{90}

The Standing Committee considered proposed Rule 32.1 at its June 2004 meeting.\textsuperscript{91} After Judge Alito reported on the proposed rule—noting that it would “merely prevent a court of appeals from prohibiting the citation of unpublished opinions” but would not “require a court to give unpublished opinions any weight or precedential value”\textsuperscript{92}—the Standing Committee voted to take no action on the proposed Rule and returned it to the Appellate Rules Committee, with the expectation that the Appellate Rules Committee would work with the FJC to conduct empirical research on unpublished opinions.\textsuperscript{93} At its November 2004 meeting, the Appellate Rules Committee noted the June 2004 decision by the Standing Committee, and announced that the FJC had designed a study and

\textsuperscript{88} See Schiltz, \textit{supra} note 6, at 23–24, 29 (“Proposed Rule 32.1 is, without question, one of the most controversial proposals in the history of federal rulemaking.”)


\textsuperscript{90} See Minutes of the April 13–14, 2004 Meeting of the Advisory Committee on Rules of Appellate Procedure 2–12 (Apr. 13–14, 2004), \textit{available at} http://www.uscourts.gov/rules/Minutes/app0404.pdf. At the meeting Judge Roberts reported to the Appellate Rules Committee that during his appearance at the Standing Committee’s January 2004 meeting (which he attended as a replacement for Judge Alito), he had “stressed that the rule and accompanying Committee Note were drafted to take no position on the issue of whether it is lawful for a court to refuse to give binding precedential effect to one of its opinions.” \textit{Id.} at 2. It is also reported that Roberts spoke in favor of the Rule at the April 2004 Appellate Rules Committee meeting, but the official minutes do not reflect such a statement, nor do they identify the two Committee members who did not favor the proposed Rule. See \textit{id.} at 2–12; \textit{see also} Tony Mauro, \textit{Court Opinions No Longer Cites Unseen}, \textit{Legal Times}, Sept. 26, 2005, at 8 (quoting Roberts as stating that “[a] lawyer ought to be able to tell a court what it has done”).


\textsuperscript{92} \textit{Id.} at 8.

\textsuperscript{93} \textit{Id.} at 11. The Standing Committee approved that action by a voice vote, without objection. See \textit{id.}. 
was considering proposed Rule 32.1 with the assistance of the Administrative Office of the U.S. Courts.  

By the Appellate Rules Committee’s next meeting in April 2005, the FJC had prepared a preliminary report analyzing citations to unpublished opinions. Apparently convinced that most of the concerns about permitting citation to unpublished opinions were not borne out by the FJC’s investigation, the Committee voted 7–2 to approve proposed Appellate Rule 32.1, again sending it to the Standing Committee for consideration.

At its June 2005 meeting, the Standing Committee unanimously approved proposed Rule 32.1 and forwarded it to the Judicial Conference, with any concerns about the Rule seemingly dispelled by the studies of the FJC and Administrative Office of the U.S. Courts.

Three months later, the Judicial Conference approved the proposed Rule, after amending it to apply only to decisions issued on or after January 1, 2007, and forwarded it to the U.S. Supreme Court in November 2005. The Supreme Court approved the new Rule on April 94 See Minutes of the November 9, 2004 Meeting of the Advisory Committee on Rules of Appellate Procedure 1–2 (Nov. 9, 2004), available at http://www.uscourts.gov/rules/Minutes/app1104.pdf.


96 See Minutes of the April 18, 2005 Meeting of the Advisory Committee on Rules of Appellate Procedure 2–18 (Apr. 18, 2005), available at http://www.uscourts.gov/rules/Minutes/AP04-2005-min.pdf; see also Schiltz, supra note 6, at 64 (observing, as Appellate Rules Committee Reporter, that all of the Committee members “agreed that the studies were well done and that they failed to support the main contentions of Rule 32.1’s opponents”).


The Rule is scheduled to take effect in December 2006.\textsuperscript{99}

IV. The Foundation of the Practice of Designating Opinions as Nonprecedential

A. The Underpinnings of Unpublished Opinions

Although the specific criteria courts of appeals use in deciding which of their opinions to designate nonprecedential vary considerably, the courts’ essential mission is the same: they attempt to sort out cases that do not “break new legal ground” or “alter, modify or clarify existing law.”\textsuperscript{101} The premise that judges can and should make this determination at the moment a ruling is made, and without the benefit of input from others, is seriously flawed.\textsuperscript{102} And even though this premise rests at the foundation of the system of unpublished opinions, it has received remarkably little attention in the recent debates over the issue.\textsuperscript{103}

A report drafted for the American College of Trial Lawyers aptly describes the practice of designating certain opinions as nonprecedential: “We put the important decisions in the ‘A’ pile and the unimportant ones in the ‘B’ pile . . . . Everything you will ever need is over here in the ‘A’ pile. . . . [T]he ‘B’ pile case[s] do[ ] not

\textsuperscript{99} Order, \textit{supra} note 1.

\textsuperscript{100} See \textit{supra} note 55 (discussing process for federal rulemaking). The Court must transmit proposed amendments to Congress by May 1 of the year in which the amendment is to take effect. See 28 U.S.C. § 2074 (2000). If Congress does not reject or modify it, the Rule will become effective on December 1, 2006. \textit{See id.}

\textsuperscript{101} See \textit{supra} notes 24–26 (setting out criteria courts use to designate decisions as nonprecedential).


say anything new.”\textsuperscript{104} The report’s author (a critic of rules limiting citation to unpublished opinions) refers to the idea that cases belonging in the “B” pile can be readily and accurately identified as the “redundancy” principle.\textsuperscript{105} This principle, and the entire system of bifurcating appellate dispositions into precedential and nonprecedential categories at the time they are made, is based on a legal fiction.

First, this fiction has a temporal component. By assigning precedential value to opinions at the time they are issued, the rendering judges are making a \textit{prediction} about their future value.\textsuperscript{106} Like all other prognosticators, judges make mistakes.\textsuperscript{107} But the problem with putting judges in the position of predicting the future value and relevance of their opinions goes beyond that posed by the fallibility of judges. It is simply an impossible task to predict future value and relevance, because the information necessary for accurate forecasting does not yet exist. An opinion’s significance depends in part on future events; there are limits on judges’ ability to foresee those events and all of the ways opinions may be fairly marshaled by litigants.\textsuperscript{108} Although judges are as well-equipped as, or better-


\textsuperscript{105} Id. at 680.

\textsuperscript{106} See DuVivier, \textit{supra} note 103, at 416–17 (“By designating some cases as published and others as unpublished, the judges themselves are attempting to predict which cases will have future significance.”); Howard Slavitt, \textit{Selling the Integrity of the System of Precedent: Selective Publication, Depublication, and Vacatur}, 30 Harv. C.R.-C.L. L. Rev. 109, 125 (1995) (“The critical assumption underlying the efficiency rationale in favor of nonpublication is that judges are able to determine in advance which opinions will be valuable to future litigants.”).

\textsuperscript{107} Several commentators have remarked on judges’ fallibility in deciding whether their decisions are, or should be, precedential. See, e.g., Brian P. Brooks, \textit{Publishing Unpublished Opinions: A Review of the Federal Appendix}, 5 Green Bag 2d 259, 260 (2002) (“A careful read of the \textit{Federal Appendix} reveals just how imperfect is an appellate panel’s ability to predict in advance which decisions are pathbreaking, interesting, or important enough to merit the ‘for publication’ designation.”); Greenwald & Schwarz, \textit{supra} note 9, at 1153 (“There is evidence that judges are not unerring judges of the value of their own work product.”); Hangley, \textit{supra} note 104, at 674 (“The judges and their screening clerks are not and never will be infallible in determining what is or is not a novel holding or a helpful discussion, or what will be one when considered in the context of a legal dispute that hasn’t happened yet, and the functions for which past decisions may or must be cited are infinitely variable and largely unpredictable.”).

\textsuperscript{108} See Philip Nichols, Jr., \textit{Selective Publication of Opinions: One Judge’s View}, 35 Am. U. L. Rev. 909, 921 (1986) (“Very often we find our opinions doing duty as precedents in ways quite other than we expected . . . . Our surprise may be . . . because we have failed to anticipate the problems our opinions will be used to solve.”).
equipped than, anyone else to venture a guess, this undertaking is destined to fail if the objective is anything close to complete accuracy. An opinion’s true precedential value can only be assessed by looking backwards after time has passed and its meaning has been debated and tested.\textsuperscript{109}

Second, there is the fiction that the particular judges rendering a decision are the only meaningful participants in the process of deciding whether it will have future significance. The significance (and even the meaning) of judicial opinions cannot be ascertained by rendering judges alone, particularly not at the moment the decision is issued. Vesting decisions with meaning and significance is, by its nature, a process in which other judges and lawyers (and sometimes others, like the media) participate.\textsuperscript{110} The full meaning of a judicial opinion is, in this sense, social—it is dependent on a type of dialogue among the participants in the broader legal community.\textsuperscript{111}

Given that the system of unpublished opinions is built upon this fiction, it should be no surprise that the process of sorting decisions

\textsuperscript{109} See Cappalli, \textit{supra} note 5, at 773 (“The duty of determining the precedential impact of the decision-with-opinion belonged not to the precedent-setting court but to the precedent-applying court.”); Hangley, \textit{supra} note 104, at 686 (“[A] particular opinion, holding or statement may have more persuasive merit or significance on one day than on another day . . . .”); Martineau, \textit{supra} note 9, at 135 (“Limited publication and citation rules require judges to determine in advance the rule of law that will emerge from a case, and then to determine the effect of their decisions on the development of the law. Because our common law system emphasizes the importance of facts in each case, judges hardly can hope to predict the facts of future disputes. They cannot know today what will be crucial to litigants of tomorrow, even when they follow the standards designed to aid them in this determination.”); Steve Sheppard, \textit{The Unpublished Opinion Opinion: How Richard Arnold’s Anastasoff Opinion Is Saving America’s Courts from Themselves}, 2002 Ark. L. Notes 85, 95 (“The precedential value of an opinion has historically been assessed by the later court, not by the court that decides the case.”).

\textsuperscript{110} See Boggs & Brooks, \textit{supra} note 5, at 25 (“One such truth is that judge-made law is a ‘grown order’ rather than a ‘made order.’ ‘Made orders’—such as statutory schemes—‘originate[] from the design of [their] creator,’ while ‘grown orders’—such as judge-made law—‘arise without a plan . . . [with] orderly features [that] result from equilibrium rather than from someone’s design.’” (quoting Mark F. Grady, \textit{Positive Theories and Grown Order Conceptions of the Law}, 23 Sw. U. L. Rev. 461, 461 (1994))); Robel, \textit{supra} note 5, at 410 (“The courts look . . . to the future value of opinions. However, . . . the future value of what they write is not, culturally speaking, determined by the authoring judge alone; rather, it is determined by consumers.”); see also Jones v. Superintendent, Va. State Farm, 465 F.2d 1091, 1093 (4th Cir. 1972) (“No appellate court can ever be much better than its bar.”).

into the “A” and “B” piles reveals serious flaws.\footnote{See supra note 104 and accompanying text. Some commentators have referred to the process of deciding whether or not a decision will be designated precedential as a “sorting device.” See, e.g., Boggs & Brooks, supra note 5, at 19.} There is considerable evidence that judges are deciding important and contested issues in unpublished opinions, and that they are applying the non-precedential designation more broadly than they should.\footnote{As two commentators have rightly observed, “the consequence of not publishing a useful precedent is much greater than the consequence of publishing a redundant, useless one.” Greenwald & Schwarz, supra note 9, at 1153.}

For instance, one might expect virtually all unpublished opinions to be unanimous dispositions affirming lower court rulings because the selection criteria suggest that opinions designated as unpublished should involve routine application of existing law, about which panel members presumably would agree.\footnote{See supra notes 24–26 and accompanying text.} Yet a significant number of such opinions include dissents,\footnote{See supra, note 18, at 221–24; see also Deborah Jones Merritt & James J. Brudney, \textit{Stalking Secret Law: What Predicts Publication in the United States Courts of Appeals}, 54 \textit{Vand. L. Rev.} 71, 120 (2001) ("Unpublished decisions do not reflect routine applications of existing law with which all judges would agree. If they did, these decisions would not include a noticeable number of reversals, dissents, or concurrences, nor would they show significant associations between case outcome and judicial characteristics. . . . We know that at least some unpublished decisions reach results with which other judges would disagree . . . .")} and/or reverse district court rulings.\footnote{See Hannon, supra note 18, at 227–31, 241–50 (discussing the Supreme Court’s review of unpublished federal appeals court opinions and listing cases in which the Court has, among other things, reviewed unpublished dispositions).} One might also expect unpublished opinions rarely, if ever, to be the subject of Supreme Court review. Yet the Court has elected to review dozens of unpublished appeals court opinions,\footnote{See, e.g., Nat’l Archives & Records Admin. v. Favish, 541 U.S. 157, 175 (2004), rev’g 37 F. App’x 863 (9th Cir. 2002) (unpublished); Dastar Corp. v. Twentieth Century Fox Film Corp., 539 U.S. 23, 38 (2003), rev’g 34 F. App’x 312 (9th Cir. 2002) (unpublished); Holmes Group, Inc. v. Vornado Air Circulation Sys., Inc., 535 U.S. 826, 834 (2002), rev’g 13 F. App’x 961 (Fed. Cir. 2001) (unpublished); Sims v. Apfel, 530 U.S. 103, 112 (2000), rev’g No. 98-60126, 1998 WL 792809 (5th Cir. Nov. 6, 1998) (unpublished) (published subsequent to grant of certiorari at 200 F.3d 229); Kokkonen v. Guardian Life Ins. Co. of Am., 511 U.S. 375, 382 (1994), rev’g No. 92-16628, 1993 WL 164884 (9th Cir. May 18, 1993) (unpublished); Houston v. Lack, 487 U.S. 266, 276 (1988), rev’g No. 86-5198, 1987 WL 36042 (6th Cir. May 22, 1987) (unpublished); Baldwin County Welcome Ctr. v. Brown, 466 U.S. 147, 152 (1984), rev’g No. 82-7033, 698 F.2d 1236 (11th Cir. Jan. 31, 1983) (unpublished); Connick v. Myers, 461 U.S. 138, 154 (1983), rev’g No. 81-3203, 654 F.2d 719 (5th Cir. Jul. 23, 1981) (unpublished); Moore v. Illinois, 434 U.S. 220, 232 (1977), rev’g No. 75-1697, 534 F.2d 331 (7th Cir. Apr. 27, 1976) (unpublished); see also}
volve circuit splits, where at least one of the court of appeals decisions addressing the disputed issue was unpublished.\textsuperscript{119}

Many judges have acknowledged it is difficult to determine accurately at the time of disposition which cases should be assigned precedential value;\textsuperscript{120} others have gone further and acknowledged that the system does not work as it should.\textsuperscript{121} In a speech delivered nearly three decades ago, Justice John Paul Stevens observed that the practice itself “rests on a false premise” in that it “assumes that an author is a reliable judge of the quality and importance of his own work product.”\textsuperscript{122}


\textsuperscript{120} See, e.g., Boggs & Brooks, supra note 5, at 21 (noting that “individual panels of judges are only imperfectly able to predict future events and disputes that will influence the development of the law”); Edward A. Adams, Increased Use of Unpublished Rulings Faulted, N.Y. L.J., Aug. 2, 1994, at 1 (noting the Tenth Circuit’s decision to allow citation to unpublished opinions and quoting then-Chief Judge of the Tenth Circuit, Stephanie K. Seymour, as stating that “limiting unpublished decisions to those with no precedential value is the theory, but it’s not always the practice. Some turn out to have precedential value even when the panel of judges thought they didn’t”); see also fones, 465 F.2d at 1094 (“We concede, of course, that any decision is by definition a precedent . . . .”); Patricia M. Wald, The Rhetoric of Results and the Results of Rhetoric: Judicial Writings, 62 U. Chi. L. Rev. 1371, 1376 (1995) (“In my own D.C. Circuit, the wide gap between the number of published and unpublished opinions written by different judges gives pause.”).

\textsuperscript{121} See, e.g., Arnold, supra note 5, at 224 (“[M]any cases with obvious legal importance are being decided by unpublished opinions.”); Wald, supra note 120, at 1374 (“In a study of the D.C. Circuit’s unpublished decisions several years ago, a bar committee, applying our own written criteria, questioned the decision not to publish in 40 percent of the cases. I would guess the number would be much higher now.”); see also Bruce M. Selya, Publish and Perish: The Fate of the Federal Appeals Judge in the Information Age, 55 Ohio St. L.J. 405, 414 (1994) (“Unfortunately, . . . even in circuits with strict rules on the books, the practical criterion for publication is the amount of time that the judge puts into an opinion.”).

\textsuperscript{122} See Justice John Paul Stevens, Address to the Illinois State Bar Association’s Centennial Dinner: Some Thoughts and Reflections on the Litigation Explosion and How It Has Affected the Courts’ Ability to Cope with the Problem (Jan. 22, 1977), in 65 Ill. B.J. 508, 510 (1977). Justice Stevens has also directly or indirectly criticized unpublished opinions in decisions written while on the Supreme Court. See, e.g., Taylor v. United States, 493 U.S. 906, 907 (1989) (Stevens, J., dissenting from denial of certiorari);
The FJC’s recent study, undertaken at the behest of the Standing Committee, also calls the underpinnings of the practice of designating certain decisions as nonprecedential into question. According to a survey of all federal appeals court judges, almost 85% reported that it has at least sometimes been difficult to reconcile an unpublished opinion of their court with one of their published opinions. Presumably, most appeals court judges would not be surprised to learn that appeals courts have relied on their own unpublished opinions from time to time—a practice that makes no sense if the unpublished decision was not precedent and added nothing new to existing case law.

---

123 See supra notes 93–95 and accompanying text.
124 See FJC, UNPUBLISHED OPINIONS, supra note 95, at 11, 40.
125 See Hangley, supra note 104, at 655 (discussing a Fifth Circuit decision in which the court relied on a prior, unpublished decision of that court and included the prior decision as an attachment to the published ruling); Hannon, supra note 18, at 231–35 & tbl.6 (discussing results of author’s own study of citation to unpublished opinions by federal appeals courts); see also Robel, supra note 5, at 406–07 (discussing Hannon’s study).
126 Some courts entertain requests to convert nonprecedential opinions to precedential ones. See 1ST CIR. R. 36(b)(2)(D) (“Any party or other interested person may apply for good cause shown to the court for publication of an unpublished opinion.”); 4TH CIR. R. 36(b) (“Counsel may move for publication of an unpublished opinion, citing reasons. If such motion is granted, the unpublished opinion will be published without change in result.”); 7TH CIR. R. 53(d)(3) (“Any person may request by motion that a decision by unpublished order be issued as a published opinion. The request should state the reasons why the publication would be consistent with the guidelines for method of disposition set forth in [Seventh Circuit Rule 53(c)(1)].”); 8TH CIR. INTERNAL OPERATING P. IV.B (“Counsel may request, by motion, that an unpublished opinion be published.”); 9TH CIR. R. 36-4 (“Publication of any unpublished disposition may be requested by letter addressed to the Clerk, stating concisely the reasons for publication.”); 11TH CIR. R. 36-3 (“At any time before the mandate has issued, the panel, on its own motion or upon the motion of a party, may by unanimous vote order a previously unpublished opinion to be published.”); 11TH CIR. R. 36-3, I.O.P. 6 (“Parties may request publication of an unpublished opinion by filing a motion to that effect . . . .”); D.C. CIR. R. 36(d) (“Any person may, by motion made within 30 days after judgment, or, if a timely petition for rehearing is made, within 30 days after action thereon, request that an unpublished opinion be published.”); Fed. CIR. R. 47.6(c) (“Within 60 days after any non-precedential opinion or order is issued, any person may request, with accompanying reasons, that the opinion or order be reissued as precedential.”). That courts appear to grant such motions often also raises questions about the rigor with which they make publication decisions. See, e.g., Guan v. Bd. of Immigration Appeals, 345 F.3d 47, 48 n.1 (2d Cir. 2003) (noting decision to grant motion to publish “[b]ecause we are persuaded that this decision may have some precedential value . . . .”); In re Administrative Sub-
The premise that judges can and should determine whether they have made “new law” at the moment a ruling is made, and without the benefit of input from others, also appears to reflect certain strands of legal formalism that were displaced decades ago. Although the legacy of “legal realism” is subject to debate, it is clear that some of its lessons are firmly entrenched in mainstream contemporary legal thought. For example, the realist insights that law is not “discovered” and that judges are required to exercise judgment are widely accepted. Today it would be difficult to find someone who disagrees with Justice Oliver Wendell Holmes’s observation that “the life of the law” has been experience rather than logic. It is also widely accepted that discerning the meaning and holding of cases is often difficult, and that precedents frequently are subject to multiple reasonable interpretations. If nothing else, legal realism contributed significantly to freeing us from the illusion that deciding cases or understanding judicial decisions is easy.

Assessing whether a decision makes “new law” is an enormously challenging undertaking, even assuming (as some steadfast realists would not) that the endeavor makes sense at all. A system that asks judges to resolve this question at the moment they issue a decision, and without the benefit of input from others, rests upon antiquated ideas about how the law develops. It is hard enough for judges to get

---


128 See Dan Simon, A Psychological Model of Judicial Decision Making, 30 Rutgers L.J. 1, 7 (1998) (discussing the “pre-realist characterization of the judicial function as one of merely finding and pronouncing extant law”).

129 Oliver Wendell Holmes, Jr., The Common Law 1 (44th prtg. 1951).

130 See Horwitz, supra note 127, at 199 (“Late-nineteenth-century legal thought has often been called formalistic because of its aspiration to be able to render one right answer to any legal question. In attacking deductive legal reasoning, Legal Realists made more original and lasting contributions to legal thought than in any other area.”); Singer, supra note 127, at 470 (“The legal realists [argued] that lawyers could not use legal rules alone to predict judicial decisions. They gave at least three separate reasons for this claim. First, they argued that legal rules were often vague and therefore ambiguous. . . . Second, the realists argued that judges could not determine, in a nondiscretionary way, the holdings of decided cases. . . . Third, the realists argued that, because of the indeterminacy of abstract concepts and the manipulability of precedent, it was almost always possible to appeal to competing and contradictory rules to decide any interesting contested case.”).
cases “right” without having to take on the additional task of assessing the relationship of their just-completed opinion to the entire body of case law previously decided by their court.

B. Rethinking Unpublished Opinions

Any discussion of the problems associated with the existing system of unpublished opinions must include an assessment of the practical realities about how the judiciary operates. It is impossible to deny the enormous challenges courts of appeals face in keeping up with heavy caseloads, and the considerable time and energy required to generate thoughtful, comprehensive, and clear opinions.\textsuperscript{131} Moreover, it is undeniable that dispositions short of full-blown opinions are useful to judges in meeting those challenges. It would be a mistake, however, to conclude that the only way to achieve the advantages afforded by our current system of unpublished opinions is to retain it as is. Doing away with the ex ante designation of opinions as precedent- and nonprecedential would not require judges to write full-blown opinions in all (or even most) cases or require the highest level of attention and scrutiny by the court. Truly easy cases could be resolved with limited explanations of the court’s ruling and the reasons for it.\textsuperscript{132} Many courts already resolve some of their cases through such vehicles.\textsuperscript{133}

\textsuperscript{131} If courts of appeals continue to designate some of their decisions as nonprecedential at the time of decision, that process must be undertaken with rigor, not as an afterthought. Discerning whether and how a particular decision is precedential is often no easy task, and can be time-consuming. \textit{See generally} Michael Abramowicz & Maxwell Stearns, \textit{Defining Dicta}, 57 Stan. L. Rev. 953 (2005) (developing a model for distinguishing between holdings and dicta); Michael C. Dorf, \textit{Dicta and Article III}, 142 U. Pa. L. Rev. 1997 (1994) (discussing the holding/dictum distinction through the lens of Article III). Many supporters of unpublished opinions seem to ignore the time that should be spent carefully assessing whether a decision should be designated as nonprecedential, and how this might offset some of the efficiencies derived from writing nonprecedential opinions. \textit{See} Cappalli, \textit{supra} note 5, at 768 (“The very choice of treating an appealed case as nonprecedential, if done conscientiously, has to be preceded by thoughtful analysis of the relevant precedents.”).

\textsuperscript{132} Some might contend that this reform merely would lead to the relabeling of heretofore “unpublished opinions”—something like “summary order” or “memorandum.” There can be little doubt that many such opinions would be decided with only a cursory written disposition. But the point is that the necessary efficiencies gained by disposing of some cases with an exhaustive opinion can be retained without casting the vast majority of decisions aside as nonprecedential.

\textsuperscript{133} \textit{See} 1st Cir. R. 36(a) (order or memorandum and order); 2d Cir. R. § 0.23 (summary order); 4th Cir. Internal Operating P. 36.3 (summary opinion); 9th Cir. R. 36-1 (memorandum or order); 10th Cir. R. 36.1 (disposition without opinion); 11th
Dispensing with the practice of classifying some opinions as nonprecedential at the time they are rendered surely would greatly expand the universe of decisions that litigants could bring to a court’s attention as authority. But the nature and quality of those decisions will be unchanged—only the labels applied to them will differ. As they do now, courts would wade through lawyers’ arguments about the relevance and persuasiveness of prior decisional law, and resolve cases in the ways they think appropriate.\textsuperscript{134}

\textbf{Conclusion}

At first blush, it might seem surprising that a rule of appellate procedure became the subject of heated debate within the legal community. But the rules governing publication of and citation to judicial opinions are not only central to the judiciary’s self-identity—they are also critical to lawyers and the public, shaping how litigants’ cases are treated by the courts and how litigants communicate with courts through their counsel.

In many respects, however, the years of effort and attention surrounding Rule 32.1 have been a diversion from the more fundamental issues regarding unpublished opinions. From the moment it was conceived, the Rule was written to be “deliberately narrow”—intended to acknowledge “that the courts of appeals designate some of their decisions as unpublished or non-precedential,” but to take “no position on the ongoing debate concerning the propriety of that practice” and not “purport[ing] to dictate to courts or judges what weight should be given to unpublished decisions.”\textsuperscript{135}

\textsuperscript{134} See Panel Discussion, \textit{The Appellate Judges Speak}, 74 \textit{Fordham L. Rev.} 1, 17 (2005) (First Circuit Chief Judge Michael Boudin stating, “Precedential weight, of course, is not an ‘on’ or ‘off’ switch,” and “opinions get very different weight depending on all kinds of circumstances: who wrote them, how recently, how persuasive they are”); Berman & Cooper, \textit{supra} note 5, at 2039–40 (“If a case genuinely calls only for a straightforward application of well-settled law, then it is likely to receive summary treatment whether or not the opinion is subject to publication. Moreover, since such an opinion by definition does not add or otherwise change the law, there is no real need or reason for either a litigant or a judge to cite to it.”); Hangley, \textit{supra} note 104, at 647 (“[A] court can surely decide what weight it wishes to give to the reasoning behind [a prior] holding in a particular new factual context, rather than making the a priori judgment that nothing in the holding could possibly be pertinent to any future case.”).

\textsuperscript{135} See Letter from Seth P. Waxman to Judge Will Garwood, \textit{supra} note 4, at 5; \textit{see also} Fed. R. App. P. 32.1 (proposed) advisory committee’s note.
The deliberative process leading up to the enactment of Rule 32.1 consciously and conspicuously ignored the pink elephant in the middle of the room.\textsuperscript{136} Now that the issue of citation to unpublished opinions is resolved and behind us, it is time to rethink the system of unpublished opinions itself. Any complete reassessment of that system must include exploration of the legal fiction upon which the current system is based.\textsuperscript{137}

Judges cannot know at the time they render a decision, with anything resembling certainty, whether their opinion has advanced the development of the law or will never again interest anyone but the parties to that case. That judgment is best made with the benefit of time, and with input from lawyers, litigants, and other judges. The legal community as a whole would be better served by a system of decision-rendering reflecting that reality.

\textsuperscript{136} The Reporter of the Appellate Rules Committee recently described unpublished opinions as the “crazy uncle in the attic of the federal judiciary.” Schiltz, \textit{supra} note 6, at 72.

\textsuperscript{137} There is some question whether the rulemaking process is the appropriate forum to resolve the underlying question of whether certain opinions can be deemed non-precedential. I take no position on that issue here.
Abstract: On December 20, 2005, seven northeastern states signed the Regional Greenhouse Gas Initiative ("RGGI"), an agreement aimed at reducing greenhouse gas pollution from power plants. Once enacted by each state’s legislature or rulemaking agencies, this agreement will establish a “cap-and-trade” program to cap greenhouse gas emissions within the region and allow power plants to trade emissions allocations. This program faces a significant challenge, however. Electricity suppliers within the region may import power from outside the regulated region to avoid the constraints of the emissions cap, resulting in little or no net decrease in overall emissions—a problem known as “leakage.” Because limiting emissions imports would inevitably place burdens on the interstate trade of electricity, the regulatory approaches available to RGGI states to limit imports of power from unregulated regions may be subject to attack as violations of the Commerce Clause of the U.S. Constitution. This Note explores the possibility of applying the concepts embodied in the compensatory tax doctrine to defend a regulatory scheme that the RGGI states might employ to combat leakage.

Introduction

On December 20, 2005, seven northeastern states signed an agreement to implement the Regional Greenhouse Gas Initiative ("RGGI") in an effort to reduce greenhouse gas pollution from power plants.¹ This agreement marked the first formal commitment to im-

¹ See Regional Greenhouse Gas Initiative Memorandum of Understanding 12 (Dec. 20, 2005) [hereinafter RGGI MoU], available at http://www.rggi.org/docs/mou_final_12_20_05.pdf. Greenhouse gases are those gases in the earth’s atmosphere that contribute to the “greenhouse effect”; that is, they absorb and reradiate energy from the sun back toward the earth, causing the earth’s surface and lower atmosphere to warm more than they otherwise would. See C.C. Lee, DICTIONARY OF ENVIRONMENTAL LEGAL TERMS 284 (1st ed. 1996); Environmental Protection Agency, Terms of Environment: Glossary, Abbreviations, and Acronyms, http://www.epa.gov/OCEPAterms/gterms.html (last visited Aug. 29, 2006).

Seven states are participating in the RGGI program: Connecticut, Delaware, Maine, New Hampshire, New Jersey, New York, and Vermont. See RGGI MoU, supra, at 1, 6–7. The
plementing a market-based trading program for carbon dioxide ("CO₂") emissions in the United States. In March 2006, California’s Environmental Protection Agency released a report summarizing the emissions trading program options that the state was exploring for regulating greenhouse gas emissions. Both the RGGI agreement and the California report, however, identify a particular concern regarding regional cap-and-trade emissions programs—that of “leakage.”

Leakage—the movement of emissions from regulated to unregulated regions to avoid caps on emissions—can occur when a cap-and-trade scheme is implemented on a state or regional, rather than national, level. Electricity suppliers begin to import power from outside the regulated region to avoid the constraints of the emissions cap, resulting in little or no net decrease in overall emissions associated with the power consumed inside the region. To combat this problem, the RGGI states and California could limit emissions associated with en-
ergy imported into the region. Because limiting interstate imports places burdens on the trade of electricity, however, this approach may be subject to attack under the Interstate Commerce Clause of the U.S. Constitution.

This Note explores the possibility of applying the concepts embodied in the compensatory tax doctrine to defend a regulatory scheme that might be employed to combat leakage, focusing on RGGI as the model scheme. The compensatory tax doctrine stands for the principle that even if a state regulation burdens interstate commerce, it may survive constitutional scrutiny if it is a compensatory tax designed merely to make interstate commerce bear a burden already borne by intrastate commerce. Any regulation the RGGI states adopt to address leakage will necessarily impose burdens on interstate commerce because they will have to limit, either directly or indirectly, electricity imports from out of state. This Note argues that the RGGI states, and any reviewing court, should draw on compensatory tax doctrine principles in crafting, and supporting, a regulation that imposes burdens on imported electricity.

Part I of this Note provides an introduction to the RGGI program and the particular problem of leakage. Part II introduces the potential legal challenges to the program based on the Commerce Clause of the U.S. Constitution, and explores the compensatory tax doctrine as developed by the U.S. Supreme Court. Part III analyzes the application of Commerce Clause jurisprudence and the compensatory tax doctrine to the alternatives that RGGI may use to combat the problem.

---

9 See infra notes 18–272 and accompanying text. The arguments in this Note could also be applied to the regulatory scheme that California is currently adopting. See supra note 3.
11 See infra notes 43–66 and accompanying text.
12 See infra notes 190–195 and accompanying text.
13 See infra notes 18–66 and accompanying text.
14 See infra notes 67–189 and accompanying text.
of leakage and evaluates the likely success of those options. Part III concludes that although the compensatory tax doctrine may not be directly applicable to the regulation of emissions leakage, the legal principles it embodies should be used to uphold the regulation. In addition, if RGGI can overcome the legal and political obstacles in its path, it may serve as an effective experiment in the regulation of CO₂ emissions and eventually could be a template for a national regulatory program aimed at slowing global warming.

I. Climate Change and the Regional Greenhouse Gas Initiative

The states participating in RGGI are taking action because they recognize that climate change poses serious risks to human health and global ecosystems. Climate change is a result of global warming, which in turn is caused by the accumulation of greenhouse gases in the earth’s atmosphere, principally CO₂. Various scientific models indicate that the average global temperature could rise by up to 7.7 degrees Fahrenheit by the middle of this century. This expected temperature increase could greatly exacerbate shortages of food, water, and energy supplies, and increase the number of refugees around the world—not to mention raise national security concerns relating to

---

15 See infra notes 190–272 and accompanying text.
16 See infra notes 190–272 and accompanying text.
17 See infra notes 190–195 and accompanying text.
18 See RGGI MoU, supra note 1, at 1.
19 See Eileen Claussen, An Effective Approach to Climate Change, SCIENCE, Oct. 29, 2004, at 816; Joint Science Academies’ Statement: Global Response to Climate Change 1 (June 7, 2005), http://nationalacademies.org/onpi/06072005.pdf. Carbon dioxide is the most abundant anthropogenic greenhouse gas in the world. Env’t Ne. Overview, supra note 2, at 1. It is released into the atmosphere when carbon-based fuel is burned. See id. In 1780, the level of CO₂ in the earth’s atmosphere was approximately 280 parts per million (“ppm”) and had been for at least 6000 years. See Elizabeth Kolbert, The Climate of Man III: What Can Be Done?, THE NEW YORKER, May 9, 2005, at 54. As the industrial age took hold, CO₂ concentrations began to rise—slowly at first and then more rapidly. See id. By the 1970s, the CO₂ concentration in the atmosphere was approximately 330 ppm and in 2000 it reached 369 ppm. Id.; see also Lester R. Brown, Growing . . . Growing . . . Gone?, MOTHER EARTH NEWS, Dec.–Jan. 2004, at 70, available at http://www.motherearthnews.com/Nature_and_Environment/2003_December_January/Growing___Growing___Gone_.
20 See Elizabeth Kolbert, The Climate of Man II: The Curse of Akkad, THE NEW YORKER, May 2, 2005, at 69. This predicted increase in temperature is based on model predictions that show that if we continue to produce greenhouse gases at the rates necessary to meet increasing demand, atmospheric CO₂ will reach 500 ppm around the middle of this century. Id. There is evidence that CO₂ concentrations in the earth’s atmosphere were that high about fifty million years ago when crocodiles lived in Colorado and ocean levels were three hundred feet higher than they are today, putting much of today’s inhabited land underwater. Kolbert, supra note 19, at 54.
nuclear proliferation, terrorism and the potential for war.\textsuperscript{21} For this reason, many countries, as well as state and local governments and private economic entities in the United States, are taking action to limit greenhouse gas emissions.\textsuperscript{22}

A. The Regional Greenhouse Gas Initiative: Regulatory Structure

The purpose of the RGGI program is to regulate CO\textsubscript{2} emissions using a market-based approach, commonly referred to as cap-and-trade.\textsuperscript{23} The cap-and-trade approach allows facilities subject to the regulation to achieve emission reduction targets, and thus avoid potential penalties, in an economically efficient manner.\textsuperscript{24} Under the

\textsuperscript{21} See generally Peter Schwartz & Doug Randall, An Abrupt Climate Change Scenario and Its Implications for United States National Security (2003), available at http://www.environmentaldefense.org/documents/3566_AbruptClimateChange.pdf. This report, commissioned by the U.S. Department of Defense, identified these and other consequences as the possible and even likely results of an abrupt climate change event that could be caused by the collapse of the Atlantic conveyor as a result of global warming. See id. at 1–3; see also RGGI MoU, supra note 1, at 1.


\textsuperscript{23} See RGGI MoU, supra note 1, at 2.

\textsuperscript{24} See Env’t NE. OVERVIEW, supra note 2, at 2. To ensure compliance with the cap, the RGGI states will develop a method of enforcement that may be imposed on the regulated
standard cap-and-trade model, the government sets an aggregate cap on the amount of allowable emissions in the region.\textsuperscript{25} The cap is then distributed either through allocation or sale to each emitting facility in the form of allowances, where one unit (usually a ton) of pollutant equals one allowance.\textsuperscript{26} Each facility must own the same number of allowances as the number of tons of pollutant it emits.\textsuperscript{27} The current proposal is to implement the RGGI cap in two phases.\textsuperscript{28} Between the years 2009 and 2015, the RGGI states will cap CO\textsubscript{2} emissions at approximately 120 million tons, which is approximately equivalent to the average emissions of the highest three years between 2000 and 2004.\textsuperscript{29} In the second phase the cap will be reduced by 10\% from 2015 through 2020.\textsuperscript{30}

The cap-and-trade approach creates a market for the allowances when a cleaner power facility has more allowances than it needs to cover its emissions.\textsuperscript{31} The cleaner facility can then sell its surplus allowances to dirtier facilities that do not have enough allowances to cover their emissions.\textsuperscript{32} If the demand for allowances increases, the market price for allowances also will increase.\textsuperscript{33} Dirtier facilities then face the choice of either reducing emissions or purchasing allowances facilities. See \textit{Model Rule} § XX-6.5 (Reg’l Greenhouse Gas Initiative 2006), http://www.rggi.org/docs/model_rule_8_15_06.pdf (setting out a model compliance scheme).


\textsuperscript{26} See \textit{Env’t Ne. Overview, supra} note 2, at 2; \textit{EPA Guide, supra} note 25, at 1-2.

\textsuperscript{27} See \textit{Env’t Ne. Overview, supra} note 2, at 2; \textit{EPA Guide, supra} note 25, at 1-2.

\textsuperscript{28} Memorandum from the RGGI Staff Working Group to RGGI Agency Heads 2 (Aug. 24, 2005) [hereinafter \textit{RGGI Staff Memorandum}], http://www.rggi.org/docs/rggi_proposal_8_24_05.pdf.

\textsuperscript{29} See \textit{id.; see also RGGI MoU, supra} note 1, at 2. If Massachusetts and Rhode Island join the program, the cap will be increased to approximately 150 million tons. See \textit{RGGI MoU, supra} note 1, at 8.

\textsuperscript{30} \textit{RGGI Staff Memorandum, supra} note 28, at 2.

\textsuperscript{31} See \textit{EPA Guide, supra} note 25, at 1-2 to -3.

\textsuperscript{32} See \textit{id.} Cap-and-trade is a workable solution in the case of CO\textsubscript{2} because CO\textsubscript{2} is a uniform pollutant; it has the same atmospheric impact regardless of where the source is located. See \textit{Env’t Ne. Overview, supra} note 2, at 2. By contrast, localized pollutants such as mercury and particulate matter directly impact the health of the local communities and ecosystems surrounding the emission source. \textit{Id.} This localized impact raises concerns over the creation of “hotspots” of pollution and related social and environmental justice issues. See \textit{EPA Guide, supra} note 25, at 2-2. Therefore, for localized pollutants it is usually necessary to implement site-specific command-and-control regulation, which does not allow the flexibility of cap-and-trade programs. See \textit{id.}

\textsuperscript{33} See \textit{EPA Guide, supra} note 25, at 1-2 to -4.
because the net emissions from the region cannot exceed the cap.\textsuperscript{34} This approach gives facilities flexibility not available to them under traditional command and control regulations; each facility can design its own compliance strategy based on economic efficiency and adjust its strategy over time in response to changes in technology and the market.\textsuperscript{35} In fact, the federal government used a cap-and-trade program to regulate the emissions that cause acid rain largely because of the flexibility the approach offers.\textsuperscript{36}

Because the federal government has not implemented a national regulatory program for CO\textsubscript{2} emissions, the northeastern states, through RGGI, may act without fear of preemption by existing federal law.\textsuperscript{37} This does not mean, however, that RGGI lacks legal obstacles.\textsuperscript{38} For example, although each state has the individual authority to regulate CO\textsubscript{2} emissions, each must determine how to fit that regulation within its state regulatory framework.\textsuperscript{39} Some states can adopt the RGGI regulations through the rulemaking authority vested in their

\textsuperscript{34} See id. at 1-2 to -3. Dirtier facilities that exceed the emissions cap may be subject to additional penalties. See supra note 24.

\textsuperscript{35} See EPA GUIDE, supra note 25, at 1-2. In addition to investing in non-emitting forms of energy generation such as wind and solar energy, regulated facilities will have an incentive to improve end-use efficiency, transition to cleaner fossil fuels, invest in more efficient generation and transmission technology, and even utilize carbon capture and sequestration techniques to offset their emissions if it is economically efficient to do so. See id. at 1-3; see also CALIFORNIA REPORT, supra note 3, at 24.

\textsuperscript{36} See 42 U.S.C. § 7651(b) (2000) (aiming to reduce emissions of sulfur dioxide and nitrogen oxide through an emissions allocation and transfer system); see also CALIFORNIA REPORT, supra note 3, at 11–12 (noting the success of the federal government’s acid rain reduction program). The largest-scale use of the cap-and-trade model in the United States to date is the federal government’s acid rain program under Title IV of the Clean Air Act, but the model has also been used in regional programs. See 42 U.S.C. § 7651 (2000); CALIFORNIA REPORT, supra note 3, at 11–15. Two regional programs, the Northeast NOx Budget Program and the Regional Clean Air Incentives Program (“RECLAIM”), used a cap-and-trade emissions program to regulate ozone and smog, respectively, with varying degrees of scope and success. See CALIFORNIA REPORT, supra note 3, at 13–14.

\textsuperscript{37} See ENV’T NE. OVERVIEW, supra note 2, at 3. This statement assumes that RGGI does not attempt to regulate electricity transmission or wholesale transactions per se that are regulated by the Federal Energy Regulatory Commission under the Federal Power Act. See 16 U.S.C. § 824 (2000); see also Note, Foreign Affairs Preemption and State Regulation of Greenhouse Gas Emissions, 119 HArv. L. REV. 1877, 1878 (2006) (arguing that state regulation of greenhouse gases should not be preempted by the federal foreign affairs power).

\textsuperscript{38} See Peter Glaser, Troutman Sanders LLP, Regional Greenhouse Gas Initiative: A Contrarian Perspective, Presentation to the American Bar Association’s Environment, Energy and Resources Section 23 (Jan. 26, 2006), http://www.abanet.org/environ/committees/renewableenergy/teleconarchives/012606/1-26-06GlaserPPT.PPT (contending that the RGGI states face Compact Clause and Commerce Clause hurdles and questioning whether the states have the political will to regulate CO\textsubscript{2} if legislation is required).

\textsuperscript{39} See RGGI MoU, supra note 1, at 7.
respective state agencies, while others require enabling legislation to give effect to the RGGI rules. Once the cap-and-trade regulations have been adopted, each RGGI state will monitor and enforce those rules. That enforcement may raise additional challenges—in particular, the problem of leakage.

B. The Problem of Leakage and Regulations That Burden Interstate Commerce

Implementing a cap-and-trade program at the regional level presents problems that do not arise when a similar program is implemented at the national level. The RGGI cap-and-trade program will operate on the supply side—that is, CO₂ emission allowances will be allocated to, and traded among, fossil fuel-fired electricity generators within the region that supply electricity to the grid. Because the emissions cap will apply only to in-region generators, the RGGI plan will not limit emissions from electricity that is imported into the region and used by consumers within RGGI states. Generators outside the capped region will be able to export power freely to the entities inside the region that are responsible for procuring and delivering electric power to consumers without concern for the cap—resulting in emissions leakage. Entities that deliver electricity to the consumer are referred to as load-serving entities (“LSEs”).

---

40 See id.
41 See id.
42 See id. at 9.
43 See Policy Options, supra note 8, at 4–5, 8 (discussing problems of regulatory programs that do not cap emissions outside the regulating state, including the problem of leakage); Robert R. Nordhaus & Stephen C. Fotis, Pew Ctr. for Global Climate Change, Analysis of Early Action Crediting Proposals 31 (1998), http://www.pewclimate.org/document.cfm?documentID=237 (identifying displacement of emissions from sources within the program to sources outside the program as a potential problem facing non-national programs).
44 See RGGI MoU, supra note 1, at 2; Regional Greenhouse Gas Initiative, About RGGI, http://www.rggi.org/about.htm (last visited Aug. 23, 2006); supra notes 23–27 and accompanying text.
45 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.
46 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5. The program cap will cover all in-region fossil fuel-fired electricity-generating units having a rated capacity equal to or greater than twenty-five megawatts. See RGGI MoU, supra note 1, at 2. The program may be expanded in the future to include other sources of greenhouse gas emissions and greenhouse gases other than CO₂. See Regional Greenhouse Gas Initiative, About RGGI, http://www.rggi.org/about.htm (last visited Aug. 23, 2006).
47 See California Report, supra note 3, at 21; Cowart, supra note 7, at 2.
The leakage problem presents two related problems for regulators in the RGGI region. First, generators outside the RGGI region will have a competitive advantage over generators within the region because they will have little incentive to invest in the cleaner technologies required to achieve the emission cap. As a result, electricity outside the region will become less expensive than electricity produced inside the region. This leads to the second problem. The resulting increase in cheaper, imported electricity will undermine the goal of the program because imported emissions will not count towards the region’s emission limits even though they are directly associated with the region’s electricity consumption.

As expressed in their Memorandum of Understanding, the RGGI states already are committed to a supply side cap-and-trade program. During the initial phase of the program, the RGGI states have decided not to take direct regulatory action to stem leakage, but have agreed to implement measures to monitor electricity imports and reevaluate whether action is required at a later date. In the meantime, the RGGI states will establish a working group to consider potential options for addressing leakage.

Various options are available to address the problem of leakage. One option is to supplement the initial domestic cap-and-trade program (imposed on in-state electricity generators as a source of emissions) with a second, load-side cap-and-trade regulation imposed on

---

48 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.
49 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.
50 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.
51 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.
52 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.
53 See RGGI MoU, supra note 1, at 2.
54 See id. at 9.
55 Id.
56 See California Report, supra note 3, at 22–24; Cowart, supra note 7, at 5. For example, California is exploring a variation of the cap-and-trade approach referred to as allocation-to-load. California Report, supra note 3, at 21. Under this approach emission allowances are allocated to electricity providers, or LSEs, rather than to electricity generators. Id. Each LSE must hold allowances equal to the emissions created by the electricity it distributes to consumers. Id. Under a complete allocation-to-load program, the regulated LSE must hold allowances for all emissions associated with the electricity it sells to consumers, regardless of where the producing generator is located, and the cap applies to total emissions associated with all electricity consumed in the state. Id. In this way, imported energy, as well as domestic energy, is accounted for in the cap. See California Report, supra note 3, at 21–23; Cowart, supra note 7, at 5.
LSEs, but only on the electricity they import into the region.\footnote{Cf. California Report, supra note 3, at 21–23 (discussing the possibility of regulating CO\textsubscript{2} emissions through LSEs, rather than generators). This Note does not address the question of whether individual states have the authority to regulate the emissions associated with imported power purchased by regulated LSEs. See id. (discussing existing regulatory authority and additional authority that would require legislative action to regulate all LSEs in California). This question depends heavily on state-specific legal issues. See id.; see also RGGI MoU, supra note 1, at 7.} This load-side regulation would treat electricity imports as an additional source of emissions included in the CO\textsubscript{2} emissions cap for the region.\footnote{See California Report, supra note 3, at 21–23; Cowart, supra note 7, at 5.} LSEs would initially be allocated CO\textsubscript{2} allowances on the same basis as that of in-region generators—LSEs would receive allocations based on historic imports just as generators receive allocations to cover their historic generation.\footnote{See California Report, supra note 3, at 21–23; Cowart, supra note 7, at 5. The current proposed cap for the total RGGI region—approximately 120 million short tons—is based on the average emissions of the highest three years between 2000 and 2004 for each state. See RGGI MoU, supra note 1, at 2; RGGI Staff Memorandum, supra note 28, at 2. Generators and LSEs alike would receive allowances on this same historic basis. See California Report, supra note 3, at 21–23; Cowart, supra note 7, at 5.} In-state generators would be legally responsible for their own emissions under the first regulation.\footnote{See California Report, supra note 3, at 21–23; Cowart, supra note 7, at 5.} Under the second regulation, LSEs would be legally responsible for the emissions associated with the electricity they import from states outside the RGGI region and distribute to in-state consumers.\footnote{See California Report, supra note 3, at 21. The assignment of CO\textsubscript{2} attributes to imported electricity for the purposes of measuring emissions associated with consumption in the state is a complicated issue that is not addressed in this Note. See id. at 23. Several methods are available. See id. Before choosing a method, the regulating community must consider the impact that each method could have on the legal analysis presented herein. See id.} The total cap on CO\textsubscript{2} emissions would therefore cover those emissions produced in the region by electricity generators, as well as those emissions produced outside the region that are directly associated with consumer demand for electricity inside the region.\footnote{See id. at 21–23.} The LSEs would engage in the same market for allowances with electricity generators and make operating choices based on economic efficiency.\footnote{See id.}

This regulatory scheme would likely face challenges based on the Interstate Commerce Clause of the U.S. Constitution, however, because the regulation imposed on LSEs would place restrictions only
on imported electricity. Electricity generators outside the RGGI region wishing to sell into the region at lower prices likely will challenge the regulation as a violation of the dormant Commerce Clause, which prohibits any state from enacting regulations that discriminate against (or place burdens on) interstate commerce. The RGGI states should thoroughly consider potential Commerce Clause challenges before implementing a cap-and-trade program on electricity imported by LSEs to stem leakage.

II. Commerce Clause Jurisprudence and the Development of the Compensatory Tax Doctrine

The Commerce Clause of the U.S. Constitution provides that “the Congress shall have Power . . . to regulate commerce . . . among the several States.” Although phrased as an affirmative grant of power to Congress, the Commerce Clause has long been recognized to have a negative aspect which denies states the power to discriminate against, or burden, interstate commerce. A variety of reasons are given for this negative aspect of the Commerce Clause (called the “dormant Commerce Clause”); two are of particular interest here. First, it prohibits economic protectionism by the states—that is, regulatory measures designed to benefit in-state economic interests by burdening out-of-state competitors. Second, it promotes economic

64 See U.S. Const. art. I, § 8, cl. 3; Or. Waste Sys., Inc. v. Dep’t of Envtl. Quality, 511 U.S. 93, 99 (1994) (holding that state laws placing burdens on interstate commerce are subject to challenge based on the Commerce Clause of the U.S. Constitution); Robert B. McKinstry, Jr., Laboratories for Local Solutions for Global Problems: State, Local and Private Leadership in Developing Strategies to Mitigate the Causes and Effects of Climate Change, 12 Penn St. Envtl. L. Rev. 15, 67 (2004) (noting potential Commerce Clause challenges to state and regional regulatory programs).

65 See U.S. Const. art. I, § 8, cl. 3; Or. Waste, 511 U.S. at 99; McKinstry, supra note 64, at 67.

66 See infra notes 190–272 and accompanying text; see also Kirsten H. Engel, The Dormant Commerce Clause Threat to Market-Based Environmental Regulation: The Case of Electricity Deregulation, 26 Ecology L.Q. 243, 250–52 (1999) (noting the Commerce Clause objections to market-based environmental regulation and arguing that such regulation should be upheld based on the logic of the market participant exception and because it promotes economic efficiency and interstate harmony, and is not motivated by economic protectionism).

67 U.S. Const. art. I, § 8, cl. 3.


69 See Fulton, 516 U.S. at 330; Or. Waste, 511 U.S. at 98–99.

70 Fulton, 516 U.S. at 330 (quoting Assoc. Indus. of Mo. v. Lohman, 511 U.S. 641, 647 (1994)).
efficiency that would be undone if states were free to place burdens on the flow of commerce across their borders. The Supreme Court has stated that in granting Congress authority over interstate commerce, the Framers sought “to avoid tendencies toward the economic Balkanization that had plagued relations among the colonies and later among the states under the Articles of Confederation.”

The first step in evaluating the constitutionality of a state law under dormant Commerce Clause jurisprudence is to determine whether the challenged statute regulates evenhandedly with only “incidental” effects on interstate commerce, or discriminates against interstate commerce either on its face or in practical effect. Where the regulation is “evenhanded” and the effects are “incidental,” the statute will be upheld if the regulation passes the Pike v. Bruce Church, Inc. balancing test. This test examines whether: (1) the law effectuates a legitimate local purpose, (2) the burden imposed on interstate commerce is not clearly excessive in relation to the putative local benefits, and (3) there are alternative means for promoting the local purpose as well without discriminating against interstate commerce. Where, however, the state regulation is discriminatory, meaning that it provides differential treatment of in-state and out-of-state interests, it is virtually per se invalid. The regulation’s proponent will only overcome the per se rule of invalidity if it can show that the regulation advances a legitimate local purpose that cannot be adequately served by reasonable nondiscriminatory alternatives. Facial discrimination by itself may be a fatal defect, and invokes the strictest scrutiny of any purported legitimate local purpose and of the absence of nondiscriminatory alternatives.

71 See Or. Waste, 511 U.S. at 98–99.
73 See Fulton, 516 U.S. at 331; Or. Waste, 511 U.S. at 99; Hughes, 441 U.S. at 336.
75 See id.
76 See Fulton, 516 U.S. at 331; Or. Waste, 511 U.S. at 99.
77 See Or. Waste, 511 U.S. at 100–01 (quoting New Energy Co. of Ind. v. Limbach, 486 U.S. 269, 278 (1988)); Hughes, 441 U.S. at 336; see also Maine v. Taylor, 477 U.S. 131, 151–52 (1986) (upholding a facially discriminatory law banning the importation of out-of-state bait fish into Maine because the fish were subject to parasites completely foreign to Maine baitfish and could jeopardize the health of the Maine fish population, and no nondiscriminatory alternatives existed).
78 See Or. Waste, 511 U.S. at 99; Hughes, 441 U.S. at 337; see also Justin M. Nesbit, Note, Commerce Clause Implications of Massachusetts’ Attempt to Limit the Importation of “Dirty” Power in the Looming Competitive Retail Market for Electricity Generation, 38 B.C. L. Rev. 811, 842 (1997) (concluding that an outright ban on imported power would likely be invalidated under the
The hybrid regulatory approach described above could be challenged as a facially discriminatory regulation because the regulation covering LSEs only regulates emissions associated with imported electricity and therefore expressly treats in-state and out-of-state interests differently. The LSE regulation imposes burdens on electricity crossing state lines only. Therefore it burdens out-of-state generators wishing to sell into the RGGI region, but does not itself impose the same burdens on in-state generators. Thus, the regulation would most likely be subject to strict scrutiny under the dormant Commerce Clause.

### A. An Exception to the Dormant Commerce Clause Rule of Invalidity: The Compensatory Tax Doctrine

The U.S. Supreme Court has recognized a narrow exception to the per se rule of invalidity for facially discriminatory regulations, in the form of the compensatory tax doctrine. Under the compensatory tax doctrine, a facially discriminatory regulation may survive strict scrutiny if it is a compensatory tax designed merely to make interstate commerce bear a burden already borne by intrastate commerce. Although often expressed as an independent doctrine unto itself, the compensatory tax doctrine is merely a specific way of justifying a facially discriminatory tax because it achieves a legitimate local purpose that cannot be achieved through nondiscriminatory means.

The Supreme Court laid the groundwork for the compensatory tax doctrine in the 1869 case of *Hinson v. Lott*. In *Hinson*, the state of Alabama imposed a tax on all liquor imported into the state equal to the tax imposed on all liquor distilled within the state. The Supreme Court stated that the tax on imported liquor was merely a complementary provision necessary to make the tax equal on all liquors sold

---

Commerce Clause but that a surcharge on sales of “dirty” electricity could pass the *Pike* balancing test).

79 *See Fulton*, 516 U.S. at 331; *Or. Waste*, 511 U.S. at 99.
80 *See supra* notes 56–66 and accompanying text; *see also* Armco, Inc. v. Hardesty, 467 U.S. 638, 644 (1984) (finding that wholesale tax imposed only on imported goods burdened interstate commerce).
81 *See supra* notes 56–66 and accompanying text.
82 *See Fulton*, 516 U.S. at 331; *Or. Waste*, 511 U.S. at 99.
83 *See Fulton*, 516 U.S. at 331; *Or. Waste*, 511 U.S. at 102.
84 *See Fulton*, 516 U.S. at 331 (quoting *Lohman*, 511 U.S. at 647).
85 *See Or. Waste*, 511 U.S. at 102.
86 *See* 75 U.S. 148, 153 (1869).
87 *Id.* at 150.
in the state.\textsuperscript{88} Therefore, the Court held that this was not an attempt to regulate commerce, but an appropriate and legitimate exercise of the state’s taxing power.\textsuperscript{89}

Since \textit{Hinson}, the Court has more clearly defined and significantly limited the compensatory tax doctrine through a line of cases beginning in 1937 with \textit{Henneford v. Silas Mason} and culminating in the decision of \textit{Fulton Corp. v. Faulkner} in 1996.\textsuperscript{90} Modern application of the compensatory tax doctrine involves a three-part test set out in 1994 in \textit{Oregon Waste Systems v. Department of Environmental Quality of the State of Oregon} and refined in \textit{Fulton Corp}.\textsuperscript{91} The three conditions necessary for a valid compensatory tax are: (1) a state must identify the intrastate burden for which the state is attempting to compensate; (2) the tax on interstate commerce must be shown roughly to approximate—but not to exceed—the amount of the tax on intrastate commerce; and (3) the events on which the interstate and intrastate taxes are imposed must be substantially equivalent—that is, they must be substantially similar in substance to serve as mutually exclusive proxies for each other.\textsuperscript{92} Given the relatively short life and limited application of the formalized three-part test, it is necessary to examine earlier cases, which address each of the prongs only implicitly, to analyze the compensatory tax doctrine fully.\textsuperscript{93}

\textbf{B. The History of the Compensatory Tax Doctrine}

\textbf{1. \textit{Henneford v. Silas Mason}: Formal Validation of the Compensatory Tax}

Nearly seventy years after its decision in \textit{Hinson}, the Supreme Court formally recognized the validity of a “compensating tax” in 1937 in \textit{Henneford v. Silas Mason}.\textsuperscript{94} In \textit{Silas Mason}, the State of Washington imposed two taxes, a 2\% tax on retail sales and a compensating 2\% tax on the privilege to use any article of tangible personal property within the state.\textsuperscript{95} The use tax did not apply to articles for which a

\textsuperscript{88} \textit{Id.} at 153.
\textsuperscript{89} \textit{Id.}
\textsuperscript{91} \textit{See Fulton}, 516 U.S. at 332–33; \textit{Or. Waste}, 511 U.S. at 103.
\textsuperscript{92} \textit{See Fulton}, 516 U.S. at 332–33.
\textsuperscript{93} \textit{See infra} notes 94–189 and accompanying text.
\textsuperscript{94} \textit{See Silas Mason}, 300 U.S. at 583–84; \textit{Hinson}, 75 U.S. at 152–53.
\textsuperscript{95} \textit{Silas Mason}, 300 U.S. at 579.
tax equal to or greater than 2% had already been applied out-of-state.96 The plaintiffs in Silas Mason brought machinery, materials and supplies into Washington that were purchased at retail in other states for use on the construction of a dam on the Columbia River.97 Washington assessed a use tax on the items because they had not been subject to a sales tax out of state.98

The Court first acknowledged that the regulatory scheme was discriminatory on its face; the tax would never be payable on items purchased within the State of Washington because those items would be subject to a sales tax.99 The burden of paying the use tax, however, was imposed equally on residents and non-residents who used their property within the state.100 The Court noted that when the account was made up, the stranger from afar was subject to no greater burdens as a consequence of ownership than the dweller within the gates.101 The Court reasoned that while one paid upon one activity or incident, and the other upon another, the sum was the same.102 This reasoning implied that the sale and use of articles within the state were substantially similar events because the burdens fell on similarly described people—those taxpayers using articles in the state—and the taxes were therefore functionally equivalent.103 The Court concluded that the scheme was not an unlawful burden on interstate commerce because it did not in fact burden commerce; it did not place a greater burden on goods purchased outside the state than those purchased inside the state.104 The Court also rejected the proposition that the scheme amounted to protectionism of local retailers.105 Because the tax was imposed on use, rather than import of the goods, and there was equality in the laying of the tax, there was no protectionism.106

96 Id. at 580–81.
97 Id. at 579.
98 Id.
99 See id. at 581.
100 See Silas Mason, 300 U.S. at 583–84.
101 Id. at 584.
102 Id.
103 See id. at 584–85. This reasoning also implied that the State of Washington had a legitimate sovereign interest in taxing the use of property within the state once commerce was at an end. See id. at 582.
104 See id. at 584–85.
105 Silas Mason, 300 U.S. at 586–87.
106 See id.
2. Maryland v. Louisiana: Rejection of the “First-Use” Tax

Since the Supreme Court’s validation of the compensatory use tax in Silas Mason, it has steadfastly refused to apply the compensatory tax doctrine to areas outside the realm of sales and use taxes.107 For example, in its 1981 decision in Maryland v. Louisiana, the Supreme Court struck down a Louisiana “first use” tax imposed on any natural gas imported into the state that was not subject to taxation by another state.108 In effect this tax meant that only gas from the outer continental shelf (the “OCS”)—an area of ocean that lies beyond state, but within federal, boundaries—was subject to the tax.109 The tax imposed was equal to the severance tax the state imposed on gas producers in the state.110 Because of numerous tax exemptions and credits, however, the net effect of the tax scheme was to tax OCS gas moving through and eventually out of the state but not to tax Louisiana consumers of OCS gas.111

As an initial matter, the Court addressed the State of Louisiana’s claim that the taxable “uses” within the state broke the flow of commerce and were wholly local events, subject to state regulation.112 The Court rejected this reasoning, stating that gas crossing a state line at any stage of its movement to the ultimate consumer was in interstate commerce during the entire journey from the wellhead to the consumer, even though interrupted by certain events within a particular state.113

Finding the tax scheme facially discriminatory towards interstate commerce, the Court set out to determine whether it could be upheld as a compensatory tax under Silas Mason.114 The Court held that the compensatory tax doctrine requires identification of the burden for which the state is attempting to compensate.115 Louisiana claimed that it was attempting to compensate for the burden of the severance tax on local production of natural gas.116 The Court rejected this ar-

---

107 See Fulton, 516 U.S. at 342 (noting that the court has consistently declined to extend the compensatory exception beyond sales and use taxes); see also Or. Waste, 511 U.S. at 105; Armco, 467 U.S. at 644; Maryland v. Louisiana, 451 U.S. 725, 760 (1981).
108 451 U.S. at 731, 760.
109 Id.
110 Id.
111 Id. at 733.
112 Louisiana, 451 U.S. at 754–55.
113 Id. at 755–56.
114 See id. at 756, 758.
115 Id. at 758.
116 Id.
gument, stating that although Louisiana has an interest in protecting its natural resources and therefore could impose a severance tax on domestic producers, it had no comparable sovereign interest in being compensated for the severance of resources from land outside its boundaries. Therefore, the first-use tax could not have been designed to meet the same ends as the severance tax—it could not have been designed to protect Louisiana’s natural resources. The Court said that the “use” of gas and severance of gas could not be considered “substantially equivalent events,” reasoning implicitly that because the burden of the two taxes fell on differently described taxpayers (in-state producers and out-of-state consumers) and did not meet the same ends, the taxes were not functionally equivalent. The Court differentiated these circumstances from the case of sales and use taxes, where a state attempts to ensure uniform treatment of goods to be consumed in the state by imposing taxes on substantially similar events occurring wholly within the state.

The Court concluded that the common thread running through the cases upholding compensatory taxes was equality of treatment between local and interstate commerce. Because the pattern of credits and exemptions principally burdened gas moving out of state, the tax was not a valid compensatory tax.

3. Armco, Inc. v. Hardesty: An Emphasis on Substantially Equivalent Events

The Supreme Court zeroed in on the notion of substantially equivalent events in 1984 in Armco v. Hardesty. In that case the Court struck down a tax imposed by West Virginia on gross receipts of tangible property sold at wholesale. The Court found the regulation to be facially discriminatory because it exempted local manufacturers from the tax. West Virginia defended the tax, which was 0.27% of the wholesale price, as a compensatory tax for the far higher

117 Louisiana, 451 U.S. at 759.
118 Id.
119 See id.
120 Id.
121 Id.
122 See Louisiana, 451 U.S. at 759–60.
123 See 467 U.S. at 643.
124 Id. at 640, 646.
125 See id. at 642.
0.88% manufacturing tax on local manufacturers.\textsuperscript{126} The Court rejected the argument, holding that manufacturing and wholesale were not substantially similar events.\textsuperscript{127} The Court noted that the manufacturing tax was not reduced when the goods were sold out of state, providing evidence that the tax was in fact a manufacturing tax and not a proxy for the gross receipts tax imposed on wholesalers from out of state.\textsuperscript{128} In addition, the Court found that it would be impossible to determine which portion of the manufacturing tax was attributable to manufacturing and which portion to sales, and therefore it would be impossible to do an accounting to determine whether the tax on intrastate commerce roughly approximated the alleged compensating tax on interstate commerce.\textsuperscript{129}

The Court also noted that when the two taxes were considered together, discrimination against interstate commerce persisted because there was no exception in the wholesale tax regulation for manufacturers who had already paid a manufacturing tax in their home state.\textsuperscript{130} If the scheme were upheld, manufacturers from out of state would pay both a manufacturing tax and a wholesale tax, while a West Virginia resident would pay only a manufacturing tax.\textsuperscript{131} The Court indicated that this would clearly violate the anti-protectionist purposes of the Commerce Clause.\textsuperscript{132}

C. The Modern Compensatory Tax Doctrine and the Three-Part Test


The Supreme Court set out the three elements of the compensatory tax doctrine distinctly for the first time in 1994 in Oregon Waste Systems, Inc. v. Department of Environmental Quality of the State of Oregon.\textsuperscript{133} In that case, an Oregon-based solid waste disposal company challenged an Oregon regulation that imposed a $2.25-per-ton surcharge on out-of-state waste disposed of at landfills within Oregon.\textsuperscript{134}

\textsuperscript{126} See id.
\textsuperscript{127} Id. at 643.
\textsuperscript{128} Armco, 467 U.S. at 643.
\textsuperscript{129} Id.
\textsuperscript{130} Id. at 644.
\textsuperscript{131} Id.
\textsuperscript{132} See id.
\textsuperscript{133} See Or. Waste, 511 U.S. at 103.
\textsuperscript{134} Id. at 96–97.
The Oregon-based company regularly shipped waste from neighboring Washington into Oregon for disposal.\textsuperscript{135} The Oregon Supreme Court upheld the surcharge as a compensatory fee with an express nexus to actual costs incurred by state and local governments associated with disposing of the waste.\textsuperscript{136} The U.S. Supreme Court reversed and invalidated the surcharge.\textsuperscript{137}

The Court held that because the rule was facially discriminatory, it was per se invalid unless it advanced a legitimate local purpose that could not be adequately served by reasonable nondiscriminatory alternatives.\textsuperscript{138} The Court began by recognizing the settled principle that interstate commerce may be made to “pay its way” and that “[i]t was not the purpose of the commerce clause to relieve those engaged in interstate commerce from their just share of state . . . burdens.”\textsuperscript{139} The Court noted that since \textit{Hinson v. Lott} in 1869 the compensatory tax doctrine had been used to express these principles, while also ensuring that no state exacts more than a just share from interstate commerce, which is a central purpose of the Commerce Clause.\textsuperscript{140}

The Court set out the first and second prongs of the compensatory tax analysis requiring the state to (1) identify the intrastate burden for which the state is attempting to compensate and (2) show that the burden on interstate commerce roughly approximated, but did not exceed, the burden on intrastate commerce.\textsuperscript{141} Applying these two requirements, the Court held that Oregon’s failure to identify a \textit{specific} compensating charge on intrastate commerce equal to or exceeding the surcharge was fatal to its claim.\textsuperscript{142} Oregon claimed that the surcharge compensated for general taxes paid by Oregon residents who disposed of in-state waste.\textsuperscript{143} The Court rejected this claim because it was impossible to determine which portion of the general taxes were attributable to the disposal of waste, and therefore the Court could not determine whether the two burdens were roughly equivalent.\textsuperscript{144} Accordingly, the state failed the first two prongs of the analysis.\textsuperscript{145}

\begin{enumerate}
\item[135] \textit{Id.} at 97.
\item[136] \textit{Id.}
\item[137] \textit{Id.} at 98.
\item[138] \textit{Or. Waste}, 511 U.S. at 99.
\item[139] \textit{Id.} at 102 (citations omitted); \textit{see also Louisiana}, 451 U.S. at 753.
\item[140] \textit{See Or. Waste}, 511 U.S. at 102.
\item[141] \textit{See id.} at 103.
\item[142] \textit{Id.} at 104.
\item[143] \textit{Id.}
\item[144] \textit{See id.}
\item[145] \textit{See Or. Waste}, 511 U.S. at 104.
\end{enumerate}
The Court further stated that even if it were possible to calculate the portion of the general taxes that contributed to an intrastate burden roughly equivalent to the interstate burden, the surcharge would still be invalid because the general tax and the surcharge were not imposed on substantially equivalent events.\textsuperscript{146} Thus, the surcharge also violated the third prong of the analysis.\textsuperscript{147} Under the “equivalent events” analysis, the Court reasoned that earning income and disposing of waste were even less equivalent than wholesale and manufacturing, which were found not to be substantially equivalent in \textit{Armco}.\textsuperscript{148} The court reasoned implicitly that the two taxes were not designed to meet the same ends, because income taxes cover far more than disposal of waste and the two could not be functionally equivalent to each other.\textsuperscript{149} Moreover, the fact that Oregon-based shippers of out-of-state waste were charged the surcharge \textit{and} income tax refuted the argument that the events were substantially (or functionally) equivalent.\textsuperscript{150} The Court noted that the prototypical example of substantially equivalent events is the sale and use of articles within the state and that the only compensatory taxes upheld had been use taxes on products purchased out of state.\textsuperscript{151} The Court refused to weigh comparative burdens imposed on dissimilar events.\textsuperscript{152}

2. \textit{Fulton Corp. v. Faulkner}: The Modern Test Summarized

The modern embodiment of the compensatory tax doctrine was summarized most recently in \textit{Fulton Corp. v. Faulkner} in 1996.\textsuperscript{153} The case involved an “intangibles tax” on the fair market value of corporate stock owned by North Carolina residents.\textsuperscript{154} Residents were entitled to take a deduction equal to the fraction of the issuing corporation’s income that was subject to tax in North Carolina.\textsuperscript{155} Therefore, if a resident owned stock in an in-state corporation the stock was not subject to the tax because the taxable percentage deduction was 100%, but stock in an out-of-state corporation was subject to the

\begin{itemize}
\item \textsuperscript{146} See id.
\item \textsuperscript{147} See id. at 103.
\item \textsuperscript{148} See id. at 105; \textit{Armco}, 467 U.S. at 643.
\item \textsuperscript{149} See \textit{Or. Waste}, 511 U.S. at 104–05.
\item \textsuperscript{150} See id. at 105.
\item \textsuperscript{151} Id.
\item \textsuperscript{152} See id.
\item \textsuperscript{153} See 516 U.S. at 332–33.
\item \textsuperscript{154} Id. at 327.
\item \textsuperscript{155} Id. at 327–28.
\end{itemize}
Therefore, the Court first determined that the regulatory scheme was facially discriminatory.\textsuperscript{157}

The Supreme Court again recognized that there may be cases where a facially discriminatory tax may be upheld if the combined effect of the multi-tax scheme is to subject intrastate and interstate commerce to equivalent burdens.\textsuperscript{158} The Court then reiterated the three conditions necessary for a valid compensatory tax: (1) a state must identify the intrastate burden for which the state is attempting to compensate; (2) the tax on interstate commerce must be shown roughly to approximate—but not to exceed—the amount of the tax on intrastate commerce; and (3) the events on which the interstate and intrastate taxes are imposed must be so substantially similar in substance as to serve as mutually exclusive proxies for each other.\textsuperscript{159}

To meet its burden under the first prong, North Carolina argued that the taxable percentage deduction (i.e. the tax on out-of-state stock interests) compensated for the burden of the general corporate income tax paid by corporations doing business in North Carolina.\textsuperscript{160} The Court rejected this argument, holding that in addition to merely identifying the intrastate burden for which it seeks to compensate, the state must also show that the intrastate tax serves some purpose for which the state may otherwise impose a burden on interstate commerce.\textsuperscript{161} The Court held that because North Carolina had no general sovereign interest in taxing income earned out of state, it would fail the first prong of the analysis unless the state could identify some in-state activity or benefit to justify the compensatory tax.\textsuperscript{162} North Carolina attempted to cure this deficiency by pointing out that the out-of-state corporations benefited from the use of the state’s capital markets without paying corporate income tax and that the intangible tax compensated for this loss.\textsuperscript{163} The Court declined to create a precedent that would allow the imposition of a tax on entities involved in interstate commerce any time they happened to use facilities supported by general state tax funds.\textsuperscript{164}

\textsuperscript{156} Id. at 328.
\textsuperscript{157} Id. at 333.
\textsuperscript{158} Fulton, 516 U.S. at 331.
\textsuperscript{159} Id. at 332–33.
\textsuperscript{160} See id. at 334.
\textsuperscript{161} See id.
\textsuperscript{162} Id.
\textsuperscript{163} Fulton, 516 U.S. at 334–35.
\textsuperscript{164} See id. at 335.
Under the second prong of the analysis, the Court in *Fulton* addressed the problem of interstate burdens that are imposed as a compensatory measure for generally defined intrastate burdens. The second prong requires that the burden on interstate commerce be shown roughly to approximate, but not exceed, the amount of the burden on intrastate commerce. North Carolina justified the intangibles tax and corresponding taxable percentage deduction as a measure for maintenance of the capital market for the shares of both foreign and domestic corporations. The Court noted that the tax for which the state purported to compensate was a general corporate income tax that paid for a wide range of things, including construction and maintenance of a transportation network, institutions to educate a workforce, and local fire and police protection. The state could not say what percentage of that general tax was allocated to support the capital market and whether that proportion was greater or smaller than the one imposed on interstate commerce by the intangibles tax. The Court emphasized the point made in *Oregon Waste*, namely that it is generally unwilling to make the complex quantitative assessments required by the compensatory tax doctrine when general forms of taxation are involved. The Court confirmed its unwillingness to permit discriminatory taxes on interstate commerce to compensate for charges purportedly included in general forms of intrastate taxation.

In addressing the third prong of the analysis, the Court noted that recent decisions expressed an extreme reluctance to recognize new compensatory categories outside the sales/use tax combination recognized in *Silas Mason*. The third prong requires that the compensating burdens fall on “substantially equivalent events.” The Court explained that to meet this requirement the taxed activities must be sufficiently similar in substance to serve as mutually exclusive proxies for each other and that the two taxes must be functionally equivalent. The Court held that actual incidence of the tax upon the same class of taxpayers is a necessary precondition for a valid

---

165 See id. at 337–38.
166 Id. at 336.
167 See id. at 338.
168 See *Fulton*, 516 U.S. at 337.
169 Id. at 338.
170 See id. (citing *Or. Waste*, 511 U.S. at 105 n.8).
171 *Fulton*, 516 U.S. at 338 (quoting *Or. Waste*, 511 U.S. at 105 n.8).
172 *Fulton*, 516 U.S. at 338.
173 Id.
174 See id. at 339 (quoting *Or. Waste*, 511 U.S. at 103).
compensatory tax, reasoning that if the burden falls on differently described entities then the taxes cannot be functionally equivalent. The Court recognized that the ultimate distribution of burdens may be different from the statutory distribution of burdens, particularly when the nominal taxpayer can pass the burden to other parties, such as consumers. The Court held that a state defending a compensatory tax scheme has the burden of showing that, at a minimum, the actual incidence of the two burdens is such that the real taxpayers are within the same class, so that a finding of combined neutrality as to interstate commerce is at least possible.

North Carolina argued that because corporate earnings influence the price of stock, the intangibles tax and the income tax are essentially taxing the same event. The Court held that this was insufficient, and that the difference between the parties on which the taxes fell was of great significance. The Court noted that in Silas Mason, the use tax was acceptable because the effect of the regulatory regime was to help in-state retailers to compete on terms of equality with retailers in other states who are exempt from a sales tax or other corresponding burden. In Silas Mason, all taxpayers using their property within the state bore an equal burden whether paying a use tax or a sales tax. This equality did not exist in Fulton because the allegedly compensating taxes fell on taxpayers who were differently described. The income tax paid by corporations doing business in the state would be reflected in the stock price, and the actual burden of the tax would be borne by other parties such as consumers of the corporations’ products. By contrast, the Court stated, it was unlikely that the stock price of corporations doing business outside the state would reflect the impact of the incidence tax because North Carolina investors make up a small por-

175 See Fulton, 516 U.S. at 340.
176 Id. at 341.
177 Id. at 340. The Court noted that a finding that the burden falls on the same class of taxpayers is a condition precedent for a finding that the two taxes are complementary, and declined to decide whether mere incidence is sufficient to compel the conclusion that the two burdens fall on substantially equivalent events. Id. at 340 n.6 (citing Armco, 467 U.S. at 643).
178 See Fulton, 516 U.S. at 339.
179 Id. at 340.
180 See id. at 340 (citing Silas Mason, 300 U.S. at 581).
181 See Fulton, 516 U.S. at 340.
182 See id.
183 See id. at 343.
tion of the national market. The economic incidence of that tax would fall on the resident shareholder.

The Court noted that the objective of the “equivalent event” requirement is to enable in-state and out-of-state businesses to compete on equal footing. The combination of the two tax schemes violated this objective because the actual incidence of the intangibles tax fell squarely on the shareholder and thus encouraged North Carolina investors to favor investment in corporations doing business within the state. The Court stated that the compensatory tax doctrine is fundamentally concerned with equalizing competition between in-staters and out-of-staters. The Court cautioned, however, that the difficulty in comparing the economic incidence of allegedly complementary tax schemes on different taxpayers and different transactions leads to the conclusion that courts will be unable to evaluate equivalency outside the context of traditional sales/use taxes.

III. Applying Commerce Clause Jurisprudence to the Problem of Leakage

Although addressing climate change at the state and regional levels is certainly suboptimal, it could eventually have important effects on national policy. Individual state actions create a patchwork of policies around the country that is both inefficient for businesses and risky for the acting states, which risk driving business out of state. This patchwork, however, often inspires the regulated community to lobby the federal government for national action. In addition, states can serve as laboratories for experimenting with various regulatory options to determine the best model for national action.

184 Id.
185 Id.
186 Fulton, 516 U.S. at 340.
187 See id. at 343.
188 See id. at 342 n.8.
189 See id.
191 Id.
192 See id. at 4.
193 See New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting) (“There must be power in the states and the nation to remould, through experimentation, our economic practices and institutions to meet changing social and economic needs.”); see also McKinstry, supra note 64, at 15–16 (noting that although states serve as laboratories for environmental policy change and often serve as a template for federal action, they face unique challenges).
RGGI may serve this experimental function well.\textsuperscript{194} For these reasons the RGGI states should use every avenue available to them, including the compensatory tax doctrine, to protect the regulatory scheme from invalidation based on the Interstate Commerce Clause.\textsuperscript{195}

The compensatory tax doctrine may not be accepted by a court as applicable legal doctrine for emissions regulations because the burden imposed is not in the form of a tax.\textsuperscript{196} As the Supreme Court noted in \textit{Oregon Waste, Inc. v. Department of Environmental Quality of the State of Oregon}, however, the compensatory tax doctrine is merely a specific way of justifying a facially discriminatory tax that achieves a legitimate local purpose that cannot be achieved through nondiscriminatory means.\textsuperscript{197} The RGGI states should therefore argue for the expansion of compensatory tax doctrine principles to cover important state and regional environmental regulations such as the RGGI program.\textsuperscript{198}

A regulatory approach that adequately addresses leakage may require a second regulation imposing a cap on the emissions associated with electricity imported into the region by LSEs.\textsuperscript{199} This regulation would be passed after the implementation of the first cap on domestic electricity generators and only if it was determined that leakage was undermining the goals of the program.\textsuperscript{200} If the RGGI states decide to use this hybrid regulatory approach to address leakage, they should employ the compensatory tax doctrine to defend the scheme by arguing that the regulation of imported electricity is necessary to further a legitimate local purpose and that the combination of the two regulations is nondiscriminatory in effect.\textsuperscript{201} The purpose of

\textsuperscript{194} See M.J. Bradley & Associates, \textit{supra} note 22, at 2; \textit{supra} notes 1–8, 18–22 and accompanying text. In addition, if California adopts a different program for reduction of CO\textsubscript{2} emissions from that adopted by RGGI, the impact of the two programs on the national regulated community could be significant and force federal action. See M.J. Bradley & Associates, \textit{supra} note 22, at 2; \textit{supra} notes 1–8, 18–22 and accompanying text.

\textsuperscript{195} See \textit{supra} notes 190–194 and accompanying text.

\textsuperscript{196} See \textit{supra} note 107 and accompanying text.

\textsuperscript{197} \textit{See 511 U.S. 93, 102 (1994).}

\textsuperscript{198} \textit{See id; supra} notes 190–194 and accompanying text.

\textsuperscript{199} See \textit{supra} notes 57–63 and accompanying text.

\textsuperscript{200} See \textit{supra} notes 57–63 and accompanying text.

\textsuperscript{201} \textit{See Fulton Corp. v. Faulkner, 516 U.S. 325, 342 (1996). There is an obvious distinction between a tax and a regulation limiting CO\textsubscript{2} emissions. See \textit{supra} note 107 and accompanying text. The emissions cap does, however, ultimately impose burdens on generators of electricity that wish to participate in interstate commerce with the RGGI states. See \textit{supra} notes 57–66 and accompanying text. This burden on electricity crossing regional borders is analogous to the burden imposed by the traditional taxes considered under the compensatory tax doctrine. See \textit{supra} notes 57–66 and accompanying text.}
the initial emissions regulation imposed on generators is to reduce CO₂ emissions associated with in-state electricity consumption in order to protect the state’s interests in public health and welfare and preservation of natural resources.\textsuperscript{202} The regulation of imported electricity through LSEs is a compensatory measure designed simply to make interstate commerce bear a burden already borne by intrastate commerce.\textsuperscript{203} In other words, the combination of the two regulations merely levels the playing field across all electricity generators serving the region.\textsuperscript{204} Because the LSE regulation is necessary to effectuate the purpose of the initial regulation, the combination of the two does not have a discriminatory effect, and is therefore a legitimate compensatory “tax” or burden on interstate commerce.\textsuperscript{205}

A. Application of the Compensatory Tax Doctrine to the Hybrid Approach

Under the first step of the dormant Commerce Clause analysis, a court would likely determine that the second regulation imposed on LSEs does not regulate evenhandedly with only incidental effects on interstate commerce.\textsuperscript{206} Rather, because the LSE regulation only regulates emissions associated with electricity that crosses state lines while exempting domestically generated electricity, a court would likely de-
termine that the regulation of imported electricity discriminates against interstate commerce on its face.\(^{207}\)

As noted above, courts review facially discriminatory regulations under a strict scrutiny test based on the assumption that they are per se invalid.\(^{208}\) To overcome this assumption, the proponent of the regulation must show that the regulation advances a legitimate local purpose that cannot adequately be served by reasonable nondiscriminatory alternatives.\(^{209}\) This may be shown by applying the principles embodied in the three prongs of the compensatory tax doctrine.\(^{210}\)

1. Application of the First Prong of the Compensatory Tax Doctrine: Identifying the Intrastate Burden Requiring Compensation

Under the first prong of the compensatory tax doctrine, the RGGI states must identify the intrastate burden for which the regulation of emissions associated with electricity imports seeks to compensate.\(^{211}\) It is reasonable to assume that a court would accept the assertion that the states adopted the regulation in good faith as compensation for the domestic burden of the emission cap placed on in-state generators, rather than suspecting some ulterior motive.\(^{212}\) The states will also be required to show that the intrastate burden serves some purpose for which the state may otherwise impose a burden on interstate commerce.\(^{213}\)

One purpose of the intrastate burden on electricity generators is to protect the RGGI states’ natural resources and public health by reducing CO\(_2\) emissions associated with electricity consumption.\(^{214}\) The

\(^{207}\) See Fulton, 516 U.S. at 333 (finding a statute that burdens interstate commerce but not intrastate commerce to be facially discriminatory); Or. Waste, 511 U.S. at 100 (stating that a law that taxes interstate activities more heavily is facially discriminatory); supra notes 56–63 and accompanying text; see also Kirsten H. Engel, Mitigating Global Climate Change in the United States: A Regional Approach, 14 N.Y.U. Env’tl. L.J. 54, 77–78 (2005) (concluding that an outright ban on importation of electricity would be a facially discriminatory Commerce Clause violation unless it was expressly authorized by Congress).

\(^{208}\) See supra notes 73–92 and accompanying text.

\(^{209}\) See supra notes 73–92 and accompanying text.

\(^{210}\) See supra notes 83–92 and accompanying text.

\(^{211}\) See Fulton, 516 U.S. at 332; Or. Waste, 511 U.S. at 103.

\(^{212}\) See Fulton, 516 U.S. at 337 (expressing suspicion that the reason given for imposing an allegedly compensatory tax was illusory).

\(^{213}\) See Maryland v. Louisiana, 451 U.S. 725, 759 (1981); see also Fulton, 516 U.S. at 334 (holding that North Carolina could not impose a tax on foreign corporations compensating for the burden of income tax on domestic corporations because North Carolina had no sovereign interest in taxing the income of a foreign corporation).

\(^{214}\) See RGGI MoU, supra note 1, at 1.
RGGI states do not, however, have a sovereign interest in protecting other states’ natural resources or public health and safety, and no court has yet held that a state has a legitimate interest in reducing global pollutants outside its borders.\textsuperscript{215} Therefore, a court could invalidate the regulation because the intrastate burden that the generator regulation imposes serves a purpose for which the state may not otherwise burden interstate commerce.\textsuperscript{216} Similarly, the Court in \textit{Maryland v. Louisiana} rejected the argument that because the state imposed a severance tax on gas extracted from its own soil, it could impose a compensating first-use tax on imported gas.\textsuperscript{217} The Court held that Louisiana had no sovereign interest in the severance of resources from land outside its borders, that the alleged compensating tax was invalid, and that the state had to identify an in-state activity in order to justify the first-use tax.\textsuperscript{218}

The LSE regulation that the RGGI states may impose, however, is unlike the Louisiana regulation in \textit{Louisiana} because it is necessary to promote the states’ legitimate interest while the Louisiana regulation was not.\textsuperscript{219} In both \textit{Louisiana} and \textit{Fulton Corp. v. Faulkner}, the Court held that the state must identify an in-state activity or benefit to justify the compensatory levy, a task that neither of the states could do.\textsuperscript{220} The RGGI states, on the other hand, may be able to overcome the sovereign interest argument by showing that the regulation of emissions associated with imported electricity is necessary to carry out the purposes of the in-state regulation, and that the two regulations are designed to meet the same end.\textsuperscript{221} The RGGI states should argue that the combination of the two regulations serves the legitimate local interest of protecting the natural resources and the health and welfare

\textsuperscript{215} See \textit{Louisiana}, 451 U.S. at 759; \textit{cf.} Massachusetts v. EPA, 415 F.3d 50, 54–56 (D.C. Cir. 2005), \textit{cert. granted}, 126 S. Ct. 2960 (2006). In \textit{Massachusetts v. EPA}, Judge Randolph, writing for a three-judge panel of the U.S. Court of Appeals for the District of Columbia Circuit, assumed without deciding that a state has standing to bring an action based on the generalized grievance of harms associated with global warming. 415 F.3d at 54–56. Judge Sentelle, dissenting in part but concurring in the judgment, stated that the state did not have standing because it did not assert a specific harm associated with CO\textsubscript{2} emissions. \textit{Id.} at 59–60 (Sentelle, J., dissenting in part but concurring in the judgment). Judge Tatel, dissenting, stated that the state did have standing, in part because it had successfully shown injury caused by global warming. \textit{Id.} at 64 (Tatel, J., dissenting).

\textsuperscript{216} See \textit{Louisiana}, 451 U.S. at 759.

\textsuperscript{217} See \textit{id.}

\textsuperscript{218} See \textit{id.}

\textsuperscript{219} See \textit{id.} at 759; supra notes 43–63 and accompanying text.

\textsuperscript{220} See \textit{Fulton}, 516 U.S. at 334; \textit{Louisiana}, 451 U.S. at 759.

\textsuperscript{221} See \textit{Or. Waste}, 511 U.S. at 99; \textit{Louisiana}, 451 U.S. at 759.
of their citizens, and that a regulatory scheme that places some burden on interstate commerce is necessary to effectuate that purpose.\textsuperscript{222} If the states cannot regulate emissions from imported electricity, then the regulation of domestic emissions will not be effective.\textsuperscript{223} This argument thus directly addresses the requirement set out in \textit{Louisiana} that the two regulations be designed to meet the same end, because both regulations are ultimately designed to protect the states’ public health and natural resources.\textsuperscript{224} Based on this reasoning, the RGGI states could persuade a court that the regulation satisfies the first prong of the compensatory tax doctrine.\textsuperscript{225}

2. Application of the Second Prong of the Compensatory Tax Doctrine: Equivalent Burdens on Interstate and Intrastate Commerce

The second prong of the compensatory tax analysis requires that the burden on interstate commerce roughly approximate, but not exceed, the burden on intrastate commerce.\textsuperscript{226} The RGGI states will not encounter the problems faced by states that sought to compensate for burdens imposed on intrastate commerce by general forms of taxation as Oregon did in \textit{Oregon Waste} and North Carolina did in \textit{Fulton}, because the LSE regulation compensates for a specific regulation focused on emissions from fossil fuel-fired generators, rather than a generally applicable resident taxation.\textsuperscript{227} The complexity of the accounting in the case of emissions trading, however, is sure to raise its own challenges.\textsuperscript{228}

It is well-established that pure economic protectionism is not considered a legitimate local purpose under Commerce Clause juris-

\textsuperscript{222} \textit{See} Maine v. Taylor, 477 U.S. 131, 151–52 (1986) (upholding a facially discriminatory law banning the importation of out-of-state bait fish into Maine because the fish were subject to parasites completely foreign to Maine baitfish and could jeopardize the health of the Maine fish population, and no nondiscriminatory alternatives existed).

\textsuperscript{223} \textit{See supra} notes 43–52 and accompanying text.

\textsuperscript{224} \textit{See} Louisiana, 451 U.S. at 759.

\textsuperscript{225} \textit{See supra} notes 211–224 and accompanying text.

\textsuperscript{226} \textit{See supra} notes 141–145, 165–171 and accompanying text.

\textsuperscript{227} \textit{See Fulton}, 516 U.S. at 338; \textit{Or. Waste}, 511 U.S. at 104–05; \textit{supra} notes 141–145, 165–171 and accompanying text.

\textsuperscript{228} \textit{See} Armco, Inc. v. Hardesty, 467 U.S. 638, 643 (1984) (striking down an allegedly compensatory tax, in part because the court could not determine which part of the tax was meant to be compensatory); \textit{California Report}, \textit{supra} note 3, at 21, 23 (discussing the current lack of a robust emissions-tracking system for LSEs); \textit{see also} Fulton, 516 U.S. at 338; \textit{Or. Waste}, 511 U.S. at 104–05.
Therefore, if a regulation had the effect of putting in-state generators at a competitive advantage over out-of-state generators, that regulation would be struck down. It is therefore critical that the RGGI states consider this issue from the beginning of the program if they intend to address the problem of leakage in the future. Even under a regulatory scheme that only targets domestic generators, the RGGI states should set the initial cap on emissions for the region at a level that includes the emissions associated with historic imports on the same basis as historic in-region generation.

Under the second regulation, LSEs should receive allowance allocations on the same basis that generators are given allowances. Any inequality in the method by which allowances are distributed to LSEs for their imports as compared to domestic generators could lead a court to detect economic protectionism. The RGGI states must be able to show that the allocation of allowances to imported electricity under the second regulatory measure is nondiscriminatory because it is based on the same historic baseline as the allocation of allowances to domestic generators.

Also critical to the defense of the regulation on imported electricity will be the method by which actual emissions associated with imported electricity and domestic generator emissions will be measured. If the methods used are not the same, then the states will run a greater risk of having the regulation of imported electricity struck down, because the court will not be able to weigh the burdens quanti-


230 See Or. Waste, 511 U.S. at 106; Silas Mason, 300 U.S. at 586.

231 See California Report, supra note 3, at 21–23. Because the initial phase of the program will only regulate in-region generators, the RGGI states must determine how to allocate the allowances associated with historic imports so as to avoid allocation problems later in the event that the regulation of imported electricity is required. See id. If the RGGI states allocate the entire cap of allowances including those associated with historic imports to generators during the initial phase of the program and leakage becomes a problem, then there will be the serious issue of reallocating those allowances associated with historic imports to the newly regulated LSEs. See id.; cf. Armco, 467 U.S. at 645 (rejecting an allegedly compensatory tax in part because it was unclear which part of the tax was intended to be compensatory).

232 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.

233 See California Report, supra note 3, at 22–23; Cowart, supra note 7, at 5.

234 See Fulton, 516 U.S. at 338; Or. Waste, 511 U.S. at 104–05.

235 See Fulton, 516 U.S. at 338; Or. Waste, 511 U.S. at 104–05.

236 See Fulton, 516 U.S. at 338; Or. Waste, 511 U.S. at 104–05.
For example, if the states use actual emission rates for in-state generators because they are able to inspect those plants, but use assumed rates based on megawatt-hour output for imported electricity because they are unable to inspect out-of-state plants, a court could find that either the burdens were not equivalent or that it was too cumbersome to attempt to weigh them. For this reason, RGGI should use a common system of assigning CO$_2$ attributes to electricity for both generators and LSEs.

Both the allocation of allowances and measurement of emissions will likely raise the sort of difficult quantitative questions that the Supreme Court has continually used to strike down compensatory regulations. For example, in Armco, Inc. v. Hardesty, the state attempted to impose a wholesale interstate tax to compensate for a manufacturing intrastate tax. The Court complained that it could not determine what part of the manufacturing tax was attributable to manufacturing and what part to sales and therefore it struck down the burden on interstate commerce, even though the burden on intrastate activities was arguably the greater of the two. To survive the second prong of the analysis, the RGGI states must ensure that the accounting of allowances and emissions reveals the actual burdens imposed and that the burden on interstate commerce is no greater than the burden on intrastate commerce.

3. Application of the Third Prong of the Compensatory Tax Doctrine: Substantially Equivalent Events

Under the third prong of the compensatory tax doctrine, the RGGI states would be required to show that the events on which the

---

237 See Fulton, 516 U.S. at 342; Or. Waste, 511 U.S. at 105; Armco, 467 U.S. at 643.
238 See Armco, 467 U.S. at 643 (striking down an allegedly compensatory tax in part because the court could not determine what portion of the tax compensated for the in-state burden); California Report, supra note 3, at 23 (noting the various difficulties associated with tracking emissions).
239 See Fulton, 516 U.S. at 342; Or. Waste, 511 U.S. at 105; Armco, 467 U.S. at 643. There is currently not a robust tracking system for LSEs to monitor emissions associated with electricity they deliver to customers. See California Report, supra note 3, at 23. Options for developing a tracking system include relying on average emissions and requiring power contracts to include emissions data for electricity delivered. Id.
240 See supra notes 129, 142–144, 168–171 and accompanying text.
241 See 467 U.S. at 643.
242 See Armco, 467 U.S. at 643; see also Or. Waste, 511 U.S. at 104–05 (refusing to engage in the type of quantitative assessments that the compensatory tax doctrine requires).
243 See Silas Mason, 300 U.S. at 584; see also Fulton, 516 U.S. at 342; Or. Waste, 511 U.S. at 105; Armco, 467 U.S. at 643.
interstate and intrastate burdens fall are substantially equivalent; that is, they are sufficiently similar in substance to serve as mutually exclusive proxies for each other. 244 The states should be able to show that emissions associated with imported electricity and emissions from domestically generated electricity serve as mutually exclusive proxies for each other because they are functionally equivalent. 245 In Louisiana, the Court held that severance of natural gas and import of gas into the state for “use” were not comparable and no equality existed because the state was not ensuring uniform treatment of goods and materials to be consumed in the state. 246 Instead, goods were burdened differently depending on whether or not they were destined for interstate commerce. 247 By contrast, in the case of the regulations imposed on LSEs and in-state generators, the states are attempting to impose a burden on imported electricity equivalent to the burden on domestic electricity to ensure uniform treatment of electricity consumed in the state. 248 This treatment is unlike that in Louisiana, but similar to that in Henneford v. Silas Mason, where the Court upheld the combination of the sales and use taxes because the regulations ensured uniform treatment of goods and the burden was imposed equally on residents and non-residents making use of goods within the state. 249

Moreover, the burden of the two regulations falls on the same class of actors—either all generators selling to the state whether resident or non-resident, or ultimately on all consumers within the state—and thus the regulations are functionally equivalent. 250 This is similar to the Court’s reasoning in Hinson v. Lott, where it upheld a tax on each gallon of liquor imported into the state on the ground that it complemented a tax of equal magnitude on each gallon of liquor distilled in the state and was necessary to equalize competition between in-staters and out-of-staters. 251 Here, as in Hinson, the two “taxes” are functionally equivalent and therefore the two events serve as mutually exclusive proxies for each other. 252 The hybrid scheme is

244 See Fulton, 516 U.S. at 332–33.
245 See id.
246 See supra notes 114–122 and accompanying text.
247 See supra notes 114–122 and accompanying text.
248 See supra notes 114–122 and accompanying text.
249 See Silas Mason, 300 U.S. at 584; cf. Louisiana, 451 U.S. at 725.
251 See supra notes 86–89 and accompanying text.
252 See supra notes 86–89 and accompanying text; see also Fulton, 516 U.S. at 339; Or. Waste, 511 U.S. at 103.
unlike the scheme in *Fulton*, where the Court expressly found that the actual incidence of the burdens due to the corporate income tax and the intangibles tax fell on differently described taxpayers.\(^{253}\) The Court concluded that because one tax fell on domestic corporations while the other fell on individuals investing in out-of-state corporations, the two could not be functionally equivalent and the discriminatory regulation was invalid.\(^{254}\) By contrast, the burdens of the two regulations here fall on similarly described entities, those serving a state’s electricity market and ultimately consumers within the state, and therefore the two burdens are functionally equivalent.\(^{255}\)

As sensible as this argument seems, the fact remains that the Supreme Court has continuously refused to acknowledge any expansion of the compensatory tax doctrine beyond the sales and use tax category since its 1937 decision in *Silas Mason*.\(^{256}\) For example, in *Armco* the Court held that manufacturing and wholesale are not substantially similar events, reasoning that the taxes imposed on the two were not functionally equivalent to each other.\(^{257}\) The Court in *Fulton* stated: “*Hinson* does not alter our conclusion today that Courts will ordinarily be unable to evaluate the economic equivalence of allegedly complementary tax schemes that go beyond traditional sales/use taxes.”\(^{258}\) It appears that the Court is largely unwilling to open the door to allowing facially discriminatory regulations as alleged compensatory regulations outside sales and use taxes because the quantitative evaluations required to determine whether the burdens are equivalent are too cumbersome.\(^{259}\) The principles embodied in the doctrine, however, are still of value to the RGGI states because they may form the foundation of an argument for upholding the regulation.\(^{260}\)

For one, there is a critical distinction between the hybrid approach to regulating emissions and each of the allegedly compensatory taxes that the Court has struck down since *Silas Mason*.\(^{261}\) In every other case the Court has found that the combination of regulations either did in effect, or had the potential to, favor domestic in-

\(^{253}\) See supra notes 173–189 and accompanying text.

\(^{254}\) See supra notes 173–189 and accompanying text.

\(^{255}\) See supra notes 173–189 and accompanying text; see also *Silas Mason*, 300 U.S. at 584.

\(^{256}\) See *Fulton*, 516 U.S. at 338 (emphasizing the Court’s reluctance to extend the compensatory tax doctrine beyond the context of sales and use taxes).

\(^{257}\) See supra notes 123–132 and accompanying text.

\(^{258}\) *Fulton*, 516 U.S. at 342 n.8.

\(^{259}\) See *Fulton*, 516 U.S. at 342; *Or. Waste*, 511 U.S. at 105; *Armco*, 467 U.S. at 643.

\(^{260}\) See supra notes 210–255 and accompanying text.

\(^{261}\) See infra notes 262–266 and accompanying text.
terests over out-of-staters.262 By contrast, here it is assumed that the RGGI states will design a combination of regulations that do not favor domestically generated electricity over imported electricity.263 In Louisiana, the combined effect of the imposed tax and tax credit scheme was to burden only gas traveling out of state; therefore, the tax was invalidated.264 In Fulton, the regulations had the effect of encouraging North Carolinians to invest in domestic rather than out-of-state companies.265 By contrast, as was the case in Silas Mason, the RGGI regulations would have the effect of burdening all electricity consumed in the state equally and therefore should be upheld.266

It is also worth considering the words of the Court in Armco when evaluating the manufacturing and wholesale taxes.267 The Court noted that because no exception existed in the regulation for imported goods already subject to manufacturing tax in another state, the combination of the two regulations could have the effect of favoring domestic goods.268 If out-of-state generators are subject to emissions caps in their home states, then the LSE regulation will not further burden them because they will already be producing clean electricity and the regulation will not run into the Armco problem.269 Before implementing a hybrid approach, however, the RGGI states must consider whether the regulation on imported electricity requires some exceptions.270 For example, the regulation should account for other potential burdens associated with CO₂ emissions that are not imposed in the RGGI region but could be imposed in other states, such as CO₂ emissions taxes.271 Taking this issue into account as well as the Court’s approach to compensatory taxes, the RGGI states may be able to avoid invalidation of future attempts to address leakage.272

262 See Fulton, 516 U.S. at 343; Or. Waste, 511 U.S. at 106; Armco, 467 U.S. at 644; Louisiana, 451 U.S. at 759.
263 See supra notes 226–243 and accompanying text.
264 See Louisiana, 451 U.S. at 759.
265 See Fulton, 516 U.S. at 343.
266 See Silas Mason, 75 U.S. at 154–55; supra notes 56–63 and accompanying text.
267 See Armco, 467 U.S. at 644.
268 See id.
269 See id.
270 See id.
271 See id.; supra notes 56–63 and accompanying text.
272 See supra notes 206–266 and accompanying text.
The ultimate solution to the problem of leakage is to implement a nationwide regulatory program for greenhouse gas emissions. Until that time, state regulators must do their best to combat global warming by implementing regional programs and to prevent leakage. If RGGI is committed to a supply-side regulatory scheme, there are several factors that should be considered before implementing a regulation that covers imported electricity through LSEs. In order to meet the requirements of the compensatory tax doctrine, the RGGI states must be able to show with absolute certainty that the combined effect of the regulations is to impose equal burdens on electricity to be consumed within the state—that the burden on interstate commerce is no greater than the burden on intrastate commerce.

The initial carbon dioxide emissions cap should be set at levels that include emissions associated with historic imports as well as historic in-state generation, to avoid difficult accounting of allowances in the second phase of the program. The RGGI states must also determine how to allocate allowances associated with historic imports during the phase of the program that only subjects in-state generators to regulation. The allocation of allowances must not favor in-state electricity generators.

The RGGI states must also determine how carbon emission attributes of imported electricity should be measured. The method used should be the same as the method used for measuring emissions associated with domestically generated electricity, to avoid any differential treatment of in-state and out-of-state generators that could invalidate the regulation.

Even if these precautions are taken by the RGGI states, there is still a substantial likelihood that a court would strike down the regulation of imported electricity through in-state LSEs as a violation of the dormant Commerce Clause. Faced with such a challenge, RGGI states should argue that the principles embodied in the compensatory tax doctrine should be applied to validate the regulatory scheme, because the scheme achieves a legitimate local purpose that cannot be achieved through nondiscriminatory means. First, the regulation of imported electricity compensates for the domestic burden caused by the emission cap placed on in-state generators. Second, the regulatory scheme places equal burdens on both in-state and out-of-state actors by placing them on equal footing with regard to emissions allowances. Third, the emissions associated with imported electricity and emissions from domestically-generated electricity serve as mutually exclu-
sive proxies for each other. A court reviewing the hybrid regulatory scheme should accordingly extend the applicability of this dormant Commerce Clause exception.

In the event that the RGGI program merely regulates in-region generators and does not address the problem of leakage, the program will still be a valuable tool. The action of the RGGI states, in combination with actions taken in other states such as California, may be the catalyst required to set a national movement in motion. At the very least, the RGGI program will inform other regulators around the country of the strengths, weaknesses, and potential pitfalls of a cap-and-trade program for regulating greenhouse gas emissions.

HEDDY BOLSTER
REDEFINING PROPERTY UNDER THE DUE PROCESS CLAUSE: TOWN OF CASTLE ROCK v. GONZALES AND THE DEMISE OF THE POSITIVE LAW APPROACH

Abstract: Since Board of Regents of State Colleges v. Roth, the U.S. Supreme Court has defined property for due process purposes as a legitimate claim of entitlement rooted in a source of law independent from the Constitution and has recognized a broad variety of property interests. In 2005, however, the Court in Town of Castle Rock v. Gonzales reined in its due process property jurisprudence in determining that a court-ordered restraining order did not create property because its language was discretionary. The Court also suggested that no police protection statute, however worded, could ever constitute property because its enforcement lacks ascertainable monetary value and only indirectly benefits the protected person. This Note argues that the Court should refine the definition of property as a benefit rooted in a source of law independent from the Constitution that is conferred on a specific class subject to specific conditions and terminable only under specific conditions.

Introduction

In 1972, in Board of Regents of State Colleges v. Roth, the U.S. Supreme Court defined property for procedural due process purposes as a legitimate claim of entitlement rooted in a source of law independent from the U.S. Constitution. Using this basic framework

1 408 U.S. 564, 577 (1972) (defining property under the Due Process Clause as a “legitimate claim of entitlement” created and “defined by existing rules or understandings that stem from an independent source such as state law”); see U.S. Const. amend. XIV, § 1; Town of Castle Rock v. Gonzales, 125 S. Ct. 2796, 2803 (2005) (quoting the definition from Roth). The U.S. Constitution mentions property in three contexts: (1) in the Fourteenth Amendment Due Process Clause, which provides that “[No State] shall deprive any person of life, liberty, or property, without due process of law”; (2) in the Fifth Amendment Due Process Clause, which provides that the federal government shall not “deprive any person of life, liberty, or property, without due process of law”; and (3) in the Takings Clause of the Fifth Amendment, which states that private property shall not be taken for public use without just compensation. U.S. Const. amendments. V; XIV. Because the Takings Clause is limited to private property, a textual analysis may imply that property under either the Fifth Amendment or Fourteenth Amendment Due Process Clauses encompasses more than common-law notions of property. See Thomas W. Merrill, The Landscape of Constitutional...
since *Roth*, the Court has recognized a variety of property interests stemming from the common law and state and federal statutes.²

Then, in 2005 in *Town of Castle Rock v. Gonzales*, the Supreme Court revisited the contours of property under the Due Process Clause in determining whether the enforcement of a restraining order seemingly mandated under state law constituted a property interest.³ Ms. Gonzales, whose husband kidnapped and murdered her three children in open defiance of a court-ordered restraining order, claimed that the restraining order’s mandatory enforcement terms entitled her to police protection and that the Castle Rock Police Department’s policy of tolerating non-enforcement violated her procedural due process rights.⁴ The Court held that enforcement of the restraining order, despite its plain language, was not truly mandatory and, therefore, did not constitute a property interest.⁵ The Court also strongly suggested that no statute could ever entitle a person to police protection for procedural due process purposes because such protec-

---


³ See 125 S. Ct. at 2800. Although property interests may be recognized under the Due Process Clause of either the Fifth or Fourteenth Amendment, this Note will refer to the Due Process Clause in the singular for simplicity.

⁴ See *id.* at 2802. The Court accepted these facts as true because the appeal was over a granted motion to dismiss. *Id.* at 2800. Identifying whether a benefit fits the definition of property is only the first step in constitutional due process analysis. *See Roth*, 408 U.S. at 569–70. Once a court identifies a protected property interest, it must then examine whether there has been a deprivation of that interest and whether the procedural safeguards that accompany deprivation of that interest are adequate under the standards of due process. *See id.* *See generally* R. Wasserman, *Procedural Due Process Claims*, 16 TOURO L. REV. 871 (2000).

⁵ *Castle Rock*, 125 S. Ct. at 2809.
tion lacks ascertainable monetary value and only indirectly benefits the protected person.6

The Court’s analysis in Castle Rock significantly departs from the framework established in Roth and its progeny for identifying due process property interests in three key respects: (1) by narrowly reading a seemingly mandatory state statute, (2) by adding a superfluous monetary value requirement that calls into question some of the Court’s own precedents, and (3) by extending the test that examines whether a purported entitlement is indirect or incidental well beyond its previous application.7

It is possible that the Court’s decision in Castle Rock is simply an extension of its general refusal to recognize a constitutional right to government protection.8 In 1989, in DeShaney v. Winnebago County Department of Social Services, the Court held that there was no substantive

---

6 See id.
7 See id. at 2805–06, 2809; infra notes 170–235 and accompanying text. Several authors have criticized aspects of Castle Rock. See Christopher Roederer, Another Case in Lochner’s Legacy, The Court’s Assault on New Property: The Right to the Mandatory Enforcement of a Restraining Order Is a “Sham,” “Nullity,” and “Cruel Deception,” 54 Drake L. Rev. 321, 330–31 (2006) (disagreeing with the Court’s statutory and constitutional interpretation, analyzing the viability of a property interest under mandatory arrest statutes in other states, and concluding that the Court’s decision is akin to Lochner-era jurisprudence because it undermines the state’s efforts to protect its citizens); Kathleen K. Curtis, Comment, The Supreme Court’s Attack on Domestic Violence Legislation—Discretion, Entitlement, and Due Process in Town of Castle Rock v. Gonzales, 32 WM. MITCHELL L. REV. 1181, 1214–15 (2006) (stating that Castle Rock is contrary to both prior due process jurisprudence and the legislative intent of states that passed mandatory arrest laws to prevent domestic violence); Robert Michael Kline, Case Comment, Constitutional Law: Is There a Protected Interest in Protection (or Are Court Orders Merely Suggestions)?: Town of Castle Rock v. Gonzales, 125 S. Ct. 2796 (2005), 58 FLA. L. REV. 459, 465–469 (2006) (analyzing Castle Rock briefly and concluding that the Court should have used prior precedent to recognize a protected property interest); The Supreme Court, 2004 Term—Leading Cases, 119 HARV. L. REV. 208, 208–09 (2005) [hereinafter 2004 Term—Leading Cases] (highlighting Castle Rock’s departure from prior precedent in defining property interests and arguing that the decision ignores the importance of an individual’s reliance on an interest, which is central to property in due process).
8 See Castle Rock, 125 S. Ct. at 2810 (stating that the benefit to a third party of arresting someone for a crime does not trigger either procedural or substantive due process protections); Erwin Chemerinsky, The End of an Era: October Term 2004, 8 GREEN BAG 2d 345, 354 (2005) (contending that Castle Rock should be understood as rejecting a constitutional duty to provide government protection regardless of whether the claim is framed as substantive or procedural due process); John C.P. Goldberg, The Constitutional Status of Tort Law: Due Process and the Right to a Law for the Redress of Wrongs, 115 YALE L.J. 524, 592 (2005) (predicting that Castle Rock will be treated as emblematic of the broader idea that constitutional rights are negative, not affirmative). For an examination of the debate as to whether the Constitution protects negative or positive liberties, see generally David P. Currie, Positive and Negative Constitutional Rights, 53 U. CHI. L. REV. 864 (1986).
due process right to government protection from third-party harm.\textsuperscript{9} \textit{DeShaney}, however, explicitly left open the question of whether a statute requiring protection for specific persons could give rise to an entitlement, and hence a property interest, for procedural due process purposes.\textsuperscript{10} In addition, the Court’s jurisprudence has consistently distinguished between substantive due process rights to government benefits and procedural due process property interests in benefits rooted in a statutory entitlement.\textsuperscript{11} Given that the legislature can make restraining orders, unlike substantive rights, either mandatory or discretionary, recognizing the orders as entitlements would not necessarily conflict with \textit{DeShaney}’s holding that there is no substantive due process right to protection from third-party harm or the view that the Constitution merely provides negative liberties.\textsuperscript{12}

Rather, the Court’s reluctance to recognize a property interest in \textit{Castle Rock} seems to stem from the difficulty of determining what constitutes property under \textit{Roth}.\textsuperscript{13} The \textit{Roth} Court’s definition of property

\textsuperscript{9} 489 U.S. 189, 196–97 (1989). Lower courts have recognized limited exceptions to \textit{DeShaney}’s prohibition of governmental liability for third-party harm where there is a special relationship between the state and the claimant, and where the state creates the danger. See Gonzales v. City of Castle Rock, 307 F.3d 1258, 1262 (10th Cir. 2002), aff’d en banc, 366 F.3d 1093 (10th Cir. 2004), rev’d sub nom. Town of Castle Rock v. Gonzales, 125 S. Ct. 2796 (2005). For a critical response to \textit{DeShaney}, see generally Jack M. Beermann, \textit{Administrative Failure and Local Democracy: The Politics of DeShaney}, 1990 DUKE L.J. 1078 (arguing for a normative, constitutional basis for rejecting \textit{DeShaney}’s reasoning). \textit{DeShaney} even had superficially similar facts to \textit{Castle Rock}; a state social worker received complaints that the petitioner was being abused by his father, yet failed to remove the petitioner from his custody. 489 U.S. at 191.

\textsuperscript{10} 489 U.S. at 195 n.2; see \textit{Castle Rock}, 125 S. Ct. at 2803 (noting that the \textit{DeShaney} Court had left the procedural due process claim unanswered).


\textsuperscript{12} Compare \textit{DeShaney}, 489 U.S. at 195 (stating that the Due Process Clause is phrased as a limitation on the state’s power to act rather than as a guarantee of minimum standards of security), with \textit{Castle Rock}, 125 S. Ct. at 2813 (Stevens, J., dissenting) (noting that even though there is no federal constitutional guarantee of police protection under \textit{DeShaney}, there is no constitutional bar to the creation of such an entitlement under state law).

\textsuperscript{13} See \textit{Castle Rock}, 125 S. Ct. at 2805–06, 2809; \textit{Roth}, 408 U.S. at 577 (defining property as a “legitimate claim of entitlement” rooted in an independent source of law). The Court’s contradictory statements on the essence of property for due process purposes highlight the problem of defining due process property. \textit{Compare Coll. Sav. Bank v. Fla. Prepaid Postsecondary Educ. Expense Bd.}, 527 U.S. 666, 667 (1999) (writing that “[t]he hallmark of a constitutionally protected property interest is the right to exclude others”), with \textit{Logan}, 455 U.S. at 430 (writing that “[t]he hallmark of property . . . is an individual
as an entitlement rooted in a source of law other than the Constitution begs the question of what constitutes an entitlement.\textsuperscript{14} The Court has identified property interests as the source of entitlements, but has made only halting efforts to delimit their scope.\textsuperscript{15} \textit{Castle Rock’s} approach to defining property only confuses the issue and questions the continued viability of some of the Court’s own precedent.\textsuperscript{16}

This Note critiques the Court’s decision in \textit{Castle Rock} and suggests a clarification of \textit{Roth’s} conception of property as an entitlement under the Due Process Clause.\textsuperscript{17} This clarification defines an entitlement, and hence property for procedural due process purposes, as a benefit rooted in a source of law other than the Constitution that requires specific conditions for conferral on a discrete class of persons and that cannot be removed except on specific conditions.\textsuperscript{18} This definition attempts to clarify \textit{Roth’s} framework of property and harmonize the Court’s jurisprudence in this area without turning the

\textsuperscript{14} See \textit{Roth}, 408 U.S. at 577; Merrill, \textit{supra} note 1, at 921 (noting that the Court in \textit{Roth} never defined “entitlement” and that its meaning surprisingly was not an issue in subsequent cases).

\textsuperscript{15} See \textit{Logan}, 455 U.S. at 430; \textit{Roth}, 408 U.S. at 577; \textit{infra} notes 58–108 and accompanying text.

\textsuperscript{16} See 125 U.S. at 2809; \textit{infra} notes 170–235 and accompanying text.

\textsuperscript{17} See \textit{Castle Rock}, 125 S. Ct. at 2809; \textit{Roth}, 408 U.S. at 577; \textit{infra} notes 170–283 and accompanying text. This Note does not address what constitutes property for substantive due process purposes, an issue which the Court has rarely explored. See Merrill, \textit{supra} note 1, at 888; \textit{see also} Ronald J. Krotoszynski, Jr., \textit{Fundamental Property Rights}, 85 Geo. L.J. 555, 591, 609 (1997) (noting the Court’s lack of attention to property for substantive due process purposes and arguing that certain interests, such as reputation and bodily integrity, merit recognition as fundamental property interests). Two Supreme Court cases in recent years have touched upon the possibility of substantive due process property, but neither spelled out a comprehensive approach to defining what it would entail. See \textit{Coll. Sav. Bank}, 527 U.S. at 673 (suggesting that there could be fundamental property interests in holding that the interest of a business firm protected by a statutory cause of action for false advertising is not “property” within the meaning of the Due Process Clause); BMW of N. Am., Inc. v. Gore, 517 U.S. 559, 584–85 (1996) (holding that a two million dollar punitive damage award for failing to disclose that a car had been repainted was grossly excessive in relation to legitimate state interests and therefore in violation of substantive due process).

\textsuperscript{18} See \textit{Logan}, 455 U.S. at 430 (stating that the Court has emphasized that “the hallmark of property . . . is an individual entitlement grounded in state law, which cannot be removed except ‘for cause’”); \textit{Roth}, 408 U.S. at 577 (stating that the benefits that give rise to a legitimate claim of entitlement are protected property interests); Henry Paul Monaghan, \textit{Of “Liberty” and “Property”}, 62 \textit{Cornell L. Rev.} 405, 443 (1977) (suggesting that the Court’s doctrine could define property as “a present legal relationship, interruptible only for cause, plus a practical expectancy of its continuance”).
Due Process Clause into a “font of tort law.” It also hews closer than the Court’s current approach to property’s underlying purpose of protecting claims that people rely on in their daily lives and the Due Process Clause’s goal to limit arbitrary decision making.

Part I of this Note reviews the Court’s treatment of due process protections for liberty and property as a unitary concept prior to Roth, and then describes Roth’s approach to defining property for due process purposes as an entitlement stemming from “positive law.” Part II focuses on the Court’s post-Roth application of the entitlement definition of property and the problems it engenders, as well as the Court’s interpretation of statutes in the prisoners’ liberty interest area using Roth’s definition of liberty. Part III discusses the decision in Castle Rock, where the Court adopted a modified and more limited view of property under Roth’s definition. Part IV critiques the Court’s statutory analysis in Castle Rock, its adoption of an additional “ascertainable monetary value” test, and its misapplication of an incidental benefit test in identifying due process property interests. In place of the Court’s monetary value and misapplied direct benefit tests, Part V suggests a refined definition of property for due process purposes: a benefit for a discrete class of persons rooted in an independent source of law that requires specific conditions for its conferral and that cannot be terminated without satisfying specific conditions.

---

19 See Paul v. Davis, 424 U.S. 693, 701 (1976) (stating that the Due Process Clause should not be turned into a “font of tort law” to be superimposed upon state tort law systems).

20 See Daniels v. Williams, 474 U.S. 327, 331 (1986) (stating that “the Due Process Clause . . . was ‘intended to secure the individual from the arbitrary exercise of the powers of government’” (quoting Hurtado v. California, 110 U.S. 516, 527 (1884))); Roth, 408 U.S. at 577.

21 See infra notes 26–57 and accompanying text. The various sources of law other than the Constitution that can give rise to property interests are often collectively referred to as “positive law” or “nonconstitutional law” and the corresponding approach to identifying property interests as “positivist.” See Jerry L. Mashaw, Due Process in the Administrative State 104–06 (1985); Merrill, supra note 1, at 920.

22 See infra notes 58–108 and accompanying text.

23 See infra notes 109–169 and accompanying text.

24 See infra notes 170–235 and accompanying text.

25 See infra notes 236–283 and accompanying text.
I. TRADITIONAL AND POSITIVIST MODES OF DEFINING PROPERTY

Since the rise of the administrative state, the Supreme Court has struggled to define property in the due process context. The Court first focused on the difference between rights and privileges rather than any distinction between property and liberty. Later, the Court focused on the nature of the property or liberty interest at stake.

A. Defining Property in Traditional Due Process Jurisprudence

Prior to the New Deal, the Court did not devote attention to defining the scope of property or distinguishing it from liberty for due process purposes. By the New Deal era, however, administrative agencies increasingly made decisions previously made by courts, leading to challenges to their decision making. In an attempt to define the scope of due process protection against government action, the


[27] See infra notes 29–42 and accompanying text.

[28] See infra notes 43–57 and accompanying text.


[30] See Rubin, supra note 26, at 1048–49 (examining the implications of the rise of the administrative state for procedural due process doctrine). For example, by the 1920s, congressional legislation fixed transportation industry rates that private contract had previously determined and common law enforced. See Ariz. Grocery Co. v. Atchison, T. & S. F. Ry. Co., 284 U.S. 370, 384 (1932) (noting that the Interstate Commerce Act of 1887 altered the common law by allowing an administrative agency to determine the reasonableness of transportation rates rather than the courts).
Supreme Court applied a conceptual distinction between “rights,” which were protected by due process, and “privileges,” which were not. The common law was the source of this distinction. Because government benefits typically fell outside the common law scope of rights, they were merely privileges, unprotected by due process, that could be curtailed or abolished by the government.

The distinction between rights and privileges led to strange results; for instance, selling ice was recognized as a fundamental right not to be deprived without due process, but government employment was not. The right-privilege doctrine also seemed ill-equipped to handle arbitrary government decision making that fell outside the scope of the common law, in particular the denial of government benefits to persons suspected of subversive Communist activity in the 1940s and 1950s. Commentators lambasted the right-privilege distinction as circular in logic and unfair in practice. By the late 1960s, the Court indicated that it would discard the distinction in favor of a

31 See generally William Van Alstyne, The Demise of the Right-Privilege Distinction in Constitutional Law, 81 Harv. L. Rev. 1439 (1968) (tracing the exceptions to the right-privilege distinction and arguing for a rejection of the theory). This distinction had been famously set out by Justice Holmes in a Massachusetts case involving a policeman who challenged being fired for political activities: “[t]he petitioners may have a constitutional right to talk politics, but he has no constitutional right to be a policeman.” McAuliffe v. Mayor of New Bedford, 29 N.E. 517, 517 (Mass. 1892).

32 See Van Alstyne, supra note 31, at 1455.

33 See Monaghan, supra note 18, at 407.

34 Compare New State Ice Co. v. Liebman, 285 U.S. 262, 278 (1932) (holding selling of ice to be a right that is encompassed within the liberty interest of engaging in a lawful business and, therefore, cannot be deprived without due process), with Bailey v. Richardson, 182 F.2d 46, 57 (D.C. Cir. 1950), aff’d by an equally divided Court, 341 U.S. 918 (1951) (per curiam) (determining that due process did not apply to dismissals from federal civil service).

35 See Barsky v. Bd. of Regents, 347 U.S. 442, 451 (1954) (upholding suspension of doctor’s license after he failed to produce certain papers for the House Committee on Un-American Activities, on the grounds that a license to practice medicine is a privilege). For the Court’s application of the right-privilege distinction to the loyalty-security programs, see Rubin, supra note 26, at 1053–60.

36 See, e.g., Kenneth Culp Davis, The Requirement of a Trial-Type Hearing, 70 Harv. L. Rev. 193, 267 (1956) (noting the faulty logic of the right-privilege distinction as applied to business licenses); Charles A. Reich, The New Property, 73 Yale L.J. 733, 787 (1964) (arguing that denying due process protection for government benefits based on a “privilege” or “gratuity” concept is a rubber stamp for arbitrary government action); Van Alstyne, supra note 31, at 1458–64 (noting the erratic reach of the right-privilege doctrine and urging its demise); Stanley C. Beyer, Note, Due Process Requires That the Holder of a Cigarette Permit Be Given Notice and a Hearing Prior to Forfeiture of the Permit, 44 Tex. L. Rev. 1360, 1361 (1966) (observing that an activity might be labeled a “right” in one state and a “privilege” in another).
new approach that would allow for greater flexibility in defining the scope of due process protections.37

This new approach appeared in 1970 in Goldberg v. Kelly, where the Supreme Court held that welfare benefits deserved due process protection prior to termination.38 Considered the opening salvo in what became known as the “Due Process Revolution,” the decision tore down the last vestiges of the right-privilege distinction.39 In its stead, the Court in Goldberg stated that due process protections extended to adjudications of important rights or benefits the deprivation of which imposed a “grievous loss.”40 This approach overlooked the threshold question of whether an interest fit within the Due Process Clause’s definition of “liberty” or “property.”41 Predictably, the Court’s analysis resulted in a dramatic expansion of the scope of procedural due process.42

B. Property as Entitlements Rooted in Positive Law: Board of Regents of State Colleges v. Roth

Just two years after Goldberg, however, in 1972 in Board of Regents of State Colleges v. Roth, the Court underwent a further, and more lasting, transition in its attempt to define liberty and property.43 The plaintiff,

40 397 U.S. at 262–63 (quoting Joint Fascist Refugee Comm’n v. McGrath, 341 U.S. 123, 168 (1951) (Frankfurter, J., concurring)); see also Bell v. Burson, 402 U.S. 535, 539 (1971) (holding that a driving license “may become essential in the pursuit of a livelihood” and that its suspension “adjudicates important interests of the licensees”).
41 See Monaghan, supra note 18, at 407–08 (examining the approach to liberty and property due process in Goldberg and other cases). The Goldberg Court only addressed this issue obliquely in a footnote: “[i]t may be realistic today to regard welfare entitlements as more like ‘property’ than a ‘gratuity.’” 397 U.S. at 262 n.8.
43 See Bd. of Regents of State Colls. v. Roth, 408 U.S. 564, 576–77 (1972). The phrase “Due Process Revolution” is generally applied only to the brief period from 1970 to 1972, comprised of five Supreme Court cases. Pierce, supra note 26, at 1973 (referring to Roth, 408 U.S. 564; Morrissey v. Brewer, 408 U.S. 471 (1972); Bell, 402 U.S. 535; Wisconsin v. Constantineau, 400 U.S. 433 (1971); and Goldberg, 397 U.S. 254). Although the prevailing view is that Goldberg opened the floodgates for procedural due process claims by unmooring property from its common-law roots and placing it in the inchoate realm of weighing the importance of an interest, the degree to which Goldberg actually marked a theoretical
an untenured professor at a state university, was not rehired after his first year of teaching. 44 Although state law provided for notice and a hearing before termination of tenured professors, university officials had unfettered discretion to fire untenured professors. 45 Taking advantage of the Court’s lenient standard for due process, the plaintiff filed suit, arguing that his procedural due process rights to liberty and property had been violated. 46

In a decision that would set the standards for defining liberty and property under the Due Process Clause, the Court took the opportunity both to rein in the breadth of Goldberg’s important rights/grievous loss definition and to legitimize the Court’s efforts to expand due process beyond traditional notions prescribed by the right-privilege distinction. 47 The Court discarded Goldberg’s one-step inquiry that looked to the importance of the claimant’s interest in determining whether due process applied, in favor of an examination of the nature of the property or liberty interest at stake. 48 Roth stated that courts must first ask whether the interest fits the definition of liberty or property before determining whether the procedures governing the interest violate due process. 49

In defining property under the Due Process Clause, the Court did not rely on a definition rooted in traditional property concepts. 50 Rather, the Court explicitly recognized that property interests take many forms beyond actual ownership of real estate or goods and, in so

---

44 Roth, 408 U.S. at 566.
45 Id. at 566–67.
46 See id. at 568.
47 See id. at 567, 570–71 (stating that the Court “must look not to the ‘weight’ but to the nature of the interest at stake” to identify property interests, but upholding Goldberg because property interests “may take many forms”); Goldberg, 397 U.S. at 262–63; see also Merrill, supra note 1, at 918–19 (contending that the Roth Court’s strategy was to make Goldberg’s due process revolution more “law-like”).
48 Roth, 408 U.S. at 570–71. For further analysis of Goldberg’s approach to procedural due process, see Shapiro & Levy, supra note 1, at 124–25 (noting that the Court’s due process jurisprudence, up until and including Goldberg, “incorporated elastic concepts such as basic standards of ‘decency and fairness’”).
49 408 U.S. at 571 (stating that “[w]e must look to see if the interest is within the Fourteenth Amendment’s protection of liberty and property”).
50 See id. at 571–72.
doing, upheld Goldberg and the “New Property” cases. On the other hand, the Court stated that property interests are not infinite. To constitute a property interest, a person must have a “legitimate claim of entitlement” to the benefit, as opposed to an abstract desire or unilateral expectation. Although the Court did not define “entitlement,” the reliance interest was central to the concept because property was designed to protect claims upon which people rely in their daily lives. The Court expounded on the source of entitlements, however, in a passage that would be quoted often in years to come:

Property interests, of course, are not created by the Constitution. Rather they are created and their dimensions are defined by existing rules or understandings that stem from an independent source such as state law—rules or understandings that secure certain benefits and that support claims of entitlement to those benefits.

Henceforth, the Court would define property interests under the Due Process Clause as entitlements rooted in positive law. It was up to subsequent jurisprudence to fill in the details.

--

[51] Id. at 571–72, 576. “New Property” refers to those Roth-type property interests in government benefits that extend beyond traditional property interests in real estate, money, and chattels. See Reich, supra note 36, at 733.

[52] Roth, 408 U.S. at 572.

[53] Id. at 577. Even though Roth expressly discards the right-privilege distinction, several commentators have noted that the entitlement concept in Roth has a whiff of the old right-privilege dichotomy because entitlements can be seen as “rights” and abstract desires as merely “privileges.” See Karen H. Flax, Liberty, Property, and the Burger Court: The Entitlement Doctrine in Transition, 60 Tul. L. Rev. 889, 902–03 (1986); Merrill, supra note 1, at 921–22; Rubin, supra note 26, at 1067; Smolla, supra note 26, at 75–82.

[54] See Roth, 408 U.S. at 577.

[55] Id.


II. Defining Property and Liberty After Roth

In defining liberty and property for procedural due process purposes, Board of Regents of State Colleges v. Roth left three significant issues unresolved: (1) whether the procedures laid out in the relevant positive law can delimit liberty or property interests, (2) what interpretive method courts should use to determine whether sources of law other than the Constitution give rise to liberty or property interests, and (3) whether courts must recognize due process interests seemingly recognized in state law that are not in accord with other policy goals.  

With respect to the first issue, the U.S. Supreme Court later definitively stated that procedures cannot define the scope of a substantive interest because they, too, are subject to constitutional scrutiny. Concerning the second issue of interpretation, the Court developed a methodology in the prisoners’ due process area that has been largely retained for identifying property interests, although later discarded for identifying liberty interests. The recent decision in...
Town of Castle Rock v. Gonzales, however, demonstrates that the Court is still wrestling with the third issue.61

A. Positive Law Procedures Limiting Property Interests

The Court first addressed the issue of whether procedures could define the scope of a property interest two years after Roth in Arnett v. Kennedy in 1974.62 Writing for a three-judge plurality, Justice Rehnquist determined that a civil service employee who was fired without an evidentiary hearing had a property interest for due process purposes in employment because the governing civil service statute prohibited dismissal except “for cause.”63 Justice Rehnquist added, however, that the procedures established for removing the employee defined the limits of the employee’s interest and thus the limits of his constitutional protection, though they did not include a hearing.64 In Justice Rehnquist’s estimation, the employee could not obtain the benefits of the statute without also being subject to its limitations; the employee “must take the bitter with the sweet.”65

Although this logic seemed to be an extension of Roth’s statement that property interests are created and defined by positive law, its potential breadth alarmed commentators.66 Taken to its logical extreme, the “bitter with the sweet” reasoning meant that legislators and regulatory agencies could dictate any procedures for termination of property interests without being subject to constitutional review.67 Recognizing this potential conundrum, the Court in 1985 in Cleveland Board of Education v. Loudermill definitively rejected this reasoning and held

---

61 See Castle Rock, 125 S. Ct. at 2809 (implying that an entitlement stemming from state law does not necessarily constitute property for Fourteenth Amendment purposes); infra notes 170–235 and accompanying text.
62 See Arnett, 416 U.S. at 153–54.
63 Id. at 154.
64 Id.
65 Id.
66 See id.; Bd. of Regents of State Colls. v. Roth, 408 U.S. 564, 577 (1972) (stating that property interests are not created by the Constitution); Jerry L. Mashaw, Administrative Due Process: The Quest for a Dignitary Theory, 61 B.U. L. Rev. 885, 893 (1981); Monaghan, supra note 18, at 441–42; Rubin, supra note 26, at 1070–72; Tribe, supra note 26, at 277–78.
67 See Rubin, supra note 26, at 1091. Professor Mashaw provides the example of a school teacher fired without a hearing whose interest in employment relies exclusively on the School Board’s rules under the Arnett plurality’s reasoning. See Mashaw, supra note 66, at 893.
that procedural rules are irrelevant for purposes of determining the scope of a property interest.\textsuperscript{68}

\textbf{B. Mandatory Statutes and Prisoners’ Rights— Identifying Interests in Positive Law}

The Supreme Court has been more concerned with defining the source of property interests for due process purposes than with statutory interpretation.\textsuperscript{69} In the prisoners’ liberty interest area, however, the Court has further revealed how it identifies a procedural due process interest in sources of law other than the Constitution.\textsuperscript{70} Initially, the \textit{Roth} Court seemed poised to define liberty for procedural due process purposes as it defined liberty in its substantive sense—as those long-recognized privileges essential to the orderly pursuit of happiness.\textsuperscript{71}

In a case handed down the same day as \textit{Roth}, however, the Court moved towards adopting an approach that would look to state positive law rather than lofty common-law goals to define liberty for procedural due process purposes.\textsuperscript{72} In \textit{Morrissey v. Brewer} in 1972, the Court extended the positive law approach of identifying property interests protected by due process to the concept of liberty, by holding that a prisoner’s conditional freedom based on a state-authorized parole program could not be revoked without a hearing.\textsuperscript{73} This marked the first time that a liberty interest arose from an expectation grounded in statutory law, rather than constitutional or common law.\textsuperscript{74} Two years later in 1974, in \textit{Wolff v. McDonnell}, the Court expressly adopted

\begin{itemize}
\item \textsuperscript{68} 470 U.S. at 541 (writing that “it is settled that the ‘bitter with the sweet’ approach misconceives the constitutional guarantee” (quoting \textit{Arnett}, 416 U.S. at 167)).
\item \textsuperscript{69} See \textit{Loudermill}, 470 U.S. at 541 (discarding the bitter-with-the-sweet approach to defining substance and procedure).
\item \textsuperscript{71} See \textit{Roth}, 408 U.S. at 572 (quoting \textit{Meyer v. Nebraska}, 262 U.S. 390, 399 (1923)).
\item \textsuperscript{72} See \textit{Morrissey v. Brewer}, 408 U.S. 471, 481–82 (1972).
\item \textsuperscript{73} \textit{Id.} at 482.
\item \textsuperscript{74} See \textit{Laurence H. Tribe, American Constitutional Law} § 10-9, at 686 (2d ed. 1988). The positivist approach to identifying liberty rights in nonconstitutional law has been generally limited to prisoners’ rights cases because most other infringements of liberty occur in the arrest context and are allowable if there is a lawful conviction. \textit{See Herman, supra note 70, at 503}. 
\end{itemize}
the positivist approach for identifying liberty interests by concluding that credits for prisoners for good behavior guaranteed under a state statute could not be revoked without due process.\textsuperscript{75} Although the Court noted that the Constitution does not guarantee so-called “good-time” credits, the liberty interest protected by the Fourteenth Amendment was expansive enough “to insure that the state-created right is not arbitrarily abrogated.”\textsuperscript{76}

In an attempt to curb the potential breadth of liberty interests identified under state law, in 1976, in \textit{Meachum v. Fano}, the Court took a more restrictive approach to identifying liberty interests in upholding a prisoner transfer to a substantially less favorable prison without a hearing.\textsuperscript{77} The Court reasoned that the state regulations regarding prison transfers gave prison officials complete discretion whether to transfer prisoners; therefore, the prisoners had no state law entitlement and, in turn, no procedural due process liberty interest.\textsuperscript{78} In distinguishing this case from \textit{Wolff}, the Court focused on whether the statute was mandatory or discretionary, not on whether the prisoner’s interest was seriously affected.\textsuperscript{79}

This method of identifying an interest in state law was clarified in 1989 in \textit{Kentucky Department of Corrections v. Thompson}, where the Court found that state prison inmate visitation regulations did not give rise to a protected liberty interest in receiving certain visitors.\textsuperscript{80} The Court reasoned that the regulations did not contain explicitly mandatory language stating that if a condition has been met, a particular outcome must follow, a requirement for any due process liberty interest.\textsuperscript{81}

Nonetheless, the flood of prisoner litigation into federal courts continued unabated.\textsuperscript{82} According to the Court itself, its prior decisions spurred prisoners and their lawyers to search for mandatory

\begin{itemize}
  \item \textsuperscript{75} 418 U.S. 539, 554–55 (1974).
  \item \textsuperscript{76} Id. at 557.
  \item \textsuperscript{77} See 427 U.S. at 223–24.
  \item \textsuperscript{78} Id. at 226–29.
  \item \textsuperscript{79} See id. at 226; Rubin, supra note 26, at 1076.
  \item \textsuperscript{80} See 490 U.S. 454, 463 (1989).
  \item \textsuperscript{81} Id. The Court in \textit{Castle Rock} also recognized that the statutory language must be mandatory to create property interests. 125 S. Ct. at 2803 (stating that “[o]ur cases recognize that a benefit is not a protected entitlement if government officials may grant or deny it in their discretion”) (citing \textit{Ky. Dep’t of Corr.}, 490 U.S. at 462–63 (1989)).
\end{itemize}
language in prison regulations, rather than to frame their cases in terms of fundamental notions of liberty. In 1995, the Court responded in \textit{Sandin v. Conner} by discarding the positive law approach of identifying liberty interests in favor of a “grievous loss” approach that would examine whether the alleged deprivation of liberty represents an “atypical and significant hardship” in relation to ordinary prison life. The Court held that an inmate who, after a hearing, had been put in disciplinary segregation, during which he was not allowed to call witnesses, did not have a cognizable liberty interest. Instead of concluding that the state regulations failed to satisfy the mandatory language requirement, the Supreme Court criticized prior cases for mechanically applying a rigid mandatory-discretionary dichotomy and for neglecting the “nature” of the interest created by state law. The Court noted that, although drawing negative inferences from the mandatory language of prison legislation may be useful for most statutes defining rights for the general public, it makes little sense in examining prison administration regulations. Because such inference-drawing involved federal courts in day-to-day prison management, the Court felt compelled to abandon the parsing of statutory language in favor of its grievous loss/atypical hardship test. \textit{Sandin} appears to have reduced the amount of prisoner litigation in federal courts. The “atypical and significant hardship” approach of identifying liberty interests for prisoners, however, trades a reduction in prisoner claims for a lack of uniformity, as lower courts have struggled to adopt a coherent strategy to define this new standard. In addition, applying \textit{Sandin}’s reasoning beyond the realm of prison

\begin{enumerate}
\item See \textit{Sandin}, 515 U.S. at 481.
\item See \textit{id.} at 484.
\item \textit{Id.}
\item See \textit{id.} at 479–80.
\item \textit{Id.} at 481–82.
\item \textit{Sandin}, 515 U.S. at 482.
\item See \textit{Bureau of Justice Statistics Special Report, supra} note 82, at 1.
\item Compare \textit{Torres} v. Fauver, 292 F.3d 141, 151–52 (3d Cir. 2002) (holding that the punishment of a prisoner found guilty on escape charges solely based on his statement that he would escape if not granted housing placement violates the atypical hardship standard), with \textit{Burnsworth} v. Gunderson, 179 F.3d 771, 774–75 (9th Cir. 1999) (holding that the punishment of a prisoner found guilty on escape charges solely based on his statement that he would escape if not granted housing placement does not violate the atypical hardship standard). See generally Lee, \textit{supra} note 70 (tracing the uneven application of \textit{Sandin}’s standard); Michael Z. Goldman, Note, \textit{Sandin v. Conner} and \textit{Intraprison Confinement: Ten Years of Confusion and Harm in Prisoner Litigation}, 45 B.C. L. REV. 423 (2004) (arguing that the atypical hardship standard has led to harsh results and has led lower courts to ignore confinement that is not representative of the prison experience).
\end{enumerate}
regulation to property interests for due process purposes is suspect, both because of the unique concerns in prison regulation and because rampant litigation is not necessarily a problem with respect to property after Roth. Nevertheless, in Castle Rock, the Court cited Sandin for part of its analysis and seems to be moving away from Roth’s positive law approach to identifying property interests.

C. Recognizing Due Process Interests from State Law Not in Accord with Other Policy Goals

Although the Court has decided two of the open issues from Roth, it has left unclear whether it is obliged to follow the dictates of state law when it seems to recognize too much or too little property. Read strictly, Roth did not seem to leave room for an independent constitutional review when it stated that “[p]roperty interests, of course, are not created by the Constitution.” In 1978, however, in Memphis Light, Gas & Water Division v. Craft, the Court suggested otherwise by holding that a utility company procedure of cutting off services for nonpayment in the midst of a payment dispute did not comport with due process. The Court found that a requirement rooted in state common law and federal constitutional law barring a public utility from termination of service except for good cause created a property interest in continued utility service. The Court stated that, although sources of law other than the Constitution create the substantive law, “federal constitutional law determines whether that interest rises to the level of a ‘legitimate claim of entitlement.’”

91 Merrill, supra note 1, at 966 (stating that “there is no sign of increasing numbers of cases raising new property claims”). See generally Herman, supra note 70 (noting that liberty is distinct from property and should not be defined by looking to positive law because the state does not create liberty).
92 See Castle Rock, 125 S. Ct. at 2808.
93 See Merrill, supra note 1, at 933 (stating that the Court has wrestled with how to avoid capturing too much or too little property for due process purposes). Professor Mashaw described this dilemma as a “positivist trap”—once the Court seeks to define constitutional interests in positive law, it must address whether those rules suffice in and of themselves or must comport with other constitutional values. See Mashaw, supra note 66, at 888.
94 See Roth, 408 U.S. at 577.
95 See 436 U.S. 1, 10 (1978).
96 Id.
97 Id. at 9 (quoting Roth, 408 U.S. at 577, and Perry v. Sindermann, 408 U.S. 593, 602 (1972)). Professor Merrill has argued that this two-tiered constitutional review represents a departure from prior procedural due process jurisprudence. See Merrill, supra note 1, at 927.
Thus, at first glance *Memphis Light* appeared to establish a two-tiered review for recognizing property interests under the Due Process Clause; on one level, there is a state law examination, and on the other, a federal constitutional one.\(^\text{98}\) The Court’s constitutional review, however, resulted in the unexceptional determination that state common law precedent did not permit public utilities to terminate service if the customer was withholding payment because of a bona fide payment dispute, even though the utility company’s procedures required payment regardless.\(^\text{99}\) Thus, the constitutional review in *Memphis Light* was almost entirely grounded in an examination of a source of law independent from the Constitution—in this case, Tennessee common law precedents that clearly contradicted utility company regulations.\(^\text{100}\) More importantly, the Court did not erect a bar for constitutional recognition of a property interest, but rather ensured that constitutional norms would apply to a public utility procedure that would not otherwise recognize a property interest in continued utility service.\(^\text{101}\) In this way, the Court’s constitutional analysis stood for the proposition that the source of law at issue—in this case, discretionary utility company regulations—did not determine the limits of a property interest for due process purposes if they were in conflict with other state law or the Constitution.\(^\text{102}\)

In 1982, in *Logan v. Zimmerman Brush Co.*, the Court reaffirmed *Memphis Light*’s limited constitutional review for recognizing property interests by holding that an employee’s right to adjudication procedures was a type of property protected by due process.\(^\text{103}\) The Court held that the hallmark of constitutional property is an entitlement grounded in state law that cannot be removed except “for cause.”\(^\text{104}\) Beyond the general requirement that for a state law benefit to create a property interest it must have specific conditions governing its termination, *Memphis Light* and *Logan* do not indicate that the Court

\(^\text{98}\) See Merrill, supra note 1, at 926–27; Rubin, supra note 26, at 1079 n.193, 1091.
\(^\text{99}\) See *Memphis Light*, 436 U.S. at 10–11.
\(^\text{100}\) See id. The Court’s only citations to decisions outside of Tennessee were to Supreme Court decisions buttressing the Court’s conclusion. See id.
\(^\text{101}\) See id. at 11.
\(^\text{102}\) See id. Additionally, because *Memphis Light* was decided before the Court had definitively settled the previously discussed “bitter with the sweet” controversy from *Arnett*, the case could also point towards a retreat from the proposition that procedures can dictate whether a property entitlement exists. See *Loudermill*, 470 U.S. at 541 (writing that “it is settled that the ‘bitter with the sweet’ approach misconstrues the constitutional guarantee”).
\(^\text{103}\) 455 U.S. 422, 430 (1982).
\(^\text{104}\) Id.
should turn to additional independent criteria in defining property under the Due Process Clause.\textsuperscript{105} No clear criteria to which the Court could turn existed because, prior to \textit{Roth}, due process protections were defined by either the discredited right-privilege distinction or \textit{Goldberg’s} important interest approach.\textsuperscript{106} Perhaps for this reason, the language of \textit{Memphis Light} was not used in subsequent cases until \textit{Castle Rock}, and was interpreted narrowly by some scholars.\textsuperscript{107} As explained below, however, the \textit{Castle Rock} dicta appears to utilize this language in crafting additional criteria for state entitlements to rise to the level of constitutional property interests and, in the process, undercuts \textit{Roth’s} clearly stated positivist approach.\textsuperscript{108}

III. \textit{Castle Rock v. Gonzales: Redefining Due Process Property}

The Supreme Court’s 2005 decision in \textit{Town of Castle Rock v. Gonzales} stands as a significant milestone in procedural due process jurisprudence.\textsuperscript{109} The facts of the case are “undeniably tragic.”\textsuperscript{110} According to Jessica Gonzales’s complaint, she obtained a restraining order on May 21, 1999 against her husband in connection with divorce proceedings.\textsuperscript{111} The order commanded Mr. Gonzales not to disturb or molest his three children and to remain at least 100 yards from the family home at all times, except during certain pre-arranged visits.\textsuperscript{112} The restraining order also contained a preprinted notice to law enforcement officials: “You \textbf{shall} use every reasonable means to

\textsuperscript{105} See id.; \textit{Memphis Light}, 436 U.S. at 10–11.
\textsuperscript{106} See supra notes 26–42 and accompanying text.
\textsuperscript{107} See \textit{Mashaw}, supra note 21, at 108 (failing to mention this two-tiered approach to identification of property interests in discussing \textit{Memphis Light}). But see Rubin, supra note 26, at 1079 n.193 (stating that \textit{Memphis Light} represents the Court’s current approach to identifying property interests). The one exception to \textit{Memphis Light’s} lack of citation, prior to \textit{Castle Rock}, was O’Bannon v. Town Court Nursing Ctr., 447 U.S. 773, 788 (1980). Professor Merrill argues that \textit{Memphis Light} was not cited because it involved a well-settled question of law and did not directly implicate the “bitter-with-the-sweet” debate which attracted most scholars’ attention in this area. See Merrill, supra note 1, at 926–29.
\textsuperscript{108} See infra notes 198–235 and accompanying text.
\textsuperscript{109} See Roederer, supra note 7, at 360–63 (contending that \textit{Castle Rock} is on par with \textit{Lochner} in its efforts to undermine state attempts to alter the common-law distribution of wealth and entitlements); \textit{2004 Term—Leading Cases}, supra note 7, at 208. See generally \textit{Town of Castle Rock v. Gonzales}, 125 S. Ct. 2796 (2005).
\textsuperscript{110} 125 S. Ct. at 2803 (quoting DeShaney v. Winnebago County Dep’t of Soc. Servs., 489 U.S. 189, 191 (1989)).
\textsuperscript{111} \textit{Id.} at 2800. Because the district court dismissed Ms. Gonzales’s complaint for failure to state a claim for which relief could be granted, the Supreme Court assumed the facts as presented in her complaint were true. \textit{Id.}
\textsuperscript{112} \textit{Id.} at 2800–01.
enforce the restraining order. You shall arrest, or, if an arrest would be impractical under the circumstances, seek a warrant for the arrest of the restrained person.”  

On June 22, 1999, a day for which he had made no visitation arrangements, Mr. Gonzales abducted his three daughters while they were playing outside the family home. When Ms. Gonzales noticed that the children were missing, she suspected that her husband had taken them, and so she called the police department around 7:30 p.m. Over the next several hours, Ms. Gonzales repeatedly contacted the police station both in person and over the phone to request enforcement of the restraining order. At one point, Ms. Gonzales talked to her husband on his cellular telephone; when he said that he was at an amusement park in Denver with the three children, she asked the police to check on him or put out an all-points bulletin for him. At every request, the police rebuffed Ms. Gonzales and asked her to call again later. Around 3:20 a.m., Mr. Gonzales arrived at the police station and opened fire with a handgun he had purchased that night. Police shot back and killed him. Inside his truck, they found the bodies of all three of his daughters, whom he had murdered earlier that day.

Ms. Gonzales brought an action under 42 U.S.C. § 1983 against three police officers and the town of Castle Rock, alleging they had violated her due process rights under the Fourteenth Amendment.

---

113 Id. at 2801 (emphasis added).
114 Id. at 2805. The relevant statute in effect at the time of the conduct reads as follows:

(a) Whenever a restraining order is issued, the protected person shall be provided with a copy of such order. A peace officer shall use every reasonable means to enforce a restraining order.

(b) A peace officer shall arrest, or, if an arrest would be impractical under the circumstances, seek a warrant for the arrest of a restrained person . . . .

COLO. REV. STAT. § 18-6-803.5(3) (1999).
115 Castle Rock, 125 S. Ct. at 2801.
116 Id.
117 Id.
118 Id. at 2801–02.
119 Id.
120 Castle Rock, 125 S. Ct. at 2802.
121 Id.
122 Id.
123 Id.
claim for which relief could be granted. On appeal, a three-judge panel for the U.S. Court of Appeals for the Tenth Circuit affirmed the rejection of the substantive due process claim. The Tenth Circuit cited DeShaney v. Winnebago County Department of Social Services, a 1989 U.S. Supreme Court case holding that the state had no constitutional duty to protect an individual from third-party harm under substantive due process. DeShaney, however, had explicitly left open the question whether a statute requiring protection could give rise to an entitlement to protection, and hence a property interest, for procedural due process purposes. The three-judge panel held that the Colorado restraining order’s mandatory enforcement terms and limitation to specific protected persons created a procedural due process property interest under the Fourteenth Amendment.

Upon further review, a divided Tenth Circuit en banc panel affirmed. The majority opinion stated that the legislative history and unequivocally mandatory terms of the statute creating the restraining order, coupled with the mandatory enforcement terms on the court-issued restraining order itself, created a property interest in its enforcement. The court made clear that the mandatory language of a law enforcement statutory provision standing alone could

---

124 Id.


126 Gonzales v. City of Castle Rock, 307 F.3d 1258, 1264 (10th Cir. 2002), aff’d en banc, 366 F.3d 1093 (10th Cir. 2004), rev’d sub nom. Town of Castle Rock v. Gonzales, 125 S. Ct. 2796 (2005). There are two exceptions recognized by some lower courts to the bar imposed by DeShaney on substantive due process claims against the state for failing to protect against third-party violence. Id. at 1262. The first is where there is a special relationship between the state and the victim, such as in a foster home. Id. The second is where police action or inaction either creates or significantly increases the danger to the individual. Id. at 1262–63. The panel held that neither of these exceptions applied in Ms. Gonzales’s case. Id. at 1263.

127 See DeShaney, 489 U.S. at 195; Castle Rock, 307 F.3d at 1262.

128 489 U.S. at 195 n.2.

129 Castle Rock, 307 F.3d at 1263, 1266.

130 Castle Rock, 366 F.3d at 1106–08.

131 Id.
not give rise to a constitutionally protected property interest. The court concluded that, because the town of Castle Rock had provided inadequate safeguards for enforcing restraining orders, Ms. Gonzales was entitled to relief. There were four dissenting opinions. Judges Kelly, O’Brien, and Hartz argued that enforcement of the restraining order did not rise to the level of a property interest because, on the whole, enforcement was not truly mandatory. Judge McConnell contended that Ms. Gonzales’s claim was a substantive due process claim masquerading as a procedural due process claim that should, therefore, be rejected under DeShaney.

The Supreme Court granted certiorari and reversed. The Court rejected Ms. Gonzales’s claim that enforcement of the restraining order created a property interest under the Due Process Clause. The Court held that the restraining order was not an entitlement created under state law because it was not truly mandatory, given the long history of police discretion for enforcement of apparently mandatory statutes. Although the restraining order used the term “shall,” the Court wrote, this was hortatory language used in many general protection statutes; a “true mandate of police action would require some stronger indication from the Colorado Legislature.”

The Court also dismissed the reasoning that the legislative history of the restraining order statutes, which suggested that the drafters’ intent was to make enforcement truly mandatory, required a contrary holding because a mandatory arrest could only be made if the offender were present. Furthermore, the Court stated that, even if the restraining order made enforcement mandatory for the police, it did not necessarily mean that Ms. Gonzales herself had an entitlement...
for due process purposes to its enforcement.\textsuperscript{142} Citing \textit{Sandin v. Conner}, the Court noted that making actions mandatory could serve a variety of ends other than the conferral of a benefit on a specific class of people.\textsuperscript{143} Moreover, because Ms. Gonzales’s claim rested on a statutory, and not a common-law or contractual entitlement to enforcement, the Court expected to see a conferral of a benefit on a specific class of persons in the statute itself.\textsuperscript{144} Thus, the Court concluded that Ms. Gonzales did not have an entitlement to “something as vague and novel as enforcement of restraining orders.”\textsuperscript{145}

The Court could have ended there, but went on to state that even if the restraining order granted Ms. Gonzales an entitlement under Colorado law, it was “by no means clear” that this could constitute a property interest under the Due Process Clause.\textsuperscript{146} The Court’s reasoning rested on two points: (1) the restraining order did not have an “ascertainable monetary value,” that the Court’s “\textit{Roth}-type property-as-entitlement” cases implicitly required; and (2) the alleged property interest arises only “incidentally” out of a government function that government actors have always performed.\textsuperscript{147}

To support the first of these points the Court cited a law review article that proposed the monetary value test as a means of separating property from liberty interests.\textsuperscript{148} For the second point, the Court cited \textit{O’Bannon v. Town Court Nursing Center}, where the Court held that indirect benefits conferred on Medicaid patients did not trigger due process protections.\textsuperscript{149} The Court concluded its opinion in \textit{Castle Rock} by stating that protection against third-party violence generally does not trigger either substantive or procedural due process protections—language that seemed designed to foreclose any person’s entitlement to enforcement of protective measures against third parties under any state law.\textsuperscript{150} The Court left the recovery mechanism for police non-

\textsuperscript{142} \textit{Castle Rock}, 125 S. Ct. at 2808.
\textsuperscript{143} Id. The \textit{Sandin} Court had found no constitutionally protected liberty interest in prison regulations phrased in mandatory terms because, among other reasons, they are not set forth solely to benefit the prisoner. 515 U.S. 472, 482 (1995).
\textsuperscript{144} \textit{Castle Rock}, 125 S. Ct. at 2808.
\textsuperscript{145} Id. at 2809.
\textsuperscript{146} \textit{See id.}
\textsuperscript{147} Id.
\textsuperscript{148} Id. (citing Merrill, \textit{supra} note 1, at 964).
\textsuperscript{149} \textit{Castle Rock}, 125 S. Ct. at 2810 (citing \textit{O’Bannon v. Town Court Nursing Ctr.}, 447 U.S. 773, 775 (1980)).
\textsuperscript{150} \textit{Castle Rock}, 125 S. Ct. at 2810 (writing that “[i]n light of today’s decision and that in \textit{DeShaney}, the benefit that a third party may receive from having someone else arrested for
enforcement up to the states, stating that “the people of Colorado are free to craft such a system under state law.”

Justice Souter filed a brief concurring opinion which Justice Breyer joined. He wrote separately to stress that Ms. Gonzales’s argument should be rejected because she was claiming a property interest in process itself. Her claim could be reduced to an argument that police officers should follow procedure, but in every case where the Court has recognized a property interest, the underlying property was distinguishable from the procedural obligations sought.

Justice Stevens dissented, in an opinion joined by Justice Ginsburg. The dissent began with a repudiation of the majority’s language suggesting that the Court would rule out all entitlements under any state law. Justice Stevens posited that, although DeShaney meant that Ms. Gonzales was not entitled to police protection as a matter of substantive due process, the Constitution did not bar the recognition of an entitlement grounded in state law. Ms. Gonzales could have entered into a contract with a private security firm, and her interest in such a contract would undoubtedly constitute “property” under the Due Process Clause. Therefore, if Colorado law created the functional equivalent of such a private contract, this state-created right should qualify as property under Roth.

After admonishing the majority for answering a question of Colorado law on its own rather than certifying it to the Colorado Supreme Court, the dissent turned to the mandatory nature of the state restraining order statute and its legislative history. Justice Stevens contended that the Court ignored the “unique case” of mandatory arrest statutes in the domestic violence context passed during the 1980s and 1990s, which had the clear goal of cabining police discretion. Thus,

---

151 Id.
152 Id. at 2811 (Souter, J., concurring).
153 Id. at 2812.
154 Id.
155 Castle Rock, 125 S. Ct. at 2813 (Stevens, J., dissenting).
156 Id.
157 Id.
158 Id.
159 Id.
160 Castle Rock, 125 S. Ct. at 2815–18 (Stevens, J., dissenting).
161 Id. at 2816–17. Against this backdrop, the dissent argued that the apparent novelty of this form of property interest is not as troubling as the majority made it out to be. Id. at 2823.
although the term “shall” might be best read as meaning “may” in other Colorado police enforcement statutes, the particular history of this restraining order statute implied that it should be read literally to limit police discretion. Justice Stevens finally rebutted the majority’s focus on the lack of precise means of enforcement with the point that the police were required either to arrest or to seek a warrant for arrest; “they lacked the discretion to do nothing.”

Justice Stevens further responded to the majority’s dicta that stated that the restraining order had no ascertainable monetary value and only incidentally benefited Ms. Gonzales. The dissent noted that the Court had never previously required that a property interest possess monetary value. In fact, property interests recognized by the Court as requiring due process protections had taken a variety of forms, “as often as not, intangible” ones, such as public education. Nevertheless, if Ms. Gonzales had contracted with a private security firm, her interest in enforcement of the contract would have an underlying value. As for whether the decision in O’Bannon regarding indirect benefits was fatal to Ms. Gonzales’s complaint, Justice Stevens stated that the O’Bannon opinion, which he authored, did not actually address a situation where the underlying law created an entitlement and was thus inapplicable. The dissent concluded its discussion with a rebuke to Justice Souter’s reasoning that Ms. Gonzales merely sought enforcement of process, not substance, by remarking that she asserted an interest in enforcement of a restraining order, a “tangible, substantive act.”

IV. Analysis of Castle Rock’s Definition of Property

Judge Kelly of the U.S. Court of Appeals for the Tenth Circuit began his dissent in Gonzales v. City of Castle Rock by stating that “[t]he facts of this case give new meaning to the old adage that hard cases make bad law.” This admonishment could equally apply to the deci-
sion of the U.S. Supreme Court in 2005 in *Town of Castle Rock v. Gonzales*.\(^{171}\) The Court’s analysis of state law in holding that a restraining order does not grant its holder a property interest in its enforcement, although perhaps not surprising given the Court’s holding in *DeShaney v. Winnebago County Department of Social Services* that there is no substantive right to government protection from third-party harm, departs significantly from the framework for identifying property interests for due process purposes set out by the Court in *Board of Regents of State Colleges v. Roth*.\(^{172}\) The Court’s decision in *Castle Rock* reveals a cramped view of property interests under the Due Process Clause that is not in accord with the Court’s prior case law or the constitutional aims implicated by due process.\(^{173}\)

A. Castle Rock’s Statutory Analysis

The Court began its analysis of Ms. Gonzales’s claim in *Castle Rock* with an examination of whether enforcement of the Colorado restraining order was mandatory.\(^{174}\) Drawing on the Court’s liberty interest precedents for the proposition that a benefit cannot rise to the level of an entitlement if officials can grant or deny it in their discretion, the Court parsed the restraining order’s language and context.\(^{175}\) The Court determined that Colorado’s seemingly mandatory restraining order statute was in fact circumscribed by a long tradition of police discretion in enforcement of the laws.\(^{176}\) In support, the Court compared the restraining order statute to other seemingly mandatory statutes regulating police behavior, such as directives or-

\(^{171}\) See generally 125 S. Ct. 2796 (2005).

\(^{172}\) See *DeShaney v. Winnebago County Dep’t of Soc. Servs.*, 489 U.S. 189, 196–97 (1989) (holding that there is no substantive right to government protection against third-party violence). Compare *Castle Rock*, 125 S. Ct. at 2809 (stating that even if Colorado law created an entitlement to enforcement of the restraining order, it could not constitute a property interest for due process purposes because it lacks monetary value and only incidentally benefits the claimant), with *Bd. of Regents of State Colls. v. Roth*, 408 U.S. 564, 577 (1972) (stating that property interests are created and their dimensions are defined by the understandings stemming from independent sources such as state law).

\(^{173}\) See infra notes 174–235 and accompanying text.

\(^{174}\) 125 S. Ct. at 2804–09. In reaching its conclusion, the Court examined the language on the preprinted notice to law enforcement personnel appearing on the back of the restraining order itself: “A peace officer shall use every reasonable means to enforce a restraining order . . . . A peace officer shall arrest, or, if an arrest would be impractical under the circumstances, seek a warrant for the arrest of a restrained person.” *Id.* at 2805 (emphasis omitted).

\(^{175}\) See *id.* at 2803, 2805–06.

\(^{176}\) *Id.* at 2805–06.
dering chiefs of police to pursue and arrest any person fleeing from justice.\textsuperscript{177} Underlying the Court’s analysis was the concern that any statute that limits police discretion could potentially be transformed into a constitutionally protected property interest if the police failed to enforce it.\textsuperscript{178} The majority also noted that the precise enforcement mechanism of the restraining order was vague; it was unclear how the restraining order could be enforced if the offender was not present.\textsuperscript{179}

The logic of the majority’s statutory analysis does not stand under scrutiny.\textsuperscript{180} Although the Court relied heavily on the ambiguity of the word “shall” to determine that enforcement of the Colorado restraining order was not truly mandatory, few words have a clearer non-discretionary meaning in statutory interpretation.\textsuperscript{181} In fact, the Colorado legislature amended the state’s restraining order statute to require enforcement precisely because police officers had been failing to enforce more discretionary restraining orders.\textsuperscript{182} In so doing, Colorado joined a wave of states that had amended their restraining order statutes to combat domestic violence.\textsuperscript{183}

Even beyond the text and legislative history, there are other aspects of restraining orders that distinguish them from other general protection laws using hortatory terms to uphold a law enforcement objective.\textsuperscript{184} A restraining order’s terms are court-ordered, unlike most other statutes limiting police discretion, and are also, by their

\begin{itemize}
  \item \textsuperscript{177} Id. at 2806.
  \item \textsuperscript{178} See id.
  \item \textsuperscript{179} Castle Rock, 125 S. Ct. at 2807–08.
  \item \textsuperscript{180} See id. at 2813–25 (Stevens, J., dissenting) (noting inconsistencies in the majority’s statutory analysis).
  \item \textsuperscript{181} See Alabama v. Bozeman, 533 U.S. 146, 153 (2001) (discussing the customary use of “shall” to command a particular action); Mallard v. U.S. Dist. Court, 490 U.S. 296, 302 (1989) (categorizing both “shall” and “must” as command expressions); see also DiMarco v. Dept of Revenue, Motor Vehicle Div., 857 P.2d 1349, 1352 (Colo. Ct. App. 1993) (“The factor which most heavily weighs in favor of a mandatory construction is the use of the word ‘shall’ in the provision at issue. Unless the context indicates otherwise, the word ‘shall’ generally indicates that the General Assembly intended the provision to be mandatory.”) (citation omitted).
  \item \textsuperscript{182} See Castle Rock, 125 S. Ct. at 2816–18 (Stevens, J., dissenting).
  \item \textsuperscript{183} Id. at 2817–18 (noting that when Colorado passed its statute in 1994, it joined nineteen other states that mandated arrests for domestic restraining order violations); see Roe derer, supra note 7, at 322 n.5 (placing the current number of states and territories that have passed mandatory arrest provisions for restraining order violations at thirty-five).
  \item \textsuperscript{184} See Castle Rock, 125 S. Ct. at 2816–20 (Stevens, J., dissenting).
\end{itemize}
nature, for the benefit of a specific class of persons—the “protected persons” identified in the restraining order.\textsuperscript{185}

Additionally, contrary to the majority’s suggestion, the means of enforcement are clear and unambiguous.\textsuperscript{186} The police are required either to arrest or to seek a warrant for arrest if they have probable cause to believe the restraining order has been violated—two specific actions, routinely performed by police, that were neither attempted nor performed in Ms. Gonzales’s case.\textsuperscript{187} Although the concurrence and majority characterize these actions as merely procedures, unconnected to any articulable substantive guarantee, the enforcement of a restraining order is in fact a tangible act.\textsuperscript{188} It involves substantive action by the police and is entirely separate from the procedures that govern its treatment—namely, the police procedures for dealing with restraining order violations.\textsuperscript{189}

The legislative history and text of the restraining order statute make clear that the law guaranteed to a specific class of beneficiaries the provision of a particular service in certain defined circumstances, and Ms. Gonzales reasonably relied on that guarantee.\textsuperscript{190} Given the narrowness of this interest, it is unlikely that its constitutional recognition would federalize a large swath of state process, as the concurrence suggested, or turn the Fourteenth Amendment into a “font of tort law,” as the majority feared.\textsuperscript{191} For example, as the Tenth Circuit’s

\textsuperscript{185} Id. at 2816.
\textsuperscript{186} See id. at 2807–08 (majority opinion).
\textsuperscript{187} See id. at 2820 (Stevens, J., dissenting).
\textsuperscript{188} See id. at 2811 (Souter, J., concurring); id. at 2824 n.20 (Stevens, J., dissenting). Additionally, the Court’s precedents have recognized property interests that could be recast as merely interests in procedure under the concurrence’s analysis—for instance, a cause of action. See Logan v. Zimmerman Brush Co., 455 U.S. 422, 431 (1982) (stating that “the right to use the [statute’s] adjudicatory procedures” is a protected property interest) (emphasis added). Logan is notably absent from the litany of post-\textit{Roth} procedural due process decisions that the concurrence lists as distinguishing between property and procedural obligations. See \textit{Castle Rock}, 125 S. Ct. at 2812 (Souter, J., concurring). As Professor Roederer notes, under the Court’s analysis, almost all government obligations to provide a service—in other words, all of the interests recognized in the “New Property” cases—could arguably be collapsed into an interest in procedure, because all entitlements to services are supplemented by process. Roederer, \textit{supra} note 7, at 354.
\textsuperscript{190} See \textit{Castle Rock}, 125 S. Ct. at 2811 (Souter, J., concurring); id. at 2824 n.20 (Stevens, J., dissenting).
\textsuperscript{191} See id. at 2823 (Stevens, J., dissenting).
\textsuperscript{191} See id. at 2810 (majority opinion); id. at 2812–13 (Souter, J., concurring). From a normative perspective, although the legislature’s choice to require enforcement of restraining orders could raise concerns that the police may not pursue enforcement of other crimes, there is also evidence to suggest that encouraging police discretion discourages police action, a potentially dangerous consequence in the domestic abuse context. See
en banc opinion in *Castle Rock* suggests, a statute that required the fire department to respond to a fire would be unaffected by such a narrow ruling and would not give rise to a property interest; it is not a guarantee of service aimed towards a certain class of beneficiaries, but a promise of service to the general public.\(^{192}\)

The Supreme Court’s analysis of state law in *Castle Rock* also reveals a deep skepticism of non-traditional forms of property in contradiction to *Roth*’s stated terms and the Court’s own precedent.\(^{193}\) Undergirding much of the Court’s opinion is an aversion to recognizing a “novel” or “unconventional” form of property.\(^{194}\) The type of property interest in *Castle Rock* was relatively novel, however, because of the relative novelty of the statutes creating mandatory restraining orders.\(^{195}\) Nevertheless, the majority implied that statutes conferring benefits on a specific class of persons based on certain enforceable criteria are required to have “further indication” that they are intended to create a property interest.\(^{196}\) Instead of explaining what further statutory requirements are necessary, however, the Court constructed additional barriers for the constitutional recognition of a property interest even where the statutory terms creating the interest are mandatory.\(^{197}\)

---

\(^{192}\) See *Castle Rock*, 366 F.3d at 1109 (distinguishing these two scenarios and narrowing its holding to enforcement of restraining orders); see also *Logan*, 455 U.S. at 430 (stating that the hallmark of a property interest is an *individual* entitlement rooted in a source of law) (emphasis added); Michael Mattis, Note, *Protection Orders: A Procedural Pacifier or a Vigorously Enforced Protection Tool? A Discussion of the Tenth Circuit’s Decision in Gonzales v. Castle Rock*, 82 Denv. U. L. Rev. 519, 534–35 (2005) (arguing that the Tenth Circuit’s decision would not mean that a general protection statute would give rise to a property interest). The Tenth Circuit explicitly acknowledged the narrowness of its *Castle Rock* holding in a subsequent case, refusing to recognize a property interest in a general protection statute. See *Jennings*, 383 F.3d at 1206–07 (holding that a state statute requiring officers not to discourage prosecution of a sexual assault claim did not constitute a property interest despite the Tenth Circuit’s decision in *Castle Rock* because the mandatory statute was not coupled with a mandatory court order).

\(^{193}\) See *Logan*, 455 U.S. at 430 (stating that the types of property interests recognized by the Court are varied and often intangible); *Roth*, 408 U.S. at 576 (noting that property interests come in many forms).

\(^{194}\) See *Castle Rock*, 125 S. Ct. at 2809 (stating that an entitlement to “something as vague and novel as enforcement of restraining orders cannot ‘simply g[o] without saying’”) (citation omitted); *id.* at 2811 (Souter, J., concurring) (characterizing Ms. Gonzales’s argument as “unconventional”).

\(^{195}\) See *id.* at 2823 (Stevens, J., dissenting).

\(^{196}\) See *id.* at 2808 (majority opinion).

\(^{197}\) See *id.*
B. Constitutional Property Values in Castle Rock

Perhaps sensing the inadequacy of its statutory analysis, the Court seized upon additional grounds for the failure of Ms. Gonzales’s interest in enforcement of her restraining order to rise to the level of a cognizable property entitlement: the entitlement lacked “ascertainable monetary value” and only incidentally benefited its recipient. The first requirement not only fails by its own terms as it would apply in Castle Rock, but makes little sense in relation to the aims of protecting property under the Due Process Clause. The second requirement is stretched well beyond its previous application by the Court. These additional requirements call into question a number of Court decisions and fundamentally misconstrue the purpose of property under the Due Process Clause, which is to protect a legal relationship based on reasonable reliance on protection from arbitrary termination.

At the outset, the suggestion in Castle Rock that there is a constitutional bar to the recognition of some state-created entitlements as property appears to misunderstand the role of the Court in recognizing property for procedural due process purposes. According to Roth, state-created entitlements are property under the Due Process Clause. The Castle Rock Court’s basis for constructing additional barriers for the constitutional recognition of property interests appears to be Memphis Light, Gas & Water Division v. Craft’s two-tiered approach to identifying property interests for due process purposes. The role for constitutional review in Memphis Light, however, was merely that state-created procedures cannot foreclose constitutional recognition of a

---

198 See id. at 2809–10.
199 See infra notes 209–228 and accompanying text.
200 See infra notes 229–235 and accompanying text.
201 See Daniels v. Williams, 474 U.S. 327, 331 (1986) (stating that “the Due Process Clause . . . was ‘intended to secure the individual from the arbitrary exercise of the powers of government’” (quoting Hurtado v. California, 110 U.S. 516, 527 (1884))); Logan, 455 U.S. at 1155 (writing that “[t]he hallmark of property . . . is an individual entitlement grounded in state law, which cannot be removed except ‘for cause’”) (citation omitted).
202 See Castle Rock, 125 S. Ct. at 2809 (stating that “[e]ven if we were to think otherwise concerning the creation of an entitlement by Colorado, it is by no means clear that an individual entitlement to enforcement of a restraining order could constitute a ‘property’ interest for purposes of the Due Process Clause”).
203 See Roth, 408 U.S. at 577 (holding that property interests are not created by the Constitution).
204 See Castle Rock, 125 S. Ct. at 2803–04 (noting that “[a]lthough the underlying substantive interest is created by ‘an independent source such as state law,’ federal constitutional law determines whether that interest rises to the level of a ‘legitimate claim of entitlement’ protected by the Due Process Clause” (quoting Memphis Light, Gas & Water Div. v. Craft, 436 U.S. 1, 10–11 (1978))) (citation omitted).
property interest if state common law and constitutional norms clearly recognize such an interest. The analysis of property in *Memphis Light* does not contemplate the adoption of new tests unmoored from the examination of positive law envisioned by *Roth* in order to determine whether a state-recognized entitlement rises to a constitutional level of property. Given that Supreme Court due process jurisprudence was dominated by an entirely different paradigm prior to *Roth*, there is a notable dearth of constitutional guiding principles to which to turn. This lack of constitutional guideposts should at least encourage caution in crafting additional rules to separate state entitlements from constitutional entitlements.

Even assuming that under *Memphis Light* the Supreme Court has the ability to construct additional hurdles for recognizing procedural due process property interests, the tests that the Court employed in this case are improper. The Court first stated that Ms. Gonzales’s interest in enforcement of her restraining order fails to fit the definition of property for due process purposes because it does not “have an ascertainable monetary value, as even our *Roth*-type ‘property-as-entitlement’ cases have implicitly required.” For this proposition, the Court quoted a law review article written by Professor Merrill, but gave no further explanation beyond the inductive rationale that the Court’s precedents have implied such a test.

Professor Merrill’s stated reasons for this test, however, are threefold. The first is inductive; most of the interests that the Court has recognized in this area have monetary value. This reason fails to account for some of the Court’s precedent, as noted below, and also fails by its own terms as applied to Ms. Gonzales’s case. As the dissent notes, if Ms. Gonzales had contracted with a private security firm instead of relying on the Castle Rock Police Department for protection from her husband, that contract would undoubtedly have ascer-

---

205 See *supra* notes 98–108 and accompanying text.
207 See *supra* notes 26–42 and accompanying text.
208 See *supra* notes 26–42 and accompanying text.
209 See *Castle Rock*, 125 S. Ct. at 2809; *Memphis Light*, 436 U.S. at 10–11.
210 *Castle Rock*, 125 S. Ct. at 2809 (quoting Merrill, *supra* note 1, at 964).
211 See *Castle Rock*, 125 S. Ct. at 2809.
212 Merrill, *supra* note 1, at 964–65.
213 See *id*.
214 See *infra* notes 215–228 and accompanying text.
tainable monetary value and would qualify as property for due process purposes.\textsuperscript{215}

Merrill’s second reason for this test is that it differentiates Roth-type property interests from procedural due process liberty interests involving prisoners.\textsuperscript{216} Yet this justification fails to explain why monetary value should be used in place of other distinguishing characteristics, or whether such a distinction is necessary at all after \textit{Sandin v. Conner} established a standards-based approach to defining liberty interests.\textsuperscript{217} Indeed, there are easier grounds for distinguishing property and liberty interests; “new” liberty interests can be seen as freedom from state restraint or punishment, while “new” property entails an entitlement to a particular government benefit.\textsuperscript{218}

Merrill’s final justification for a monetary value test is that it brings the due process definition of property more closely in line with the ordinary understanding of property, which, he argues, connotes something of value that enhances individual wealth.\textsuperscript{219} Pecuniary value is hardly the only measure of value or individual wealth, however, as \textit{Castle Rock} demonstrates.\textsuperscript{220} Ms. Gonzales’s restraining order enriched her, in that she did not have to seek out a private security firm for protection or move away from her husband.\textsuperscript{221} The lack of

\textsuperscript{215} \textit{Castle Rock}, 125 S. Ct. at 2823 n.19 (Stevens, J., dissenting). Although it might seem odd that contractual rights are deemed to fit the definition of property as an entitlement, the companion case to Roth apparently assumed this very point. See Perry v. Sindermann, 408 U.S. 593, 601 (1972) (treating a cause of action for breach of contract as property for procedural due process purposes); \textit{see also} Merrill, \textit{supra} note 1, at 990–92 (examining the Court’s treatment of contractual rights as due process property interests).

\textsuperscript{216} Merrill, \textit{supra} note 1, at 964–65.

\textsuperscript{217} See 515 U.S. 472, 484 (1995) (establishing an “atypical and significant hardship” standard for prisoner due process claims).

\textsuperscript{218} \textit{See} Merrill, \textit{supra} note 1, at 964–65; \textit{2004 Term—Leading Cases}, \textit{supra} note 7, at 216.

\textsuperscript{219} Merrill, \textit{supra} note 1, at 965.

\textsuperscript{220} \textit{See Castle Rock}, 125 S. Ct. at 2823 (Stevens, J., dissenting).

\textsuperscript{221} \textit{Id.} at 2824. Professor Merrill himself measured monetary value by looking to the cost of the next-best alternative in an attempt to harmonize \textit{Goss v. Lopez} with his test. Merrill, \textit{supra} note 1, at 964 n.289; \textit{see} 419 U.S. 565, 576 (1975) (recognizing a due process property interest in public education). Remarking on \textit{Goss}, Professor Merrill stated, “[A]ny parent who has contemplated sending their children to private schools knows that public schooling has a monetary value.” Merrill, \textit{supra} note 1, at 964 n.289. The majority in \textit{Castle Rock}, however, responded to the dissent’s point that paying a private party for enforcement would have ascertainable monetary value by saying that Ms. Gonzales’s interest was in enforcement of the restraining order rather than the abstract right to protection. 125 S. Ct. at 2809 n.12. Therefore, a private party would not legally be able to enforce the restraining order through arrest unless the crime occurred in that party’s presence and therefore it was a proper citizen’s arrest. \textit{See id.} This response, however, fails to account for other situations where the next-best alternative may not be legally enforceable, such as the cause of action recognized as a type of property in \textit{Logan}. \textit{See Logan}, 455 U.S. at 430.
affirmative reasons for adopting an “ascertainable monetary value” test and the lack of logic underlying its inductive justification counsel against its application.\(^{222}\)

Perhaps the most compelling argument against an “ascertainable monetary value” requirement for recognizing property for due process purposes, however, is the Court’s own precedent.\(^{223}\) *Roth* embraced the concept that property under the Due Process Clause comes in “many forms,” extending “well beyond real estate, chattels, or money.”\(^{224}\) Decisions after *Roth* have recognized property interests as varied and intangible as entitlements to public education and causes of action.\(^{225}\) Arguably, neither of these entitlements has an ascertainable monetary value, except in the restraining order’s sense of enrichment, and, therefore, each would fail the *Castle Rock* test.\(^{226}\) Of course, imposing a monetary value requirement could make property for due process purposes more closely approximate a “traditional conception of property.”\(^{227}\) This approach, however, would contradict *Roth*’s explicit language and would contradict the underlying purpose of property under the Due Process Clause: to protect a person’s reasonable expectation of reliance on a governmental conferral of benefits without arbitrary termination.\(^{228}\)

The Court in *Castle Rock* also stated that Ms. Gonzales’s interest in enforcement failed to reach the level of a constitutional property interest because it arose only incidentally, out of a function that the police have always performed: arresting or seeking warrants for those suspected of a crime.\(^{229}\) For this proposition, the Court cited *O’Bannon v. Town Court Nursing Center*, where it held that nursing home residents’ indirect interest in continued residence based on nursing home regula-

\(^{222}\) See 2004 Term—Leading Cases, supra note 7, at 215 (noting the lack of affirmative reasons for adopting the monetary value test).

\(^{223}\) See *Logan*, 455 U.S. at 430 (writing that property interests are varied and often intangible); *Roth*, 408 U.S. at 576 (stating that property may take many forms).

\(^{224}\) 408 U.S. at 571–72, 576.


\(^{226}\) See *Logan*, 455 U.S. at 430; *Goss*, 419 U.S. at 576; see also *Merrill*, supra note 1, at 964 n.289 (acknowledging that *Goss* is a borderline case involving ascertainable monetary value).

\(^{227}\) See *Castle Rock*, 125 S. Ct. at 2809.

\(^{228}\) See *Roth*, 408 U.S. at 577 (stating that the purpose of property is to protect claims upon which people rely); *Hurtado*, 110 U.S. at 527 (stating that the purpose of the Due Process Clause is to protect persons from arbitrary government action).

\(^{229}\) 125 S. Ct. at 2809.
tions did not constitute a property entitlement for due process purposes even though their direct benefits—financial payments for certain medical services—did.\textsuperscript{230}

Despite its superficial similarity to \textit{Castle Rock}, \textit{O'Bannon} is inapplicable for several reasons.\textsuperscript{231} First, \textit{O'Bannon} did not address a situation where the underlying nursing home regulations actually created an entitlement, and, therefore, it cannot stand as a bar for the creation of an entitlement in \textit{Castle Rock}'s context.\textsuperscript{232} Secondly, the incidental-direct distinction is derived from third-party beneficiary theory, which posits that “intended” beneficiaries can enforce their rights against a promisor whereas “incidental” beneficiaries cannot.\textsuperscript{233} For the analogy to stand in \textit{Castle Rock}, however, the intended beneficiaries of the Colorado restraining order statute—apparently the general public in the eyes of the Court—would have a right to assert their interests in enforcement against the government, an entirely implausible understanding.\textsuperscript{234} Furthermore, the Court’s use of the incidental-direct distinction is unworkable because the restraining order clearly implied that it \textit{was} intended to benefit specific persons—the protected persons defined by the restraining order statute.\textsuperscript{235}

V. \textsc{Giving Substance to Entitlement—A Redefinition}

Underlying much of the U.S. Supreme Court’s decision in \textit{Town of Castle Rock v. Gonzales} is a deep skepticism of the means and results of \textit{Board of Regents of State Colleges v. Roth}’s positivist approach towards identifying property interests for due process purposes and a fear that a strict reading of an apparently mandatory state law might uncover “something as vague and novel as enforcement of restraining orders” as property interests.\textsuperscript{236} Thus, instead of resting solely on \textit{Roth}'s in-

\begin{footnotesize}
\begin{enumerate}
\item See \textit{id.} at 2810 (citing 447 U.S. 773, 786–88 (1980)).
\item See \textit{infra} notes 232–235 and accompanying text.
\item See \textit{Castle Rock}, 125 S. Ct. at 2823 n.18 (Stevens, J., dissenting); \textit{O'Bannon}, 447 U.S. at 785.
\item See \textit{Restatement (Second) of Contracts} §§ 302, 304, 315 (1981).
\item See \textit{id.} Because the majority does not indicate who would constitute the intended beneficiary of the statute, one commentator has read the statute equally implausibly as intended to benefit the party most “directly” affected—in this case, Mr. Gonzales. \textit{See 2004 Term—Leading Cases, supra} note 7, at 218.
\item See \textit{Colo. Rev. Stat.} § 18-6-803.5(1.5)(a) (1999) (defining “protected person” as “the person or persons identified in the protection order as the persons or persons \textit{for whose benefit} the protection order was issued”) (emphasis added); \textit{Castle Rock}, 125 S. Ct. at 2821–22 (Stevens, J., dissenting) (citing the text of the Colorado statute).
\item See \textit{Town of Castle Rock v. Gonzales}, 125 S. Ct. 2796, 2809 (2005). At oral argument, Justice Scalia seemed to suggest that a property interest in enforcement of a restraining order
\end{enumerate}
\end{footnotesize}
quiry as to whether the state law gives rise to a legitimate claim of entitlement, the Court looks to other sources—the law of contracts for the direct-incidental distinction and a law review article for a monetary value test—to serve as constitutional limits to recognition of certain government benefits as property for due process purposes.\footnote{Given the Court’s repudiation of \textit{Roth}’s positivist approach in defining liberty interests for prisoners after \textit{Sandin v. Conner} and its general aversion to recognition of “new” property interests after the 1970s, perhaps the Court’s deployment of other standards of law to limit the interpretation of positive law in \textit{Castle Rock} is not particularly surprising.\footnote{But neither is it inevitable.} Much of the difficulty in determining what rises to the level of a property interest for purposes of due process—beyond the traditional conceptions of ownership in real estate, money, and chattels—stems from the articulation of the constitutional test.\footnote{The definition for property laid out in \textit{Roth} is perhaps deliberately vague: an entitlement rooted in an independent source such as state law.\footnote{By allowing the constitutional guarantee to adapt to the contours of other sources of law, the Court avoided having to list precisely what property for procedural due process encompasses, and risk either omitting certain types of property or becoming outdated.\footnote{The property-as-entitlement approach also avoided requiring the Court to repudiate its own precedents that recognized a variety of property interests.}} Perhaps most would be “utterly zany.” See Transcript of Oral Argument at 14, \textit{Castle Rock}, 125 S. Ct. 2796 (No. 04-278), \textit{available at} \url{http://www.supremecourtus.gov/oral_arguments/argument_transcripts/04-278.pdf}.

\footnote{See \textit{Castle Rock}, 125 S. Ct. at 2809 (citing Merrill, \textit{supra} note 1, at 964 for the implicit ascertainable monetary value requirement, and O’Bannon v. Town Court Nursing Ctr., 447 U.S. 773, 775 (1980) for the indirect benefit concept).}

\footnote{See \textit{Sandin v. Conner}, 515 U.S. 472, 484 (1995) (replacing the positivist approach to identifying procedural due process liberty interests with an atypical hardship standard); Pierce, \textit{supra} note 26, at 1973 (predicting and applauding a due process counterrevolution in the 1990s that would overturn many of the early 1970s due process cases).}

\footnote{See \textit{infra} notes 240–283 and accompanying text.}

\footnote{See Bd. of Regents of State Colls. v. Roth, 408 U.S. 564, 577 (1972); Shapiro & Levy, \textit{supra} note 1, at 113.}

\footnote{See \textit{Roth}, 408 U.S. at 577; Merrill, \textit{supra} note 1, at 921 (noting \textit{Roth}’s deliberate vagueness).}

\footnote{See \textit{Roth}, 408 U.S. at 577 (asserting that “[p]roperty interests, of course, are not created by the Constitution”).}

\footnote{See, e.g., Bell v. Burson, 402 U.S. 555, 539 (1971) (recognizing due process protections in a driving license); Goldberg v. Kelly, 397 U.S. 254, 261–62 (1970) (recognizing a property interest in welfare benefits); see also \textit{Roth}, 408 U.S. at 576 (recognizing that “property interests . . . may take many forms”).}
importantly in the Roth Court’s view, it limited the breadth of Goldberg-era due process, which threatened to subject any and all interests deemed important to due process protections. The “new” property interests that the Court has identified since Roth, however, do have certain common attributes, the articulation of which can give clearer meaning to Roth’s concept of entitlement.

The common theme for “new” property interests since Roth is the recognition of a benefit rooted in a source of law independent from the Constitution that is conferred on a specific class subject to specific conditions and terminable only under specific conditions. Identifying these common attributes allows for a more particularized definition of entitlement, and may better guide courts attempting to define property. It also allows for an approach that embraces property’s many forms, tangible and not, without federalizing tort law. Additionally, this definition calls into question Castle Rock’s statutory interpretation and its application of monetary value and incidental-direct benefit tests.

A. Independent Source of Law

Perhaps the one undisputed element of the post-Roth definition of property under the Due Process Clause is that entitlements arise from a source of law independent from the Constitution. Commentators

---

244 Compare Roth, 408 U.S. at 571 (looking to the nature, rather than the weight, of the interests at stake to determine due process protections), with Bell, 402 U.S. at 539 (stating that a driving license “may become essential in the pursuit of a livelihood” and that its suspension “adjudicates important interests of the licensees”).

245 See Merrill, supra note 1, at 921.

246 See Roth, 408 U.S. at 577; infra notes 247–283 and accompanying text.

247 See, e.g., Castle Rock, 125 S. Ct. at 2803 (stating that the existence of an entitlement turns on whether government officials can grant or deny it in their discretion); Logan, 455 U.S. at 1155 (identifying the hallmark of property as an individual entitlement grounded in state law which cannot be removed except for cause); Roth, 408 U.S. at 577 (stating that property interests are rooted in positive law and that benefits that give rise to a legitimate claim of entitlement are protected property interests).

248 See Merrill, supra note 1, at 968 (noting that a refined definition of property for due process purposes would clarify the Court’s jurisprudence).

249 See Logan, 455 U.S. at 430 (noting that the Court has recognized a wide variety of property interests, both tangible and intangible); Roth, 408 U.S. at 576 (stating that property may take many forms); see also Castle Rock, 125 S. Ct. at 2810 (noting that the Court’s approach to due process jurisprudence avoids treating the Fourteenth Amendment as a “font of tort law”) (citation omitted).


251 See id. at 2803 (quoting Roth, 408 U.S. at 577).
have bemoaned the fact that Roth’s definition of property leaves due process protections to the whim of the legislature and administrators who can grant or decline to grant government benefits at their discretion. Nevertheless, the difficulty of establishing a workable framework that incorporates a person’s subjective expectation of or reliance on a benefit without grounding that reasonable expectation in a source of law has made the Court wary of straying from Roth’s positivist approach. In addition, the Court’s repudiation of the “bitter with the sweet” hypothesis—the notion that the procedures in a law could provide all the due process protections the Constitution requires—blunted some of the criticism that the Court was allowing legislative and administrative bodies to avoid constitutional review. Furthermore, government entities are arguably not acting arbitrarily if they are proceeding within the letter of the law, thereby avoiding the pitfall that the Due Process Clause was designed to prevent.

B. Benefit or Service

The post-Roth definition of property also clearly contemplates that the independent source of law must confer a benefit on the claimant for the interest to rise to the level of a procedural due process property interest. The interests that the Court has recognized as

---

252 See generally Mashaw, supra note 21 (advocating for a broader, non-instrumentalist view of due process); Shapiro & Levy, supra note 1 (contending that a standards-based approach to applying due process protections to government benefits better comports with rule of law principles); Van Alstyne, supra note 26 (stating that under the Court’s analysis, the Due Process Clause protects only what the legislature wants it to protect).

253 See Castle Rock, 125 S. Ct. at 2803 (reaffirming Roth’s approach to identifying property interests); Pierce, supra note 26, at 1988 (noting and praising the Court’s complete retreat from the “important interest” weighing approach to identifying due process protections); see also Notes, Statutory Entitlement and the Concept of Property, 86 YALE L.J. 695, 709–14 (1977) (arguing that a concept of property based on a person’s subjective reliance would be inconsistent with the underlying purpose of property, would be inconsistent with the Court’s jurisprudence, and would radically expand the ambit of the Due Process Clause).


255 See Daniels v. Williams, 474 U.S. 327, 331 (1986) (writing that “the Due Process Clause . . . was ‘intended to secure the individual from the arbitrary exercise of the powers of government’” (quoting Hurtado v. California, 110 U.S. 516, 527 (1884))).

256 See Roth, 408 U.S. at 577 (stating that, although property under the Due Process Clause does not protect all benefits, it does protect those benefits to which the party can claim a legitimate entitlement).
due process property all connote value, as opposed to freedom from restraint.\textsuperscript{257} Although \textit{Castle Rock} attempted to limit this concept of “benefit” to pecuniary value, “benefit” should instead be defined as an enrichment for two reasons.\textsuperscript{258} First, the Court’s own precedents recognize property interests that do not have a measurable monetary value, except insofar as a person’s reliance on the benefit provided allows that person to avoid the cost of the next-best alternative.\textsuperscript{259} Additionally, the idea of a benefit as enrichment better comports with the underlying purpose of subjecting benefits to due process protection—to protect people’s reliance on a relationship rooted in the law.\textsuperscript{260} Taken out of context, the notion of “value” of “benefit,” unconnected to pecuniary value, does have the potential to be overbroad.\textsuperscript{261} However, the other proposed requirements for property interests under the Due Process Clause—the conditions imposed in the law for provision of the benefit and the conditions providing for termination of the benefit—prevent the entitlement concept from ballooning to include “all interests valued by sensible men.”\textsuperscript{262}

C. Direct Benefit

The Court’s precedents also make clear that for an asserted interest to constitute a property interest for procedural due process purposes, the source of law must confer a benefit \textit{on the claimant}.\textsuperscript{263} \textit{Castle Rock} seemed to suggest that even if the enforcement of a restraining order constituted an entitlement under state law, the indirect nature of the benefit on a claimant could bar constitutional recognition of that entitlement.\textsuperscript{264} The indirect benefit test, however, as applied in \textit{O’Bannon v. Town Court Nursing Center} and stated in contract principles, presupposes that there is a direct beneficiary of the

\textsuperscript{257} See \textit{Castle Rock}, 125 S. Ct. at 2809; Merrill, \textit{supra} note 1, at 964.

\textsuperscript{258} See \textit{Castle Rock}, 125 S. Ct. at 2809; \textit{Logan}, 455 U.S. at 430 (noting that the Court had recognized both tangible and intangible property interests “relating to the whole domain of social and economic fact”).


\textsuperscript{260} See \textit{Roth}, 408 U.S. at 577 (stating that the purpose of property is to protect claims upon which people rely).

\textsuperscript{261} See Merrill, \textit{supra} note 1, at 964 (proposing an ascertainable monetary value test to limit the breadth of the entitlement concept).

\textsuperscript{262} See \textit{Monaghan}, \textit{supra} note 18, at 409; \textit{infra} notes 263–283 and accompanying text.

\textsuperscript{263} See \textit{Castle Rock}, 125 S. Ct. at 2809–10.

\textsuperscript{264} See \textit{id.} at 2809 (noting that even if Colorado law created an entitlement, it is by no means clear that this would create a due process property interest).
statute who could claim an entitlement to the benefit. The key questions for the Court to determine, therefore, are first whether the underlying source of law creates an entitlement, and then to whom it confers the entitlement.

D. Satisfaction of Specific Conditions for Conferral of Benefit

The Court has recognized property interests rooted in sources of law other than the Constitution where the claimant fulfills certain criteria according to which the government provides a benefit. This judicial requirement is generally satisfied because almost all claimants have already satisfied the conditions for the provision of the benefit and are now trying to prevent termination of that benefit. A key condition, however, is whether the claimant is part of the class receiving the benefit as defined in the source of law. Although the requirement that the government confer a benefit on a defined class does not explicitly appear in any previous Court formulations of how to define property, several cases have impliedly required it. Such a requirement would avoid the Castle Rock Court’s palpable fear that recognizing a property interest in enforcement of a restraining order would open the floodgates to property interests in general protection laws. Such a requirement also makes intuitive sense; for example, a fire department regulation requiring that the department respond to any fire, absent directives for its conferral on a discrete class of

\[\text{} \]

\[\text{2006} \]

\[\text{Redefining Property Under the Due Process Clause} \]

\[\text{811} \]

\[\text{statute who could claim an entitlement to the benefit. The key questions for the Court to determine, therefore, are first whether the underlying source of law creates an entitlement, and then to whom it confers the entitlement.} \]

\[\text{D. Satisfaction of Specific Conditions for Conferral of Benefit} \]

\[\text{The Court has recognized property interests rooted in sources of law other than the Constitution where the claimant fulfills certain criteria according to which the government provides a benefit.} \]

\[\text{This judicial requirement is generally satisfied because almost all claimants have already satisfied the conditions for the provision of the benefit and are now trying to prevent termination of that benefit. A key condition, however, is whether the claimant is part of the class receiving the benefit as defined in the source of law. Although the requirement that the government confer a benefit on a defined class does not explicitly appear in any previous Court formulations of how to define property, several cases have impliedly required it. Such a requirement would avoid the Castle Rock Court’s palpable fear that recognizing a property interest in enforcement of a restraining order would open the floodgates to property interests in general protection laws. Such a requirement also makes intuitive sense; for example, a fire department regulation requiring that the department respond to any fire, absent directives for its conferral on a discrete class of} \]

\[\text{See O’Bannon, 447 U.S. at 786–88; Restatement (Second) of Contracts §§ 302, 304, 315 (1981); supra notes 229–235 and accompanying text.} \]

\[\text{See Restatement (Second) of Contracts §§ 302, 304, 315; supra notes 229–235 and accompanying text.} \]

\[\text{See Mathews v. Eldridge, 424 U.S. 319, 332 (1976) (recognizing a property interest in Social Security disability benefit payments based on the plaintiff’s satisfaction of statutory criteria); see also Massey, supra note 1, at 542–45 (noting in passing that the Court has recognized property interests in statutory entitlements based on the fulfillment of criteria under which the government provides a benefit or the inapplicability of the conditions depriving the person of that benefit).} \]

\[\text{See generally Castle Rock, 125 S. Ct. 2796 (refusing to examine whether Ms. Gonzales was qualified to receive the benefit of police enforcement of the restraining order).} \]

\[\text{See id. at 2821 (Stevens, J., dissenting) (noting that Ms. Gonzales was identified as the protected person in the restraining order statute); see also Logan, 455 U.S. at 430 (declaring that the hallmark of property is an individual entitlement).} \]

\[\text{See Castle Rock, 125 S. Ct. at 2809–10 (majority opinion) (stating that the benefit to Ms. Gonzales is only incidental); id. at 2822 (Stevens, J., dissenting) (arguing that enforcement of the restraining order benefits Ms. Gonzales directly); see also Logan, 455 U.S. at 430 (declaring that the hallmark of property is an individual entitlement).} \]

\[\text{See Castle Rock, 125 S. Ct. at 2808 (stating that the criminal law performs functions other than conferring a benefit on a specific class of persons).} \]
beneficiaries, should not rise to the level of a procedural due process property interest, but a government contract with a company providing fire protection should. 272

E. Conditions for Termination

The Court’s jurisprudence likewise clearly requires that a protected property interest can only be terminated under certain conditions. 273 There are two varieties of conditions for termination: (1) non-discretionary termination of a benefit, or (2) termination made “for cause” or its equivalent. 274 As the Court in Castle Rock stated and the prisoners’ liberty cases demonstrated, the existence of a protected interest often turns on whether officials can grant or deny a benefit at their discretion because a discretionary benefit cannot lead to a reasonable expectation of its continuance. 275 Therefore, this definition of entitlement follows the Court’s lead in recognizing that a benefit rooted in positive law must be couched in non-discretionary terms. 276

Another iteration of a termination condition is the “for cause” variety appearing in Logan and Memphis Light: “[t]he hallmark of property for procedural due process purposes is an individual entitlement grounded in state law, which cannot be removed except ‘for cause.’” 277 The “for cause” requirement has its roots in employment common law, but, more generally, it stands for the idea that if an entitlement can only be terminated upon certain acceptable conditions, this is enough

272 See supra notes 190–192 and accompanying text; see also Gonzales v. City of Castle Rock, 366 F.3d 1092, 1101–05 (10th Cir. 2004) (en banc) (basing its holding that enforcement of a restraining order is a protected property interest in part on the restraining order’s specific conferral on protected persons defined in the statute), rev’d sub nom. Town of Castle Rock v. Gonzales, 125 S. Ct. 2796 (2005).

273 See Castle Rock, 125 S. Ct. at 2803 (declaring that a benefit is not a protected entitlement if a government official can grant or deny it in his or her discretion); Logan, 455 U.S. at 430 (stating that the hallmark of property is entitlement that can only be terminated for cause).

274 See Castle Rock, 125 S. Ct. at 2803 (stating that a benefit is not a protected entitlement if a government official can grant or deny it in his or her discretion); Bishop v. Wood, 426 U.S. 341, 344–47 (1976) (finding no property interest in continued employment because, by ordinance, there were no exclusive conditions under which the petitioner could be deprived of the job).

275 See Castle Rock, 125 S. Ct. at 2803 (stating that a discretionary benefit is not a protected entitlement). See generally Meachum v. Fano, 427 U.S. 215 (1976) (denying a prisoner’s liberty interest claim based on the degree of discretion granted to the prison official by the regulation).

276 See Castle Rock, 125 S. Ct. at 2803.

277 See Logan, 455 U.S. at 430 (citations omitted); Memphis Light, 436 U.S. at 10.
to create an entitlement to its continuation.\textsuperscript{278} As both \textit{Logan} and \textit{Memphis Light} demonstrate, this requirement not only allows the removal of property interests if good cause is provided, but also can serve as a shield against the arbitrary termination of entitlements if the written procedures providing for their termination do not comport with those required under the common law or the Constitution.\textsuperscript{279}

Lastly, it should be noted that this approach to entitlement serves other goals of property and due process that do not explicitly appear as elements within the definition—namely, protecting one’s reasonable expectation of reliance on the law and avoiding arbitrary governmental decision making.\textsuperscript{280} The definition of entitlement under \textit{Roth}, however, assumes that these elements are only measured indirectly.\textsuperscript{281} The Court, in reaction to the excesses of \textit{Goldberg}’s important interest approach, has assiduously avoided any measurement of the reliance of the individual on the government benefit or the arbitrariness of a decision in determining whether an interest rises to the level of property.\textsuperscript{282} Nevertheless, a statute that confers a direct benefit subject to conditions for its conferral and only terminable for cause, under the definition of entitlement fleshed out by the Court’s jurisprudence, is likely to serve these goals.\textsuperscript{283}

\begin{itemize}
\item \textsuperscript{278} See \textit{Logan}, 455 U.S. at 430 (requiring a for-cause termination provision in order to recognize a property interest); \textit{Bishop}, 426 U.S. at 344–47 (declining to find a property interest in continued employment because there were no conditions under which the petitioner could be deprived of the job). This concept also implicitly acknowledges that if the denial of the benefit did not follow the stated conditions for termination, the termination was likely done arbitrarily, and hence in violation of due process. \textit{Logan}, 455 U.S. at 430; \textit{Bishop}, 426 U.S. at 344–47
\item \textsuperscript{279} See \textit{Logan}, 455 U.S. at 430 (holding that an employee’s property interest in an adjudicatory procedure could not be deprived without cause, notwithstanding company procedures); \textit{Memphis Light}, 436 U.S. at 10–11 (recognizing a property interest in continued utility service based on common-law precedents, notwithstanding the utility company’s procedures).
\item \textsuperscript{280} See \textit{Daniels}, 474 U.S. at 331 (writing that “the Due Process Clause . . . was ‘intended to secure the individual from the arbitrary exercise of the powers of government’” (quoting \textit{Hurtado}, 110 U.S. at 527)); \textit{Roth}, 408 U.S. at 577 (noting that a purpose of property in due process is to protect those claims upon which people rely).
\item \textsuperscript{281} See \textit{Roth}, 408 U.S. at 577.
\item \textsuperscript{282} See \textit{id}.
\item \textsuperscript{283} See \textit{Merrill}, supra note 1, at 967 (contending that a refined definition of the property-as-entitlement concept would protect one’s reasonable reliance on the law); \textit{supra} notes 236–282 and accompanying text.
\end{itemize}
Conclusion

The U.S. Supreme Court’s decision in Town of Castle Rock v. Gonzales marked a major stride towards undoing much of the so-called “Due Process Revolution.” Through its narrow reading of a seemingly mandatory statute, and its adoption and expansion of both a monetary value and an incidental benefit test, the Court put sharp limits on Board of Regents of State Colleges v. Roth’s continued viability in identifying property for procedural due process purposes. The two requirements it adopted in dicta—an ascertainable monetary value and a direct benefit on the recipient—are on shaky constitutional ground and should be used sparingly, if at all, in determining whether a benefit rises to the level of a property interest. In their place, the Court should adopt a clearer articulation of the defining characteristics of Roth’s definition of entitlement—a benefit rooted in a source of law independent from the Constitution that is conferred on a specific class subject to specific conditions and terminable under specific conditions—to guide them, as well as to bring the Court’s jurisprudence more closely in line with the underlying purposes of protecting property under the Due Process Clause.

Joel Hugenberger
LIFE-SUSTAINING TREATMENT LAW: A MODEL FOR BALANCING A WOMAN’S REPRODUCTIVE RIGHTS WITH A PHARMACIST’S CONSCIENTIOUS OBJECTION

Abstract: In recent years, there have been several incidents where pharmacists refused to dispense prescriptions for emergency contraception, as well as other types of contraceptives, because of their ethical, moral, or religious beliefs. State law has attempted to address this problem in various ways, but frequently fails to balance adequately the rights of a woman to access lawful contraceptive prescriptions against a pharmacist’s right to conscientiously object. This Note argues that pharmacist refusal laws should seek guidance from a similar conflict in the life-sustaining treatment context. Life-sustaining treatment law permits a health care provider to refuse to comply with a patient’s decision regarding life-sustaining treatment, but imposes additional duties on the health care provider who does so. These additional duties—requiring the health care provider to notify the patient of its policy in advance and transfer the patient to another health care facility—prevent both the patient’s and health care provider’s rights from being compromised. The Note concludes that analogous transfer and notice requirements should be placed on pharmacists who conscientiously object to dispensing contraceptive prescriptions.

Introduction

In 2002, a woman entered the pharmacy at a Kmart store in Menomie, Wisconsin to refill her birth control prescription.1 The only pharmacist on duty refused to refill it because of his religious beliefs.2 He also refused to transfer the prescription to a nearby pharmacy.3 As a result, the woman was forced to wait two days to receive her prescription, exposing her to an increased risk of preg-

---

1 Sarah Sturmon Dale, Can a Pharmacist Refuse to Dispense Birth Control?, TIME, June 7, 2004, at 22.
2 Id.
3 Id.
nancy. Although the Wisconsin Pharmacy Examining Board reprimanded the pharmacist, this was not an isolated incident—pharmacists have declined to fill birth control prescriptions because of their ethical, moral, or religious beliefs in several other states. There have been reports of similar incidents in California, Washington, Georgia, Illinois, Louisiana, Massachusetts, Texas, New Hampshire, Ohio, and North Carolina.

State legislation deals with this issue in various ways. For example, four states—Arkansas, Georgia, Mississippi, and South Dakota—have so-called “pharmacist refusal laws” explicitly allowing pharmacists to refuse to dispense contraceptive prescriptions because of their religious, moral, or ethical beliefs. Illinois, on the other hand, explicit-


5 See Amanda Paulson, Culture Wars Hit Local Pharmacy, CHRISTIAN SCIENCE MONITOR, Apr. 8, 2005, at 1 (reporting several instances of pharmacists refusing to dispense contraceptives and “morning-after” pills on moral or religious grounds); Kimberlee Roth, Pharmacists, Doctors Refuse to Dispense Pill on Moral Grounds, CHI. TRIB., Nov. 17, 2005, at C1 (same); Rob Stein, Pharmacists’ Rights at Front of New Debate, WASH. POST, Mar. 28, 2005, at A1 (same). See generally In re Disciplinary Proceedings Against Neil T. Noesen, No. LS0310091PHM (Wis. Pharmacy Examining Bd. April 13, 2005), http://drl.wi.gov/dept/decisions/docs/0405070.htm (reprimanding the pharmacist for failing to inform the pharmacy of his refusal to transfer oral contraceptives to another pharmacy and requiring him to notify future employers of his objection prior to providing pharmacy services).

6 See Paulson, supra note 5, at 1; Stein, supra note 5, at A1.

7 Compare Ark. Code Ann. § 20-16-304(4) (West 2005) (exempting pharmacists as well as doctors from being required to dispense contraceptive prescriptions when the refusal is based on religious or conscientious objection), MISS. CODE ANN. §§ 41-107-3(a), (b), -5(1) (2005) (stating that no health care provider, including pharmacists, shall be required to participate in a health care service that violates his or her conscience), S.D. CODIFIED LAWS § 36-11-70 (2004) (declaring that no pharmacist will be required to dispense medication if there is reason to believe it will be used to cause an abortion), and GA. COMP. R. & REGS. 480-5-.03(n) (2001) (stating that a pharmacist’s refusal to fill a prescription based upon ethical or moral beliefs will not be considered unprofessional conduct), with CAL. BUS. & PROF. CODE § 733(b)(3) (West Supp. 2006) (allowing a pharmacist to refuse to dispense a drug on religious, moral, or ethical grounds only if the pharmacist notifies his or her employer beforehand in writing and only if protocols are in place to ensure that the patient has timely access to the drug), and ILL. ADMIN. CODE tit. 68, § 1330.91 (2005) (requiring a pharmacy to dispense valid prescriptions for contraceptives without delay, unless the contraceptive is out of stock, in which case the pharmacy must provide an alternative drug or transfer the prescription to another pharmacy).

8 See Ark. Code Ann. § 20-16-304(4); MISS. CODE ANN. §§ 41-107-3(a), (b), -5(1); S.D. CODIFIED LAWS § 36-11-70; GA. COMP. R. & REGS. 480-5-.03(n). This Note will refer to laws
Reproductive Rights & Pharmacists’ Conscientious Objections

Itly does not allow pharmacists to refuse to dispense contraception. In 2005, in response to an incident in downtown Chicago where a pharmacist at a drugstore refused to give emergency contraception to two women, Governor Rod Blagojevich made permanent a temporary order requiring Illinois pharmacies to dispense all prescriptions for emergency contraception. Under the law, any pharmacy in the state that sells contraceptives approved by the U.S. Food and Drug Administration (the “FDA”) must fill all contraceptive prescriptions. If the contraceptive is not in stock, the pharmacy must provide an alternative drug, order the drug, or transfer the prescription to another local pharmacy.

California also restricts a pharmacist’s right to refuse to dispense prescriptions to which he or she is ethically, morally, or religiously opposed. In order to refuse to dispense a prescription, the pharmacist must first notify his or her employer in writing of the drugs that he or she objects to dispensing. The pharmacist will then only be permitted to refuse to dispense the prescription if the employer can reasonably accommodate this refusal without creating undue hardship on the employer. The employer is also required to ensure that the patient has timely access to the drug, despite the pharmacist’s refusal to dispense it.

Although the specific issue of a pharmacist’s right to refuse to dispense contraception is relatively new, the larger debate over whether a health care provider can refuse to perform or provide services because

---

9 ILL. ADMIN. CODE tit. 68, § 1330.91.
10 See ILL. ADMIN. CODE tit. 68, § 1330.91; John Chase, Legislators Back Edict on Birth Pills, CHI. TRIB., Aug. 17, 2005, at 3; Paulson, supra note 5, at 1.
11 ILL. ADMIN. CODE tit. 68, § 1330.91.
12 Id.
13 CAL. BUS. & PROF. CODE § 733(b)(3) (West Supp. 2006).
14 Id.
15 Id. (noting that “undue hardship” shall have the same meaning as applied to “undue hardship” under subdivision (l) of section 12940 of California’s Government Code). Section 12940 cross-references to the definition of “undue hardship” in Section 12926. CAL. GOV’T CODE § 12940 (West 2005). Subsection (s) of section 12926 defines “undue hardship” as an action requiring significant difficulty or expense, when considered in light of the financial resources of the facility involved, among other factors. CAL. GOV’T CODE § 12926(s) (West 2005).
16 CAL. BUS. & PROF. CODE § 733(b)(3).
of moral or religious objections has swirled for years.\footnote{Adam Sonfield, \emph{Rights v. Responsibilities: Professional Standards and Provider Refusals, Guttmacher Rep’t on Pub. Pol’y}, Aug. 2005, at 7, 7, available at http://www.guttmacher.org/pubs/tgr/08/3/gr080307.pdf (noting that the debate over the right of pharmacists to refuse to dispense contraception is new but the larger issue of whether health care providers can refuse to perform services because of their religious convictions is not). This Note uses “health care provider” as a general term to refer to any individual person, facility, or organization involved in the health care field.} Most states have laws dating back to the 1970s that explicitly allow doctors, nurses, or health care institutions to refuse to provide or participate in certain types of reproductive health care, such as abortion and contraceptive services.\footnote{See, e.g., \textsc{Ark. Code Ann. § 20-16-304(4), (5) (West 2005); Colo. Rev. Stat. § 25-6-102(9) (2004); Del. Code Ann. tit. 24, § 1791 (1997); Fla. Stat. Ann. § 381.0051(6) (West 2002); Iowa Code Ann. § 146.1 (West 2005); Me. Rev. Stat. Ann. tit. 22, § 1903(4) (2004); Neb. Rev. Stat. Ann. § 28-337 (LexisNexis 2003); Or. Rev. Stat. § 435.225 (2003); 18 Pa. Cons. Stat. Ann. § 3213(d) (2000); S.C. Code Ann. § 44-41-40 (2002); Tenn. Code Ann. § 68-34-104(5) (2001); W. Va. Code Ann. § 16-2B-4 (LexisNexis 2001); Wyo. Stat. Ann. § 42-5-101(d) (2005).} The debate continues, however, as to how far these health care provider refusal laws should be expanded.\footnote{See, e.g., Adam Sonfield, \emph{New Refusal Clauses Shatter Balance Between Provider ‘Conscience,’ Patient Needs, Guttmacher Rep’t on Pub. Pol’y}, Aug. 2004, at 1, 1–3, available at http://www.guttmacher.org/pubs/tgr/07/3/gr070301.pdf (contending that more recent movements to expand health care refusal laws, including such a movement by pharmacists, threaten the balance between allowing for providers’ conscientious objections and protecting patients’ access to health care services); Lynn D. Wardle, \emph{Protecting the Rights of Conscience of Health Care Providers}, 14 \textsc{J. Legal Med.} 177, 177 (1993) (arguing that current health care refusal laws provide inadequate protection for health care providers with conscientious objections and proposing legislation expanding the right of health care providers to conscientiously object to performing or providing health care services).} Much of this debate centers on pharmacists, who were not explicitly included in the early refusal laws and are now insisting that they too have a right to conscientiously object.\footnote{See Pharmacists for Life International, Legal News and Information for Pro-Life Pharmacists of Conscience, http://www.pfl.org/main.php?pfl=legal (last visited Mar. 24, 2006) [hereinafter Pharmacists for Life] (advocating for an expansion of conscience clauses to pharmacists).} This issue is especially pertinent because controversial advances in medical technology, such as emergency contraception, have met both with high demand from patients and moral objections from health care providers.\footnote{See Sonfield, supra note 19, at 1.} 

This Note addresses pharmacists’ right to conscientiously object to dispensing contraception. Part I discusses the rights of women to access contraception and abortion services in the United States and the limitations placed on those rights.\footnote{See infra notes 28–74 and accompanying text.} Part II addresses the history of health care provider refusal laws, the factors contributing to the
expansion of refusal laws, and the application of these laws to life-sustaining treatment decisions.\textsuperscript{23} Part III focuses on efforts to expand health care provider refusal laws to pharmacists and recent legislation addressing this issue.\textsuperscript{24} Part IV discusses the inadequacies of current pharmacist refusal laws, which fail to balance adequately a pharmacist’s right to conscientiously object against a woman’s right to access contraception.\textsuperscript{25} This Part then looks to life-sustaining treatment court decisions and legislation to provide a model for balancing these competing interests and concludes that current laws should include transfer and notice requirements.\textsuperscript{26} Although relevant, this Note will not address possible First Amendment arguments based on the Free Exercise Clause or the Establishment Clause in relation to a pharmacist’s right to conscientiously object.\textsuperscript{27}

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{23} See infra notes 75–190 and accompanying text.
\item \textsuperscript{24} See infra notes 191–235 and accompanying text.
\item \textsuperscript{25} See infra notes 236–250 and accompanying text.
\item \textsuperscript{26} See infra notes 251–283 and accompanying text. Several authors have explored aspects of pharmacist refusal laws, but have not utilized the approach that courts have taken towards life-sustaining treatment options as a template for addressing the issue. See generally Lorraine Schmall, Birth Control as a Labor Law Issue, 13 DUKE J. GENDER L. & POL’Y 139 (2006) (arguing that legal obstructions to birth control, such as health care provider refusal laws, frustrate gender equality in the workplace and hence raise labor law issues); Amy Bergquist, Note, Pharmacist Refusals: Dispensing (With) Religious Accommodation Under Title VII, 90 MINN. L. REV. 1073 (2006) (analyzing pharmacist refusals under the religious accommodation requirements of Title VII); Bryan A. Dykes, Note, Proposed Rights of Conscience Legislation: Expanding to Include Pharmacists and Other Health Care Providers, 36 GA. L. REV. 565 (2002) (proposing that state legislatures expand and strengthen health care provider refusal laws as applied to pharmacists and other health care providers so as to protect the moral integrity and personal autonomy of health care providers); Minh N. Nguyen, Comment, Refusal Clauses & Pro-Life Pharmacists: How Can We Protect Ourselves from Them?, 8 SCHOLAR 251 (2006) (discussing the effect of pharmacist refusal laws on women’s reproductive freedom, health, and rights); Teliska, supra note 4 (arguing that broad pharmacist refusal laws will adversely affect low-income women and rural women in particular).
\item \textsuperscript{27} In Employment Division, Department of Human Resources v. Smith, the U.S. Supreme Court upheld the state’s denial of unemployment compensation for employees who had ingested illegal drugs—even though the employees had ingested the drug as part of a religious practice—because the law prohibiting the use of the drug was generally applicable. 492 U.S. 872, 890 (1990). The Court held that generally applicable laws that prohibit the exercise of religion do not offend the Free Exercise Clause of the First Amendment. Id. at 878–79. As such, it is possible that the Constitution would not require exemption for pharmacists who religiously object to dispensing contraceptives from generally applicable laws requiring contraceptive prescriptions to be filled. See id. Additionally, the Court noted that religious accommodation and exemption should be left to the political process. Id. at 890. Moreover, a health care provider’s conscientious objection may also be based on ethical or moral objections, rather than religious objections, in which case the Free Exercise Clause of the First Amendment would presumably not apply. See, e.g., Valley Hosp. Ass’n v. Mat-Su Coal. for Choice, 948 P.2d 963, 972 (Alaska 1997); Brophy v. New England Sinai
\end{itemize}
\end{footnotesize}
I. HISTORY OF ACCESS TO CONTRACEPTION AND ABORTION

A. History of Access to Contraception

The U.S. Supreme Court first recognized a right to be free from governmental intrusion with regard to contraceptive use in *Griswold v. Connecticut* in 1965.\(^{28}\) The statute at issue in *Griswold* made the use of contraception illegal, as well as the assistance and counseling of that use.\(^{29}\) The Court held that a state could not prohibit contraceptive use between married couples because such activity fell within a zone of privacy guaranteed by the Constitution.\(^{30}\) This zone of privacy was a penumbral right emanating from several fundamental constitutional guarantees in various amendments, including the First, Third, Fourth, Fifth, and Ninth Amendments.\(^{31}\)

In particular, *Griswold* noted that contraceptive use was part of the fundamental right of privacy surrounding the marital relationship.\(^{32}\) While the Court acknowledged that other types of contraception regulation could be upheld (such as regulations regarding manufacture or sales), a regulation banning contraceptive use altogether

---


\(^{29}\) *Id.* at 480.

\(^{30}\) *Id.* at 485–86.

\(^{31}\) *Id.* at 484.

\(^{32}\) *Id.* at 485–86.
was unnecessarily broad and violated the fundamental right of married couples to make a private decision to use contraceptives.\textsuperscript{33}

In 1972, the Supreme Court in \textit{Eisenstadt v. Baird} extended the right to use contraceptives to unmarried persons.\textsuperscript{34} A Massachusetts law made it illegal to distribute contraceptives to unmarried persons, while allowing distribution to married persons.\textsuperscript{35} While the Court struck down the law as violating the Equal Protection Clause in treating married and unmarried persons differently, it also observed that the right to privacy associated with using contraceptives is a fundamental right as well.\textsuperscript{36} In particular, the Court noted that the right to privacy grants the individual freedom from governmental intrusion into fundamentally personal decisions, such as whether to bear or beget a child.\textsuperscript{37}

In its 1977 decision in \textit{Carey v. Population Services International}, the Supreme Court reaffirmed the reasoning and holdings of both \textit{Griswold} and \textit{Eisenstadt}.\textsuperscript{38} In \textit{Carey}, a New York statute not only prohibited the distribution of contraceptives to anyone under the age of sixteen and the advertisement or display of contraceptives, but also prohibited anyone other than a licensed pharmacist from distributing contraceptives to persons sixteen or older.\textsuperscript{39} The Court again stated that the decision to bear or beget a child is at the very heart of the choices that are constitutionally protected by the right of privacy.\textsuperscript{40} The Court then concluded that the statute’s prohibitions on contraceptive use, distribution, and advertisement were unconstitutional because they did not serve any compelling state interests that justified burdening such a fundamental right.\textsuperscript{41}

\textbf{B. History of Access to Abortion}

The Supreme Court extended the right of privacy to abortion in 1973, in the landmark case of \textit{Roe v. Wade}.\textsuperscript{42} In \textit{Roe}, the Court struck down a statute that prohibited abortions except when necessary to

\begin{itemize}
  \item \textsuperscript{33} \textit{Griswold}, 381 U.S. at 485.
  \item \textsuperscript{34} 405 U.S. 438, 453–55 (1972).
  \item \textsuperscript{35} \textit{Id.} at 442.
  \item \textsuperscript{36} \textit{Id.} at 453–55.
  \item \textsuperscript{37} \textit{Id.} at 453.
  \item \textsuperscript{38} 431 U.S. 678, 685 (1977).
  \item \textsuperscript{39} \textit{Id.} at 681.
  \item \textsuperscript{40} \textit{Id.} at 685.
  \item \textsuperscript{41} \textit{Id.} at 690–91, 694–99, 701–02.
  \item \textsuperscript{42} 410 U.S. 113, 153 (1973).
\end{itemize}
save the life of the mother.\textsuperscript{43} The Court held that the fundamental right to privacy rooted in the Due Process Clause of the Fourteenth Amendment was broad enough to encompass a woman’s decision to terminate her pregnancy.\textsuperscript{44} The Court also held that the state had compelling interests in both the health of the mother and in the potential life of the fetus.\textsuperscript{45} These interests, however, became compelling at various stages in the pregnancy, which the Court divided into three-month periods called trimesters.\textsuperscript{46} The state’s interest in the health of the mother was paramount during the first trimester; therefore, the state could not regulate abortion during this period.\textsuperscript{47} After the first trimester, the interest in the potential life rose in importance and the state could regulate abortion in ways that were reasonably related to maternal health.\textsuperscript{48} After the point of viability, which the Court placed at the end of the second trimester, the state could choose to regulate or even proscribe abortion, except when necessary to preserve the life or health of the mother.\textsuperscript{49}

In 1992, in \textit{Planned Parenthood of Southeastern Pennsylvania v. Casey}, the Supreme Court revisited the right to access an abortion in examining various requirements under a Pennsylvania abortion statute.\textsuperscript{50} In so doing, the Court upheld the central holding of \textit{Roe}, recognizing the fundamental right of a woman to choose to have an abortion.\textsuperscript{51} The Court, however, rejected \textit{Roe’s} approach of balancing the competing interests in a pregnancy through a trimester framework.\textsuperscript{52} Instead, the Court held that state regulation of abortion is unconstitutional when it places an undue burden on a woman’s choice to have an abortion.\textsuperscript{53} Thus, a state may regulate abortions to promote its inter-

\begin{footnotesize}
\textsuperscript{43} Id. at 117–18, 166.
\textsuperscript{44} Id. at 153.
\textsuperscript{45} Id. at 163–64.
\textsuperscript{46} Id. at 162–64.
\textsuperscript{47} See Roe, 410 U.S. at 164–65.
\textsuperscript{48} Id.
\textsuperscript{49} Id.
\textsuperscript{50} 505 U.S. 833, 844 (1992). The five provisions at issue in the Pennsylvania statute were: (1) a requirement that a woman seeking an abortion give her informed consent prior to the procedure; (2) a provision mandating informed consent of one parent for a minor to obtain an abortion, but also providing for a judicial bypass procedure; (3) a requirement that, unless certain exceptions apply, a married woman seeking an abortion must sign a statement indicating that she has notified her husband; (4) an exemption from the foregoing requirements based upon a “medical emergency”; and (5) certain reporting requirements for facilities providing abortion services. \textit{Id.}
\textsuperscript{51} Id. at 871.
\textsuperscript{52} Id. at 873.
\textsuperscript{53} Id. at 874.
\end{footnotesize}
est in potential life and the health of the mother at any point in the pregnancy, so long as such regulation does not amount to an undue burden on a woman’s access to an abortion.\footnote{Id. at 878.} The Court reaffirmed that a state may choose to forbid abortions, but only after the point of viability.\footnote{Casey, 505 U.S. at 879.} Additionally, even though a state may proscribe abortions after viability, there must still be an exception to allow for an abortion when necessary for the preservation of the life or health of the mother.\footnote{Id. Subsequent Supreme Court cases have reiterated this health exception requirement. See Ayotte v. Planned Parenthood of N. New England, 126 S. Ct. 961, 967 (2006); Stenberg v. Carhart, 530 U.S. 914, 930 (2000).}

C. Limitations on the Federal Constitution Abortion Right

After the Supreme Court’s decision in \textit{Roe} holding that a woman has a fundamental right to have an abortion, some lower courts inferred that public hospitals, as state actors, could not refuse to perform elective abortions because such refusal violated a woman’s constitutional right to an abortion.\footnote{See Doe v. Charleston Area Med. Ctr., 529 F.2d 638, 643–45 (4th Cir. 1975); Doe v. Hale Hosp., 500 F.2d 144, 147 (1st Cir. 1974). Because the Fourteenth Amendment ostensibly only applies to the states, as opposed to persons or private organizations, the Court’s rulings on abortion do not apply to a privately-funded hospital. See U.S. CONST. amend. XIV; see also Hale Hosp., 500 F.2d at 147 (implying that a privately-funded hospital would not be subject to \textit{Roe}). Under the concept of state action, however, an organization that receives significant public funding or is heavily regulated by the state may be required to abide by the Constitution to the same degree as the state. See Jackson v. Metropolitan Edison Co., 419 U.S. 345, 351 (1974) (stating that there must be a “sufficiently close nexus between the State and the challenged action of the regulated entity so that the action of the latter may be fairly treated as that of the State itself”). The Court, however, has sharply limited the degree to which even a public hospital is currently required to provide abortions. See infra notes 61–68 and accompanying text.} For example, in \textit{Doe v. Hale Hospital}, a public hospital owned by the city of Haverhill, Massachusetts had a policy of allowing therapeutic abortions, but forbidding elective abortions.\footnote{500 2d. at 145.} The U.S. Court of Appeals for the First Circuit held that a public medical facility may not forbid elective abortions so long as it offers medically indistinguishable procedures, such as therapeutic abortions.\footnote{Id. at 147.} Likewise, in \textit{Doe v. Charleston Area Medical Center}, the Fourth Circuit held that Charleston Area Medical Center’s policy of prohibiting abortions except when necessary to save the life of the
mother conflicted with the holding in Roe, and thus was unconstitutional.60

The Supreme Court, however, has since rejected the argument that states are required either to provide elective abortions through public facilities or to fund elective abortions through public funds.61 For example, in its 1977 decision in Maher v. Roe, the Court clarified that Roe did not grant an unqualified right to an abortion from the state; rather, it protected women from unduly burdensome interference with a woman’s freedom to choose to terminate a pregnancy.62 Thus, a state could choose to favor a policy of childbirth over abortion and implement that policy through the allocation of funds without creating an unduly burdensome interference with a woman’s right to choose an abortion.63 Women still had the option of using private sources for abortion services.64

Additionally, in 1989, in Webster v. Reproductive Health Services, the Court held that nothing in the Constitution required states to enter or remain in the abortion business or entitled private physicians and patients to access public facilities for the performance of abortions.65 In Webster, a Missouri statute prohibited the use of public employees and facilities to perform or assist in elective abortions.66 As in Maher, the state’s decision to use public facilities and staff to encourage childbirth over abortion did not place an obstacle in the path of a woman who chooses to have an abortion.67 A woman could still choose to have an abortion, the Court reasoned, even though she had to rely on a privately employed physician to perform the procedure.68

D. Broader Abortion Rights Under State Constitutions

Supreme Court decisions thus interpreted the federal Constitution as not requiring states to fund or provide abortion services.69 Some state courts, however, have held that a state actor’s denial of

60 Doe, 529 F.2d at 643–45.
62 Maher, 432 U.S. at 473–74. At issue was a Connecticut statute that limited state Medicaid benefits to abortions that were necessary to save the life of the mother. Id. at 466.
63 Id. at 473–74.
64 Id. at 474.
65 492 U.S. at 509–10.
66 Id. at 501.
67 Id. at 509–10.
68 Id.
69 See Webster, 492 U.S. at 509–10; Maher, 432 U.S. at 474.
abortion services or a state’s refusal to fund abortion services violates a woman’s right to choose an abortion under their state constitutions. For example, in Valley Hospital Ass’n v. Mat-Su Coalition for Choice, the Supreme Court of Alaska held that a quasi-public hospital, as a state actor, violated a woman’s fundamental right to an abortion found in Alaska’s state constitution by refusing to perform elective abortions. By holding that a quasi-public institution was a state actor subject to the limitations in the state constitution, the court necessarily implied that a public hospital would be similarly subjected to constitutional limitations as a state actor.

Additionally, in Committee to Defend Reproductive Rights v. Myers, the Supreme Court of California held that the state’s refusal to fund elective abortions under Medi-Cal, California’s Medicaid program, violated California’s state constitution. The court reasoned that although the state had no constitutional obligation to pay for medical care for the poor, once it decided to provide such benefits, it could not withhold these benefits from otherwise qualified individuals simply because they chose to exercise their constitutional right to an abortion.

II. History of Health Care Provider Refusal Laws

A. Initial Wave of Health Care Provider Refusal Laws

Health care provider refusal laws initially appeared in the 1970s as a general response to the constitutional requirements posed by the U.S. Supreme Court’s 1973 decision of Roe v. Wade, and as a specific response to a court’s issuance of a preliminary injunction requiring a Catholic hospital to perform a sterilization procedure. On October 27, 1972, in Taylor v. St. Vincent’s Hospital, the U.S. District Court for

---


71 948 P.2d at 971.

72 See id. at 969–70.

73 625 P.2d at 798–99.

74 Id.

the District of Montana issued a preliminary injunction that required a Catholic hospital—which was considered a state actor by virtue of receiving certain federal funds—to perform a sterilization procedure on a woman.\textsuperscript{76} Despite the fact that performing a sterilization procedure was against the Catholic hospital’s religious beliefs, the court reasoned that, as a state actor, the hospital could not refuse to perform the sterilization without infringing a woman’s constitutional right to receive such services.\textsuperscript{77}

In June 1973, less than a year after the District of Montana’s preliminary injunction, Congress passed the Church Amendment.\textsuperscript{78} The Church Amendment prevented health care entities that received certain federal funds and individuals employed by those entities from being forced to perform certain reproductive services to which those entities and their employees were morally or religiously opposed.\textsuperscript{79} The Church Amendment opened the door for states to follow with similar laws permitting health care providers to refuse to perform reproductive services.\textsuperscript{80}

Current health care provider refusal laws reflect this initial focus on allowing providers to refuse to perform reproductive services.\textsuperscript{81} As such, most states have a health care provider refusal law referring specifically to abortion procedures.\textsuperscript{82} Some state laws simply declare that a health care facility is not required to permit abortions within the

\textsuperscript{76} See 369 F. Supp. at 950–51.
\textsuperscript{77} See id. at 950.
\textsuperscript{80} See ACLU REPORT, supra note 75, at 1.
facility. In several states, however, only private or religiously-affiliated health care facilities are permitted to decline to perform abortions. Additionally, most states allow an individual person, employee, or health care provider to refuse to perform or participate in abortion procedures. A number of states also require that the person who objects to performing or participating in the procedure either advise the facility of his or her objection verbally or file a written statement with the facility stating his or her objection.

Some states also allow individual health care providers or facilities to refuse to provide contraceptive services. As with refusal laws relating to abortion, some states limit the right to refuse to provide contraceptive services to private institutions. In contrast, some states specifically permit state employees or facilities to refuse to provide contraceptive services.

B. Lower Courts’ Interpretation of Health Care Provider Refusal Laws

Lower courts have interpreted health care provider refusal clauses in a variety of ways. One court held that the constitutional

---

right to an abortion trumps the health care provider’s statutory right to refusal. In *Valley Hospital Ass’n v. Mat-Su Coalition for Choice*, for example, the Supreme Court of Alaska rejected a quasi-public hospital’s invocation of Alaska’s refusal law, which provided that a hospital could decline to offer abortions for reasons of moral conscience. The court reasoned that it could not defer to the legislature when a statute led to a violation of a woman’s fundamental right to an abortion as embodied in Alaska’s constitution. Thus, it held that the refusal law, as applied to the quasi-public hospital in this case, was unconstitutional.

Other courts have narrowly interpreted their health care provider refusal laws by limiting their application to certain individuals or certain medicines. For instance, in *Brownfield v. Daniel Freeman Marina Hospital*, a rape victim sought declaratory and injunctive relief from a Catholic hospital for its failure to provide her with information and access to emergency contraception as part of her treatment. The hospital claimed it was exempt from providing this treatment under the state’s Therapeutic Abortion Act, which allowed a religious non-profit hospital to avoid liability for failing to perform or permit the performance of an abortion on its premises. The California Court of Appeal concluded that since emergency contraception prevents rather than terminates a pregnancy, the treatment was not considered an abortion under the meaning of the statute.

---

91 See *Valley Hosp. Ass’n*, 948 P.2d at 971–72.
92 Id.
93 Id. at 972.
94 Id.
95 See *Brownfield*, 256 Cal. Rptr. at 244–45; *Spellacy*, 1978 WL 3437, at *3–4. The Spellacy court, in an unreported case, chose to interpret refusal clauses narrowly so as not to allow certain individuals not directly involved in providing the health care service to invoke it. See 1978 WL 3437, at *3–4. In that case, an admissions clerk refused to admit patients entering the hospital for abortions because of her religious beliefs. Id. at *1. The Court of Common Pleas of Pennsylvania did not allow the clerk to invoke Pennsylvania’s health care provider refusal clause. Id. at *3–4. The court reasoned that the refusal clause only allowed individuals directly involved in performing the abortion and whose services were essential to the performance of the procedure to refuse to participate because of their religious beliefs. Id. Because the clerk’s duties were ancillary to the procedure and merely clerical, she could not be exempted from her admissions duties. Id.
96 256 Cal. Rptr. at 242.
97 Id. at 244.
98 Id. at 245.
Additionally, at least one court held that even if a health care provider refusal law is written as an absolute, an employer may still terminate an employee for refusing to perform a procedure if accommodating his or her religious beliefs would create an undue hardship on the employer. In *Kenny v. Ambulatory Centre of Miami, Florida, Inc.*, a nurse was demoted after she refused to participate in abortion procedures. Although Florida’s refusal law did not require consideration of the hardship on the employer created by the employee’s refusal to participate in certain services, the court adopted the federal standard under Title VII of the Civil Rights Act of 1964. This standard required the employer to reasonably accommodate the employee’s religious beliefs unless doing so would result in undue hardship. Because a reasonable accommodation of the nurse’s beliefs would not have created an undue hardship on the hospital’s ability to perform abortions, the nurse rightfully invoked the state’s refusal law and was wrongfully terminated.

By contrast, one court held that the language of its state health care provider refusal law gave a hospital employee an unqualified statutory right to refuse to participate in a medical service because of moral objections. In *Swanson v. St. John’s Lutheran Hospital*, a nurse refused to participate in a sterilization procedure and was subsequently dismissed from her employment. Because the refusal law did not consider the hardship a hospital might suffer as a result of an employee’s conscientious objection, the Supreme Court of Montana in turn did not consider any hardship that might be suffered by the hospital from the nurse’s refusal.

C. Factors Contributing to the Expansion of Health Care Provider Refusal Laws

Beginning in the 1990s, health care provider refusal laws attracted renewed attention as a means to protect health care employ-

---

99 See *Kenny*, 400 So. 2d at 1266.
100 Id. at 1263.
101 Id. at 1264–66.
102 Id. at 1266–67. The court noted that determining whether a good faith effort to reasonably accommodate an employee’s religious beliefs had been made and whether the employer had suffered undue hardship in his attempts to do so was a factual determination, to be made on a case-by-case basis. Id.
103 Id.
104 See *Swanson*, 597 P.2d at 710.
105 Id. at 704–05.
106 Id. at 709–10.
ees’ conscientious objections to reproductive services.\textsuperscript{107} State legislatures expanded the types of procedures covered by these laws as well as the number and types of providers who may invoke them.\textsuperscript{108} At least three factors contributed to this renewed pressure: (1) the rise of religious health care systems, (2) the expansion of managed care, and (3) the development of controversial medical technology.\textsuperscript{109}

First, religious health care systems have become more significant in the health care provider market.\textsuperscript{110} For example, as of September 2005, four of the top ten largest non-profit health care systems were Catholic-owned, including the largest non-profit health care system.\textsuperscript{111} Additionally, there are 611 Catholic hospitals in the United States, representing approximately 12\% of all hospitals nationwide.\textsuperscript{112} Because of its ethical directives that prohibit or limit controversial services such as contraceptive sterilization, in vitro fertilization, prescription or dispensation of contraceptive devices, and abortions, the prevalence of Catholic health care is significant.\textsuperscript{113} Moreover, this impact is intensified by the fact that over a quarter of the 611 Catholic

\begin{thebibliography}{99}
\item[107] ACLU Report, supra note 75, at 1; Sonfield, supra note 19, at 1.
\item[108] ACLU Report, supra note 75, at 1; Sonfield, supra note 19, at 1.
\item[110] ACLU Report, supra note 75, at 1; CATHOLIC HEALTH CARE, supra note 109, at 1–6.
\item[111] CATHOLIC HEALTH CARE, supra note 109, at 1–2. This Note focuses mainly on Catholic health care not only because of its size in the health care market, but also because it places the greatest restrictions on reproductive health care services. See Susan Berke Fogel & Lourdes A. Rivera, Saving Roe Is Not Enough: When Religion Controls Healthcare, 31 FORDHAM URB. L.J. 725, 732 (2004) (arguing that limits must be placed on corporate health entities’ ability to restrict access to health care services based on religious beliefs). For example, the United Methodist Church recognizes a woman’s right to an abortion. Donald H.J. Hermann, Religious Identity and the Health Care Market: Mergers and Acquisitions Involving Religiously Affiliated Providers, 34 CREIGHTON L. REV. 927, 959 (2001) (analyzing the health care market’s integration of services and challenges faced when secular and religious entities integrate). Also, the Presbyterian Church leaves the moral decision to have an abortion in the hands of the woman, while noting that the decision to have an abortion should be a last resort. Id. at 959–60. Additionally, Judaism is likewise not as restrictive as the Catholic Church regarding reproductive health care services. Id. at 959. For example, Judaism welcomes the option of in-vitro fertilization. Id. Jewish hospitals also do not restrict medical services. Fogel & Rivera, supra, at 732.
\item[112] CATHOLIC HEALTH CARE, supra note 109, at 1.
\end{thebibliography}
hospitals in the United States are located in rural areas. Thus, for a woman located in a rural area, the only practically available hospital may be Catholic; as such, her ability to access many reproductive services may be severely limited.

Second, the rapid expansion of managed care has also contributed to the renewed interest in health care provider refusal laws because managed care plans restrict enrollees to a limited pool of health care providers. Traditionally, a patient could seek out alternative providers if his or her primary provider refused to provide certain services; fee-for-service insurance paid for any provider a patient chose. Managed care plans, however, only pay for services performed by certain health care providers. As a result, a patient may be unable to obtain certain services if his or her primary providers refuse to perform them and the patient is unable to afford an alternate provider’s services not covered by his or her managed care plan.

Moreover, there has been marked growth in religiously affiliated managed care organizations, which often object to paying for certain reproductive services. For example, as of September 2005, abortion services were largely unavailable under Catholic HMOs. In addition, only approximately one-half of Catholic HMOs cover contraception or sterilization services. This is a significantly lower percentage than the 93% of all HMOs that cover at least one form of contraception, and the 86% that cover sterilization services. As such, it is a significant limitation for the nearly 2.5 million individuals covered by Catholic managed care plans in the United States as of 2000.

Additionally, under the Balanced Budget Act of 1997, Congress now allows Medicaid managed care organizations to refuse to cover certain services based on religious or moral objections. This statute allows a Medicaid managed care organization to refrain from provid-

114 Catholic Health Care, supra note 109, at 1.
115 See id. at 1–2.
116 ACLU Report, supra note 75, at 2; Gold, supra note 109, at 1.
117 ACLU Report, supra note 75, at 2.
118 Id.; Gold, supra note 109, at 1.
119 ACLU Report, supra note 75, at 2; Gold, supra note 109, at 1.
120 ACLU Report, supra note 75, at 2; Gold, supra note 109, at 1.
121 Catholic Health Care, supra note 109, at 4. An HMO (health maintenance organization) is a type of managed care organization. See id.
122 Id.
123 Id.
124 Id.
ing, reimbursing for, or providing coverage of a counseling or referral service if the organization objects to the provision of such services on moral or religious grounds.\(^\text{126}\)

In addition to the rise in religiously affiliated health care organizations and managed care, the development and increased availability of controversial medical technology has also contributed to the movement towards expanding refusal laws.\(^\text{127}\) While most health care provider refusal laws cover abortion, contraception, and sterilization, the range of medical technologies to which health care providers are morally or religiously opposed is broader than those services.\(^\text{128}\) In addition to abortion and sterilization services, objectionable technology can include emergency contraception, in vitro fertilization, medical research involving human embryos or fetuses, withdrawal or withholding of life-sustaining treatment, and physician-assisted suicide.\(^\text{129}\)

D. Judicial and Legislative Response to Conflict Between Health Care Providers and Patients’ Life-Sustaining Treatment Decisions

Life-sustaining treatment in particular has generated considerable conflict between patients and health care providers who conscientiously object to complying with a patient’s decision to forgo or withdraw from such treatment.\(^\text{130}\) Life-sustaining treatment refers to any treatment that serves to prolong life without reversing the underlying medical condition.\(^\text{131}\) For example, a hospital may keep a patient alive through the use of a respirator or ventilator because she cannot breathe on her own.\(^\text{132}\) Additionally, a doctor may give a patient artificial nutrition and hydration because he cannot eat or drink independently.\(^\text{133}\)

\(^{126}\) Id.

\(^{127}\) ACLU Report, supra note 75, at 3; Sonfield, supra note 19, at 1.

\(^{128}\) ACLU Report, supra note 75, at 3; Sonfield, supra note 19, at 1.

\(^{129}\) ACLU Report, supra note 75, at 3; Sonfield, supra note 19, at 1.


\(^{132}\) See id.

\(^{133}\) See id.
The right to refuse life-sustaining treatment has been recognized by both the judiciary and state legislatures as a broad fundamental right.\(^{134}\) Courts have recognized this right as emanating from a variety of sources.\(^{135}\) For example, the right to refuse life-sustaining treatment has been affirmed as a corollary to the common-law right of informed consent.\(^{136}\) It has also been recognized as rooted in the right to self-determination and personal autonomy.\(^{137}\) Additionally, some state courts have recognized that the right to refuse life-sustaining treatment can be encompassed within the right to privacy in their respective state constitutions.\(^{138}\) Lastly, in *Cruzan v. Director, Missouri Department of Health*, the Supreme Court not only noted the various sources to which the right to refuse life-sustaining treatment has been attributed, but also affirmed that the right to refuse life-sustaining medical treatment was a liberty interest encompassed in the Due Process Clause of the Fourteenth Amendment.\(^{139}\)

States have also recognized the fundamental right to refuse medical treatment in statutory terms.\(^{140}\) For instance, Alabama law declares that “competent adult persons have the right to control the decisions relating to the rendering of their own medical care, including, without limitation, the decision to have medical procedures . . . withheld, or withdrawn in instances of terminal conditions and permanent unconsciousness.”\(^{141}\) Similarly, Tennessee law states that “every person has the fundamental and inherent right to die naturally with as much dignity as circumstances permit and to accept, refuse, withdraw from,


\(^{137}\) See, e.g., *Gray*, 697 F. Supp. at 586; *Brophy*, 497 N.E.2d at 633.


\(^{139}\) *Cruzan*, 497 U.S. at 277–78.


or otherwise control decisions relating to the rendering of the person’s own medical care.”

Accordingly, all states allow a patient (or a surrogate decision maker) to decide to have life-sustaining treatment withdrawn or withheld from the patient, even though such a decision may result in the patient’s death. Patients or their surrogates are authorized to make such a decision by virtue of a state’s advance directive statute. An advance directive can take the form of a living will, drafted by the patient, which directs a health care provider as to which medical treatment to provide or cease providing. Additionally, a patient may appoint a health care proxy or agent to make a health care decision for him or her in the event the patient is incompetent.

Nevertheless, a patient’s decision to withdraw or withhold life-sustaining treatment can raise moral or ethical conflicts for a health care provider. For example, in *Brophy v. New England Sinai Hospital, Inc.*, both the patient’s physician and the hospital believed that withdrawing a feeding tube from a patient in a persistent vegetative state would constitute a harmful act that deliberately produced death. Similarly, in *Delio v. Westchester County Medical Center*, the hospital objected to withdrawing a feeding tube from a patient because it be-

142 TENN. CODE ANN. § 32-11-102(a). Additionally, the West Virginia legislature states that “[c]ommon law tradition and the medical profession in general have traditionally recognized the right of a capable adult to accept or reject medical or surgical intervention affecting one’s own medical condition.” W. VA. CODE ANN. § 16-30-2(b)(1).


144 See, e.g., Ala. Code § 22-8A-4; CAL. PROB. CODE §§ 4670, 4671(a); CONN. GEN. STAT. ANN. §§ 19a-575, 576; KY. REV. STAT. ANN. § 311.623; MASS. GEN. LAWS ANN. ch. 201D, § 2; S.D. CODIFIED LAWS § 34-12D-2; TENN. CODE ANN. § 32-11-104(a); W. VA. CODE ANN. § 16-30-4.

145 See, e.g., Ala. Code § 22-8A-4(a); CONN. GEN. STAT. ANN. § 19a-575; KY. REV. STAT. ANN. § 311.623(1)(a), (b); S.D. CODIFIED LAWS § 34-12D-2; TENN. CODE ANN. § 32-11-104(a).

146 See, e.g., Ala. Code § 22-8A-4(b); CONN. GEN. STAT. ANN. § 19a-576(a); KY. REV. STAT. ANN. § 311.623(1)(c); MASS. GEN. LAWS ANN. ch. 201D, § 2; W. VA. CODE ANN. § 16-30-4.

147 See, e.g., Gray, 697 F. Supp. at 583; Morrison, 253 Cal. Rptr. at 532; Bouvia, 225 Cal. Rptr. at 304; Brophy, 497 N.E.2d at 632; Jobes, 529 A.2d at 437; Elbaum, 544 N.Y.S.2d at 843; Delio, 516 N.Y.S.2d at 680.

148 497 N.E.2d at 632.
lieved that doing so would be contrary to its moral and ethical standards and its mission of preserving life.  

Withdrawing from life-sustaining treatment may cause religious concerns for a health care provider as well. For instance, the Ethical and Religious Directives for Catholic Health Care Services states that there should be a presumption in favor of providing nutrition and hydration to all patients, including patients who require medically assisted nutrition and hydration, as long as this is of sufficient benefit to outweigh the burdens involved to the patient. Thus, a patient may wish to withdraw or withhold life-sustaining treatment, but if the health care provider believes the benefits of treatment outweigh the burdens to the patient, the health care provider may refuse to do so on religious grounds.

1. Judicial Response to the Conflict Between a Patient’s Request to Refuse Life-Sustaining Treatment and a Health Care Provider’s Conscientious Objection

Several court decisions have addressed the situation where a health care provider conscientiously objects to a patient’s request to withdraw or withhold life-sustaining treatment. Some courts attempted to protect the patient’s right to refuse treatment while still accommodating the health care provider’s conscientious objection by requiring that the patient be transferred to another health care provider willing to carry out the patient’s wishes to withdraw life-sustaining treatment.

For example, in *Brophy*, the wife of a patient in a persistent vegetative state wished to have life-sustaining nutrition and hydration withdrawn from the patient. The patient’s physicians as well as the

---

149 516 N.Y.S.2d at 680.
150 See e.g., *Bartling*, 209 Cal. Rptr. at 223; *Requena*, 517 A.2d at 887–89; *Ethical Directives*, supra note 113, §§ 56–59.
151 *Ethical Directives*, supra note 113, § 58.
152 See id.
153 See, e.g., *Gray*, 697 F. Supp. at 583; *Morrison*, 253 Cal. Rptr. at 532; *Bouvia*, 225 Cal. Rptr. at 298; *Bartling*, 209 Cal. Rptr. at 221; *Brophy*, 497 N.E.2d at 627; *Jobes*, 529 A.2d at 437; *Requena*, 517 A.2d at 887; *Elbaum*, 544 N.Y.S.2d at 842; *Delio*, 516 N.Y.S.2d at 680.
155 497 N.E.2d at 631–32. A person in a persistent vegetative state is unaware of his or her surroundings and lacks cognitive abilities, but may still exhibit non-cognitive functions, such as breathing, circulation, and spontaneous movements. See Nat’l Institute of Neurological Disorders and Stroke, Nat’l Inst. of Health, NINDS Coma and Persistent Vegetative State Information Page, http://www.ninds.nih.gov/disorders/coma/coma.htm (last visited July 29, 2006). As the political firestorm surrounding the removal of Terri
hospital refused to withdraw the feeding tube, believing that doing so would constitute a harmful act which would deliberately produce death.\textsuperscript{156} The hospital, however, was not opposed to transferring the patient to another facility willing to remove the feeding tube.\textsuperscript{157} The Supreme Judicial Court of Massachusetts reasoned that so long as the patient’s right to self-determination was not denied and the patient’s wishes were fulfilled at another facility, it was possible to preserve the ethical integrity of the hospital and its staff.\textsuperscript{158}

More than one court has gone further to protect the patient’s right to withdraw life-sustaining treatment over the health care provider’s objections by explicitly requiring the health care provider to carry out the patient’s wishes if a transfer was not possible.\textsuperscript{159} For example, in \textit{Gray v. Romeo}, the family of a patient in a persistent vegetative state requested that the patient be removed from artificial nutrition and hydration.\textsuperscript{160} The hospital refused to comply with the request because, as an institution, it was opposed to withdrawing nutrition and hydration.\textsuperscript{161} It equated the withdrawal of nutrition and hydration with euthanasia.\textsuperscript{162} The U.S. District Court for the District of Rhode Island, however, while noting that it would be unsettling to the health care professionals who opposed withdrawing the feeding tube, held that the hospital had no choice but to acknowledge the patient’s right of self-determination.\textsuperscript{163} As a result, if the patient could not be promptly transferred to a health care facility willing to respect the patient’s wishes, the hospital was obligated to comply with the patient’s requests.\textsuperscript{164}

Furthermore, at least one court has held that, even though transferring the patient to another facility was possible, the facility had to comply with the patient’s wishes to refuse artificial feeding.\textsuperscript{165} In \textit{In re Requena}, a woman dying of amyotrophic lateral sclerosis, which rendered her unable to ingest food orally, refused to accept artificial nu-

---

\textsuperscript{156} Brophy, 497 N.E.2d at 632.
\textsuperscript{157} \textit{Id.}
\textsuperscript{158} \textit{Id.} at 639.
\textsuperscript{159} \textit{See Gray}, 697 F. Supp. at 591; Jobes, 529 A.2d at 450; Elbaum, 544 N.Y.S.2d at 848.
\textsuperscript{160} 697 F. Supp. at 583.
\textsuperscript{161} \textit{Id.}
\textsuperscript{162} \textit{Id.}
\textsuperscript{163} \textit{Id.} at 591.
\textsuperscript{164} \textit{Id.}
\textsuperscript{165} \textit{See Requena}, 517 A.2d at 889.
trition and hydration.\textsuperscript{166} The hospital had a policy against withholding or withdrawing food or fluids from patients and refused to comply with the patient’s wishes.\textsuperscript{167} It was, however, willing to assist in transferring the patient to another institution which would carry out the patient’s request.\textsuperscript{168} Although transferring the patient was both medically and practically feasible, the Superior Court of New Jersey ordered the hospital to honor the patient’s request.\textsuperscript{169} The court stated that it would be too emotionally and psychologically upsetting to force the patient to leave the hospital.\textsuperscript{170}

In deciding whether to require the facility to comply with the patient’s wishes, courts also sometimes consider whether the patient and his or her family had notice of the facility’s official policy regarding life-sustaining treatment prior to or upon admission to the facility.\textsuperscript{171} Some courts have reasoned that, without prior notice of a health care provider’s policy regarding life-sustaining treatment, a patient is entitled to presume that his or her right to refuse treatment would not be hindered.\textsuperscript{172}

For instance, in \textit{In re Jobes}, a nursing home refused on moral grounds to comply with a patient’s family’s request to remove the patient from life-sustaining artificial nutrition and hydration.\textsuperscript{173} The nursing home, however, did not inform the patient’s family of its policy regarding artificial feeding until the patient’s family requested that the feeding tube be removed.\textsuperscript{174} Nor was there any indication that the nursing home’s policy against removing artificial nutrition and hydration was ever formalized.\textsuperscript{175} The Supreme Court of New Jersey noted that the patient’s family had no notice that they were surrendering the right to choose among medical alternatives when they placed the patient in the nursing home.\textsuperscript{176} As a result, the court held that the

\begin{enumerate}
\item \textit{Id.} at 887–88.
\item \textit{Id.} at 888–89.
\item \textit{Id.} at 889.
\item \textit{Id.} at 890.
\item \textit{Requena}, 517 A.2d at 892–93.
\item \textit{Requena}, 517 A.2d at 870; \textit{Elbaum}, 544 N.Y.S.2d at 847–48.
\item \textit{Id.} at 450.
\item \textit{Id.}
\item \textit{Id.}
\end{enumerate}
family was entitled to rely on the nursing home’s apparent willingness to defer to the family’s choice of medical treatment for the patient.\textsuperscript{177}

2. Legislative Response to the Conflict Between a Patient’s Wish to Refuse Life-Sustaining Treatment and a Health Care Provider’s Conscientious Objection

Most states allow a health care provider to refuse to comply with a patient’s life-sustaining treatment decision as contained within the patient’s advance directive if the health care provider opposes the treatment decision the patient has made.\textsuperscript{178} Refusing to comply with the patient’s wishes, however, usually imposes additional duties on the health care provider.\textsuperscript{179} These additional duties ensure that a patient is not prevented from exercising his or her right to refuse medical treatment, while still accommodating the health care provider’s religious, moral, or ethical integrity.\textsuperscript{180}

For example, in most states, the physician or facility that objects to honoring the patient’s decision must aid the patient and his or her family in transferring the patient to a physician or facility that will carry out the decision.\textsuperscript{181} Additionally, a few states address the situation that arises if a transfer cannot be arranged.\textsuperscript{182} For instance, New York and Massachusetts impose an additional duty: if a transfer cannot be arranged, the health care provider must either seek judicial

\textsuperscript{177} Id.
relief or honor the patient’s or health care proxy’s decision.\textsuperscript{183} By contrast, Indiana laws declare that if an attending physician cannot find another physician willing to carry out the patient’s wishes to withdraw or withhold life-sustaining treatment, the attending physician may refuse to comply with the patient’s wishes.\textsuperscript{184} Moreover, in Kansas, Rhode Island, and Utah, failure to effect the transfer of the patient to another facility constitutes unprofessional conduct.\textsuperscript{185}

Some states also require a health care facility to give notice to the patient and his or her family, prior to or upon admission, of the facility’s policy of refusing to withdraw life-sustaining treatment.\textsuperscript{186} In fact, federal regulation has also imposed a notice requirement on health care facilities that receive Medicare and Medicaid funding.\textsuperscript{187} In order to refuse to comply with a patient’s advance directive for reasons of conscience, a health care facility must have clear and precisely written policies explaining the facility’s limitations.\textsuperscript{188} The statement, at a minimum, should (1) clarify differences between institution-wide conscience objections and those that may be raised by individual physicians, (2) identify the state legal authority permitting such objections, and (3) describe the medical conditions that would be affected by these objections.\textsuperscript{189} This information must be given when an individual is admitted to the facility.\textsuperscript{190}

### III. Pharmacist Refusal Laws

Refusal clauses have expanded not only to cover health care services other than abortion and contraception, such as life-sustaining treatment, but also to include health care providers other than doctors and nurses who were not initially included in refusal laws.\textsuperscript{191} Specifically, pharmacists have asserted their right to conscientiously object.\textsuperscript{192}

\begin{itemize}
\item \textsuperscript{184} Ind. Code Ann. § 16-36-4-13.
\item \textsuperscript{187} 42 C.F.R. § 489.102(a)(1) (2004).
\item \textsuperscript{188} Id. § 489.102(a)(1)(i).
\item \textsuperscript{189} Id.
\item \textsuperscript{190} Id. § 489.102(b)(1), (2).
\item \textsuperscript{191} Sonfield, supra note 19, at 1.
\item \textsuperscript{192} See Pharmacists for Life, supra note 20.
\end{itemize}
Some have criticized pharmacists’ motivations in asserting this right, claiming that it is an organized campaign by religious conservatives to encroach on patients’ rights to access controversial medical services.193 On the other hand, pharmacists, like physicians, play an important and unique role in the health care system.194 Among other tasks, pharmacists have the role of reviewing the appropriateness of a patient’s drug therapy, intervening when there is a potential problem, and instructing patients on how to use medications safely and effectively.195 It follows that, like physicians and nurses, pharmacists should not be forced to participate in procedures to which they have religious, moral, or ethical objections.196

A. Emergency Contraception’s Role in the Push for Pharmacist Refusal Laws

Pharmacists’ interest in invoking refusal laws has arisen largely in response to the availability of emergency contraception. In 1998 and 1999, two pharmaceutical products, Preven and Plan B, became available as products packaged and marketed specifically for use as emergency contraception.197 The arrival of these drugs thus sparked considerable controversy among pharmacists who objected to prescribing emergency contraception due to religious, ethical, or moral beliefs.198

193 See Sonfield, supra note 19, at 1; Paulson, supra note 5, at 1; Stein, supra note 5, at A1.

MACY & L. 1, 2–6 (1996).


196 Hearing, supra note 194, app. at 62.


In particular, emergency contraception raised objections from pro-life pharmacists because they believed that they would be participating in an abortion procedure by dispensing the drug.\textsuperscript{199}

The American Pharmaceutical Association, however, has stated that emergency contraception does not cause an abortion.\textsuperscript{200} According to this view, implantation of the fertilized egg establishes a pregnancy.\textsuperscript{201} Unlike RU-486, which acts after implantation of the egg into the uterus, emergency contraception acts before implantation of the fertilized egg into the uterus, and thus cannot disrupt an established pregnancy, which is measured by whether the fertilized egg has already been implanted.\textsuperscript{202} It is therefore not an abortifacient.\textsuperscript{203}

Pro-life groups, however, contend that emergency contraception does cause an abortion.\textsuperscript{204} For example, Pharmacists for Life, a pro-life association of pharmacists, believes that pregnancy begins with the fertilization of an egg rather than at the time of implantation.\textsuperscript{205} Because emergency contraception may prevent pregnancy by preventing the fertilized egg from implanting, this group believes that emergency contraception terminates an established pregnancy by acting after fertilization of the egg has occurred.\textsuperscript{206} Pharmacists for Life and other groups also believe that daily oral contraceptives can also act as chemical abortifacients because those medications can interfere with processes that occur after fertilization of the egg.\textsuperscript{207}

\begin{flushleft}
\textsuperscript{199} See Cohen, supra note 198, at 1; Bergquist, supra note 26, at 1077–78.
\textsuperscript{200} APhA Report, supra note 197, at 3.
\textsuperscript{201} See id.
\textsuperscript{202} See APhA Report, supra note 197, at 3–4; Cohen, supra note 198, at 1–2. Some pharmacists’ objections to dispensing emergency contraception may actually have grown out of confusion in the media equating emergency contraception and RU-486. See APhA Report, supra note 197, at 3–4; Cohen, supra note 198, at 1–2. For a discussion of the FDA’s approval of RU-486, which is otherwise known as mifepristone, see generally Lars Noah, \textit{A Miscarriage in the Drug Approval Process? Mifepristone Embroils the FDA in Abortion Politics}, 366 \textit{Wake Forest L. Rev.} 571 (2001).
\textsuperscript{203} See APhA Report, supra note 197, at 3.
\textsuperscript{205} Id.
\textsuperscript{206} Id.
\textsuperscript{207} See APhA Report, supra note 197, at 3, 12; The Pill, supra note 204.
\end{flushleft}
B. Pharmacist Refusal Laws

1. Broad Pharmacist Refusal Laws

The controversy concerning emergency contraception provided momentum for expanding health care provider refusal laws to include pharmacists. For example, South Dakota law allows a pharmacist to refuse to dispense medication that would cause an abortion or destroy an unborn child. The law defines an unborn child as an organism existing from fertilization to live birth. Under this definition, any type of oral contraceptive, not merely emergency contraception, may qualify as an abortifacient because these medications may interfere with processes that occur after fertilization.

Arkansas has also recognized that pharmacists may object to filling prescriptions for any type of contraceptive. Its pharmacist refusal clause states that nothing prohibits a pharmacist from refusing to furnish any contraceptive procedures, supplies, or information.

Two states enacted even broader legislation, allowing for pharmacists’ refusal to dispense any medication to which they were morally opposed. For example, Georgia’s State Board of Pharmacy’s Code of Professional Conduct simply declares that a pharmacist who, for moral reasons, refuses to fill a prescription does not violate professional conduct standards. Additionally, Mississippi’s broad refusal clause grants all health care providers, including pharmacists, the right to refuse to participate in any health care service that violates the health care provider’s conscience.

---


210 Id. § 22-1-2(50A) (2004).

211 See APhA REPORT, supra note 197, at 3, 12.


213 Id.


215 Ga. Comp. R. & REGS. 480-5-.03(n).

216 Miss. Code Ann. §§ 41-107-3(a), (b), -5(1).
2. Restrictive Pharmacist Refusal Laws

By contrast, Illinois has restricted a pharmacist’s right to refuse to dispense contraception.217 Under Illinois law, if a contraceptive is in stock at a pharmacy, the pharmacy must dispense the contraceptive to a patient without delay.218 If the contraceptive is not in stock, the pharmacy must provide a suitable alternative, order the drug, or transfer the prescription to a local pharmacy of the patient’s choice.219 This law does not mention an individual pharmacist’s right to object ethically, morally, or religiously to dispensing contraceptives.220 Thus, in a situation where a pharmacy stocks contraceptive medications, but the only pharmacist on duty has a religious, moral, or ethical opposition to dispensing contraceptives, the pharmacist may have no choice but to dispense the prescription.221

Moreover, Illinois’s Health Care Right of Conscience Act most likely does not protect objecting pharmacists.222 The statute offers broad protection from liability for health care personnel who refuse to perform or assist in any way in any health care service that is contrary to their conscience.223 It appears, however, that pharmacists are not considered “health care personnel” for purposes of the statute.224 In fact, Republican State Senator Dan Rutherford said that attempts to include pharmacists in the definition of “health care personnel” have failed.225

3. A More Balanced Approach Taken by the American Pharmacists Association and California

Rather than choosing to give broad rights to either pharmacists or patients, the American Pharmacists Association (the “APhA”) responded to Illinois’s law by taking a more balanced position on a pharmacist’s right to refuse to dispense contraceptives based on

217 Ill. Admin. Code tit. 68, § 1330.91 (2005); Paulson, supra note 5, at 1.
218 Ill. Admin. Code tit. 68, § 1330.91.
219 Id.
220 See id.
221 See id.
223 See id.
224 See id. 70/3(c) (defining “health care personnel” as “any nurse, nurses’ aide, medical school student, professional, paraprofessional or any other person who furnishes, or assists in the furnishing of, health care services”); see also Chase, supra note 10, at 3 (noting the lack of statutory protection for conscientiously objecting pharmacists).
225 See 745 Ill. Comp. Stat. Ann. 70/3(c); Chase, supra note 10, at 3.
moral objections. The APhA recognizes a pharmacist’s right to refuse, but contends that there should also be a system in place to ensure that the patient’s health care needs are served. Among other things, the APhA suggests that a pharmacy should be staffed such that if an objecting pharmacist is on call, another pharmacist in the same pharmacy could dispense the medication. If that is not possible, the pharmacy could transfer the prescription to a different pharmacy that will dispense the medication. Additionally, the APhA suggests that prescribers direct patients to pharmacies that they know will dispense the medication.

Similarly, California allows a pharmacist to refuse to dispense prescriptions to which he or she is ethically, morally, or religiously opposed. California law requires the pharmacist and pharmacy, however, to ensure that the patient is able to receive her prescription in a timely manner. To refuse to dispense a prescription, the pharmacist must have previously notified his or her employer in writing of the drugs that he or she objects to dispensing. The pharmacist will then only be permitted to refuse if the employer can reasonably accommodate this refusal without creating undue hardship on the employer. The employer must also establish means to ensure that the

226 Hearing, supra note 194, app. at 61–64.
227 Id. app. at 62.
228 Id.
229 Id.
230 Id.
232 Id.
233 Id.
234 Id. (noting that “undue hardship” shall have the same meaning as applied to “undue hardship” pursuant to subdivision (l) of section 12940 of California’s Government Code). Section 12940 cross-references to the definition of “undue hardship” in Section 12926. Cal. Gov’t Code § 12940 (West 2005). Subsection (s) of section 12926 defines “undue hardship” as an action requiring significant difficulty or expense, when considered in light of several factors, including (1) the nature and cost of the accommodation needed; (2) the overall financial resources of the facilities involved in the provision of the reasonable accommodations, the number of persons employed at the facility, and the effect on expenses and resources or the impact otherwise of these accommodations upon the operation of the facility; (3) the overall financial resources of the covered entity, the overall size of the business of a covered entity with respect to the number of employees, and the number, type, and location of its facilities; (4) the type of operations, including the composition, structure, and functions of the workforce of the entity; and (5) the geographic separateness, administrative, or fiscal relationship of the facility or facilities. Cal. Gov’t Code § 12926(s) (West 2005).
patient has timely access to the drug, despite the pharmacist’s refusal
to dispense it.\textsuperscript{235}

IV. LIFE-SUSTAINING TREATMENT LAW AS A MODEL FOR BALANCING
A PHARMACIST’S RIGHT TO CONSCIENTIOUSLY OBJECT WITH A
WOMAN’S RIGHT TO OBTAIN CONTRACEPTIVES

A. A Failure to Balance Competing Interests in Current
Pharmacist Legislation

A few states have explicitly addressed the issue of a pharmacist’s
right to refuse to dispense contraceptives based on religious, moral,
or ethical views.\textsuperscript{236} Of those states, Georgia, Mississippi, South Dakota,
and Arkansas fail to protect adequately a woman’s fundamental right
to use contraceptives.\textsuperscript{237} In these states, a pharmacist may flatly refuse
to dispense a prescription for a contraceptive and the pharmacist has
no obligation to assist the woman in filling her prescription.\textsuperscript{238} Moreover,
there is no requirement that pharmacists or their employers give
notice to women that their prescriptions may not be filled because of
the pharmacists’ religious, moral, or ethical beliefs.\textsuperscript{239} In sum, the legis-
lation leaves women with no recourse for filling their contraceptive
prescriptions in a timely manner.\textsuperscript{240}

The inability to fill contraceptive prescriptions in a timely man-
ner significantly hinders a woman’s fundamental right to choose to
access and use contraceptives.\textsuperscript{241} Timely access to birth control pre-

\textsuperscript{235} \textit{Cal. Bus. \\& Prof. Code} § 733(b)(3).

\textsuperscript{236} \textit{See Ark. Code Ann.} § 20-16-304(4) (West 2005); \textit{Cal. Bus. \\& Prof. Code}
§ 733(b)(3) (West Supp. 2006); \textit{Miss. Code Ann.} §§ 41-107-3(a), (b), -5(1) (2005); S.D.

\textsuperscript{237} \textit{See Ark. Code Ann.} § 20-16-304(4); \textit{Miss. Code Ann.} §§ 41-107-3(a), (b), -5(1);
S.D. \textit{Codified Laws} § 36-11-70; Planned Parenthood of Se. Pa. v. Casey, 505 U.S. 833, 851
\\& Regs.} 480-5-03(n).

\textsuperscript{238} \textit{See Ark. Code Ann.} § 20-16-304(4); \textit{Miss. Code Ann.} §§ 41-107-3(a), (b), -5(1);
\\& Regs.} 480-5-03(n).

\textsuperscript{239} \textit{See Ark. Code Ann.} § 20-16-304(4); \textit{Miss. Code Ann.} §§ 41-107-3(a), (b), -5(1);

\textsuperscript{240} \textit{See Ark. Code Ann.} § 20-16-304(4); \textit{Miss. Code Ann.} §§ 41-107-3(a), (b), -5(1);

\textsuperscript{241} \textit{See} \textit{Casey}, 505 U.S. at 851; \textit{Carey}, 431 U.S. at 685; \textit{Eisenstadt}, 405 U.S. at 453–54; \textit{Gris-
wold}, 381 U.S. at 485; \textit{APhA Report}, supra note 197, at 4; \textit{Julie Cantor \\& Ken Baum, The
Limits of Conscientious Objection—May Pharmacists Refuse to Fill Prescriptions for Emergency Con-
scriptions is essential to women who use contraceptives because the effectiveness of many contraceptives depends on the ability to take them within a certain time frame.\textsuperscript{242} Timeliness is an especially critical issue for women seeking to fill a prescription for emergency contraception; the medication is most effective at preventing pregnancy if taken within seventy-two hours or fewer of unprotected sexual intercourse.\textsuperscript{243}

If a woman experiences a delay in receiving emergency contraception because she cannot find a pharmacist who will dispense it, the time within which she must take the medication may pass, dramatically increasing the chance of pregnancy.\textsuperscript{244} For a woman who lives in an urban area with a multitude of pharmacies within a small radius, or a woman with the means to travel to another pharmacy, this may be no more than an inconvenience.\textsuperscript{245} For low-income women and women who live in rural areas, however, such a law places a significant obstacle in the way of exercising the fundamental right to access and use contraceptives.\textsuperscript{246} If the fundamental right to use and access contraceptives is to have any meaning, all women who choose to exercise their right need to be able to fill their contraceptive prescriptions in a timely manner, so that their choice to exercise this right will not be rendered ineffective or moot by the passage of time.\textsuperscript{247}

Illinois, on the other hand, has failed to protect adequately the pharmacist’s right to conscientiously object to dispensing such prescriptions by requiring a pharmacy to dispense a contraceptive, without delay, if it is in stock.\textsuperscript{248} A pharmacist who conscientiously objects to dispensing contraceptives may be forced to dispense a contraceptive if he or she is employed by a pharmacy that stocks contraceptives.\textsuperscript{249} This could happen, for example, if the pharmacy has contraceptives in stock and the pharmacist is the only pharmacist on duty at

\begin{itemize}
\item \textsuperscript{242} See APhA Report, \textit{supra} note 197, at 4; Cantor & Baum, \textit{supra} note 241, at 2010.
\item \textsuperscript{243} See APhA Report, \textit{supra} note 197, at 4; Cantor & Baum, \textit{supra} note 241, at 2010.
\item \textsuperscript{244} See APhA Report, \textit{supra} note 197, at 4; Cantor & Baum, \textit{supra} note 241, at 2010.
\item \textsuperscript{245} See Casey, 505 U.S. at 851; Carey, 431 U.S. at 685; Eisenstadt, 405 U.S. at 453; Griswold, 381 U.S. at 485; Cantor & Baum, \textit{supra} note 241, at 2010; Cohen, \textit{supra} note 198, at 2.
\item \textsuperscript{246} See Casey, 505 U.S. at 851; Carey, 431 U.S. at 685; Eisenstadt, 405 U.S. at 453; Griswold, 381 U.S. at 485; Cantor & Baum, \textit{supra} note 241, at 2010; Cohen, \textit{supra} note 198, at 2.
\item \textsuperscript{247} See Casey, 505 U.S. at 851; Carey, 431 U.S. at 685; Eisenstadt, 405 U.S. at 453–54; Griswold, 381 U.S. at 485–86; Cantor & Baum, \textit{supra} note 241, at 2010; Cohen, \textit{supra} note 198, at 2.
\item \textsuperscript{248} See ILL. ADMIN. CODE tit. 68, § 1330.91 (2005).
\item \textsuperscript{249} See id.
\end{itemize}
a particular time, or if there is no other willing pharmacist working at the same time to fill the prescription.\textsuperscript{250}

B. Model for Balancing Competing Interests in Pharmacist Refusal Clause Legislation

1. Fundamental Rights May Present Moral Conflict

Life-sustaining treatment court decisions and legislation provide a model for pharmacist refusal clause legislation because the use of contraceptives implicates a fundamental liberty interest in the same way that refusing life-sustaining treatment does.\textsuperscript{251} The Supreme Court has recognized that the Due Process Clause of the Fourteenth Amendment protects both the right to use and access contraceptives and the right to refuse life-sustaining treatment.\textsuperscript{252} In its 1992 \textit{Planned Parenthood of Southeastern Pennsylvania v. Casey} decision, the Supreme Court affirmed that the law affords protection to personal decisions relating to contraception.\textsuperscript{253} It noted that personal decisions regarding the use of contraceptives involve choices central to personal dignity and autonomy.\textsuperscript{254}

Similarly, in \textit{Cruzan v. Director, Missouri Department of Health}, the Court assumed that the liberty interest in the Due Process Clause of the Fourteenth Amendment encompassed the right to refuse life-sustaining medical treatment.\textsuperscript{255} The Court also noted that the right to refuse medical treatment has been recognized by states as a common-law right to bodily integrity and self-determination, as a constitutionally protected right to privacy within state constitutions, and, in some cases, as a statutory right.\textsuperscript{256} In sum, at their core, both the right to use and access contraception and the right to refuse life-sustaining treatment speak to an individual’s right to self-determination and to control fundamental decisions involving his or her own body.\textsuperscript{257}

\textsuperscript{250} See id.


\textsuperscript{253} 505 U.S. at 851.

\textsuperscript{254} Id.

\textsuperscript{255} 497 U.S. at 278.

\textsuperscript{256} Id. at 277–78.

Although both are protected as fundamental rights under the U.S. Constitution, the refusal of life-sustaining treatment and the use of contraceptives are also religiously, morally, and ethically controversial. For example, the teachings of the Catholic Church limit health care providers’ ability to allow refusal of life-sustaining treatment and provide contraceptives. Thus, guided by their own religious convictions, health care providers may oppose both withdrawing or withholding life-sustaining treatment and dispensing contraceptives in certain circumstances. Even if a health care provider is not particularly religious, his or her conscientious objection may be based on strong ethical or moral beliefs.

Thus, a patient’s request to withdraw or withhold treatment may cause religious, moral, or ethical conflict for a health care provider in the same way that a patient’s request for contraceptives does. As a result, examining the conflict in the life-sustaining treatment context provides guidance in resolving the conflict between patients seeking to have their contraceptive prescriptions timely filled and conscientiously objecting pharmacists.

2. Proposed Components of Pharmacist Refusal Clause Legislation

Both the transfer and notice requirements embraced by courts and legislatures allow the patient to exercise his or her right to self-determination by refusing unwanted medical treatment, without vio-


260 See id.

261 See, e.g., Brophy, 497 N.E.2d at 632; Jobes, 529 A.2d at 437; Delio, 516 N.Y.S.2d at 680.

262 See, e.g., Gray, 697 F. Supp. at 583; Morrison, 253 Cal. Rptr. at 532; Bouvia, 225 Cal. Rptr. at 304; Bartling, 209 Cal. Rptr. at 222–23; Brophy, 497 N.E.2d at 632; Jobes, 529 A.2d at 437; Requena, 517 A.2d at 888–89; Elbaum, 544 N.Y.S.2d at 843; Delio, 516 N.Y.S.2d at 680.

lating the religious, moral, or ethical integrity of the objecting health care provider.\textsuperscript{264} A notice requirement gives patients and their families advance knowledge of a health care provider’s policy regarding life-sustaining treatment, notifying patients that they cannot rely on that facility to fulfill their wishes and enabling patients to find another provider if they so choose.\textsuperscript{265} A transfer requirement ensures that the patient’s wishes will be carried out in the event the health care provider refuses the patient’s request.\textsuperscript{266}

Imposing similar duties on pharmacists who conscientiously object to filling prescriptions for contraceptives would provide for a compromise between the two competing interests of the patient and the objecting pharmacist.\textsuperscript{267} A transfer requirement for pharmacists who conscientiously object to dispensing contraceptives would allow a pharmacist to maintain his or her ethical integrity while ensuring that a woman is able to exercise her fundamental right to self-determination in choosing to use contraceptives.\textsuperscript{268} The APhA’s testimony to the House Small Business Committee notes that pharmacies already commonly use such systems in order to accommodate the objecting pharmacist’s beliefs while still allowing the woman to access contraceptives.\textsuperscript{269}

\hspace{1cm}

\begin{itemize}
\item \textsuperscript{269} Hearing, supra note 194, app. at 62.
\end{itemize}
For example, a transfer requirement could require the pharmacist who objects to dispensing contraceptives to avoid doing so by transferring the prescription to another pharmacist within the same pharmacy who does not object to dispensing contraceptives.\textsuperscript{270} To facilitate this process and ensure access to contraceptives that must be taken within a few days of being prescribed, pharmacies would be required to have on staff at all times a pharmacist willing to dispense contraceptives.\textsuperscript{271} If, however, this is not possible because there are no other pharmacists within the pharmacy willing to dispense contraceptives, the pharmacist would then be required to transfer the customer’s prescription to a different pharmacy that is willing to dispense contraceptives.\textsuperscript{272}

A requirement that pharmacies have at least one pharmacist on staff at all times willing to dispense contraceptives may be especially important in rural areas where there may be no pharmacy within a short distance.\textsuperscript{273} Emergency contraception, in particular, must be taken within seventy-two hours to be most effective in preventing pregnancy.\textsuperscript{274} If transferring to another pharmacy would result in a delay of more than seventy-two hours before the patient could take the medication, the medication would become ineffective in preventing pregnancy.\textsuperscript{275}

Pharmacists and pharmacies conscientiously objecting to dispensing contraceptives should also be required to notify their customers of their policy so that customers will not mistakenly rely on a pharmacist’s or pharmacy’s willingness to dispense contraceptives.\textsuperscript{276} A notification requirement would require an individual pharmacist who conscientiously objects to dispensing contraceptives to inform his or her employer of such objections before he or she is permitted to refuse to fill any prescriptions.\textsuperscript{277} This would enable the employer to

\begin{footnotes}
\footnote{270} See id.
\footnote{271} See id.
\footnote{272} See id; Cantor & Baum, supra note 241, at 2011; Cohen, supra note 198, at 2–3.
\footnote{274} APhA Report, supra note 197, at 12; Cantor & Baum, supra note 241, at 2010; Cohen, supra note 198, at 2.
\footnote{275} See APhA Report, supra note 197, at 12; Cantor & Baum, supra note 241, at 2010; Cohen, supra note 198, at 2.
\end{footnotes}
develop a system so that the pharmacist’s objection could be accommodated without burdening any customers.\textsuperscript{278} For example, California requires the pharmacist to notify his or her employer, in writing, of the drugs that he or she objects to dispensing, before the pharmacist is permitted to conscientiously refuse to dispense these drugs.\textsuperscript{279} The pharmacy must then establish alternate means to ensure that the customer receives her prescriptions in a timely manner.\textsuperscript{280}

Additionally, a notice requirement as applied to pharmacies would require pharmacies that choose not to stock contraceptives to notify potential customers by displaying a sign that clearly indicates that they do not provide contraceptives.\textsuperscript{281} With such notification, customers would know prior to filling certain prescriptions that they cannot rely on this particular pharmacy.\textsuperscript{282} Thus, they would be able to find an alternate pharmacy before they need the particular medication, avoiding a conflict with that pharmacy, and balancing their rights against those of the pharmacy.\textsuperscript{283}

**Conclusion**

Refusal clauses were first enacted to ensure health care providers would not be forced to perform certain reproductive services, such as abortion, to which they were religiously, morally, or ethically opposed. Several factors, such as the growth of religious health care organizations and managed care, as well as the development of controversial medical technology, have led states to expand health care provider refusal laws. In particular, pharmacists, who are not protected under

\textsuperscript{278} See id.

\textsuperscript{279} Id.

\textsuperscript{280} Id.


most abortion and contraception refusal clauses, have pushed for a right to conscientiously object to dispensing contraception.

A few states have explicitly addressed a pharmacist’s right to conscientiously object. Most of this current legislation, however, fails to balance adequately a pharmacist’s right to conscientiously object against a woman’s right to access contraceptives. Court decisions and legislation addressing a similar conflict between health care providers and patients’ refusal of life-sustaining medical treatment provide guidance for future pharmacist refusal laws. The transfer and notice requirements imposed on conscientiously objecting health care providers in the life-sustaining treatment context should likewise be imposed on pharmacists who conscientiously object to dispensing contraceptives. This approach best balances the rights of both pharmacist and patient.

Natalie Langlois
A HOSTILE ENVIRONMENT: HOW THE “SEVERE OR PERVERSIVE” REQUIREMENT AND THE EMPLOYER’S AFFIRMATIVE DEFENSE TRAP SEXUAL HARASSMENT PLAINTIFFS IN A CATCH-22

Abstract: This Note argues that the combination of the “severe or pervasive” requirement and the employer’s affirmative defense, as applied in lower federal courts, makes it very difficult for a hostile work environment sexual harassment plaintiff to prevail against his or her employer. Courts require actionable harassment to consist of one extremely severe incident or to continue long enough to become cumulatively severe or pervasive. But once the harassment has gone on long enough to become severe or pervasive, the employer’s affirmative defense is increasingly likely to bar the plaintiff’s prima facie case. The result is that as the plaintiff’s prima facie case grows stronger, the probability that the employer will prevail on its affirmative defense also increases. This Note argues that such a contradictory approach is unfair to plaintiffs and proposes a more equitable standard that would place plaintiffs and employers on equal footing.

Introduction

Hostile work environment sexual harassment cases have troubled lower federal courts since the cause of action was first recognized by the U.S. Supreme Court in 1986.1 Two of the major issues have been (1) articulating the criteria for evaluating prima facie sexual harassment claims and identifying the appropriate threshold for an actionable environment, and (2) determining when to impute liability to an employer for the conduct of its supervisory employees.2 Courts have struggled with these issues and with achieving the goal of equal opportunities in the workplace for men and women—a goal that underlies Title VII of the Civil Rights Act of 1964.3

2 See infra notes 55–79 and accompanying text.
As the Supreme Court stated in its first hostile work environment sexual harassment case, conduct must be “severe or pervasive” to be actionable. The Court later clarified that the harassing conduct had to “create an objectively hostile or abusive work environment—an environment that a reasonable person would find hostile or abusive.” This broad language and the lack of a clear definition of “severe or pervasive” led lower federal courts to require varying degrees of abusiveness and persistence to find a hostile work environment.

The Supreme Court hoped to effectuate Title VII’s primary objective of preventing workplace discrimination by granting employers an affirmative defense to liability for the conduct of supervisory employees. The affirmative defense was designed to encourage employers to implement effective complaint mechanisms and grievance procedures. It requires an employer to prove that (1) the employer exercised reasonable care to prevent and correct workplace harassment, and (2) the victim unreasonably failed to take advantage of the preventive and corrective mechanisms established by the employer.

Lower federal courts have interpreted these requirements loosely. For the first prong, many courts require employers to show little more than promulgation of an anti-harassment policy that addresses sexual harassment and a grievance procedure that allows an employee to bypass a harassing supervisor. Some courts find that an employer has satisfied this requirement even when the employer’s response to the complaint fails to stop the harassment. Because many courts find that the existence of such a policy and procedure satisfies the duty of rea-

---

4 Meritor, 477 U.S. at 67.
5 Harris, 510 U.S. at 21.
6 See infra notes 80–158 and accompanying text.
8 See id.
9 Id. at 807; Burlington Indus., Inc. v. Ellerth, 524 U.S. 742, 765 (1998).
10 See infra notes 160–212 and accompanying text.
11 See, e.g., Faragher, 524 U.S. at 807–08 (finding against employer only because it failed to disseminate its policy against sexual harassment among employees, even though the policy did not include any assurance that a complainant could bypass the harassing supervisor, and also stating that demonstration that the plaintiff unreasonably failed to use any adequate complaint procedure provided by the employer will normally suffice to satisfy the employer’s burden); Madray v. Publix Supermarkets, Inc., 208 F.3d 1290, 1300 (11th Cir. 2000) (noting that, once an employer has promulgated an effective anti-harassment policy, it is incumbent upon the employees to utilize the procedural mechanisms established by the company).
12 See Crenshaw v. Delray Farms, Inc., 968 F. Supp. 1300, 1306–07 (N.D. Ill. 1997) (finding no evidence that employer should have been aware after the first of three incidents that a warning and reprimand would not sufficiently deter harasser).
sonable care to prevent and correct harassment under the first prong, an employee’s failure to use that policy and procedure is usually found to be unreasonable under the second prong.\textsuperscript{13} This is true even if the employee merely delays a few weeks in reporting the harassment.\textsuperscript{14}

These cases place an early burden on victims of sexual harassment to determine whether the conduct in question is serious enough to justify or require a complaint to their employer.\textsuperscript{15} An employee who is subjected to an isolated incident of trivial or low-level harassment may prefer to resolve the situation herself or simply look past it, rather than endure the complications and acrimony that often accompany the filing of a formal complaint.\textsuperscript{16} Such an employee does so at her own risk, however.\textsuperscript{17} If she does not report the incident, or merely delays in reporting it, and is later subjected to further harassment, many courts may bar her from asserting a cause of action because they will find that her failure to complain promptly following the first incident was unreasonable.\textsuperscript{18} In many cases, however, the same courts will also find that the initial incident was insufficiently severe to be actionable, thereby precluding her from establishing a cause of action even if she does report the incident.\textsuperscript{19} Therefore,

\textsuperscript{13} See, e.g., \textit{Faragher}, 524 U.S. at 807–08 (stating that demonstration that plaintiff unreasonably failed to use any complaint procedure provided by the employer will normally suffice to satisfy the employer’s burden); \textit{Ellerth}, 524 U.S. at 765 (same); \textit{Madray}, 208 F.3d at 1300.


\textsuperscript{15} See \textit{Faragher}, 524 U.S. at 807–08; \textit{Ellerth}, 524 U.S. at 765; \textit{Madray}, 208 F.3d at 1300.

\textsuperscript{16} See, e.g., \textit{Matvia v. Bald Head Island Mgmt., Inc.}, 259 F.3d 261, 270 (4th Cir. 2001) (noting that plaintiff claimed she reasonably feared retaliation by co-workers if she reported the harassment); \textit{Hanna v. Boys & Girls Home & Family Servs., Inc.}, 212 F. Supp. 2d 1049, 1069 (N.D. Iowa 2002) (finding a factual issue whether plaintiff was terminated in retaliation for complaining of sexual harassment); \textit{McGraw v. Wyeth-Ayerst Labs., Inc.}, No. CIV. A. 96-5780, 1997 WL 799437, at *8 (E.D. Pa. Dec. 30, 1997) (same); \textit{Pereira v. Schlage Elecs.}, 902 F. Supp. 1095, 1099, 1100–01 (N.D. Cal. 1995) (noting that after plaintiff’s complaint about harassment, co-workers threatened to kill her, burn down her house, and kidnap her and leave her where she would be raped and killed).

\textsuperscript{17} See \textit{Faragher}, 524 U.S. at 807–08; \textit{Ellerth}, 524 U.S. at 765; \textit{Madray}, 208 F.3d at 1300.

\textsuperscript{18} See \textit{Faragher}, 524 U.S. at 807–08; \textit{Ellerth}, 524 U.S. at 765; \textit{Madray}, 208 F.3d at 1300.

\textsuperscript{19} See, e.g., \textit{Faragher}, 524 U.S. at 788 (stating that isolated incidents, unless extremely severe, will not create an actionable environment); \textit{Alfano v. Costello}, 294 F.3d 365, 374 (2d Cir. 2002) (stating that a plaintiff must demonstrate either that a single incident was extraordinarily severe, or that a series of incidents was sufficiently continuous and con-
these courts require a plaintiff to report incidents of harassment that are likely not actionable and also find that such a plaintiff will have no cause of action even if she reports the harassment the moment it becomes actionable.  

This Note analyzes the relationship between the “severe or pervasive” standard in hostile work environment sexual harassment cases and the employer’s affirmative defense. Part I of this Note demonstrates how the combination of the “severe or pervasive” requirement and the employer’s affirmative defense makes it difficult for victims to successfully assert a cause of action.  

Section A of Part I reviews the Supreme Court’s decisions articulating the hostile work environment sexual harassment cause of action under Title VII of the Civil Rights Act of 1964 that established the “severe or pervasive” standard for an actionable environment and the affirmative defense to employer liability. Section B of Part I then examines lower federal courts’ interpretation and application of these precedents. This Section reviews cases finding that: (1) the plaintiff failed to allege severe or pervasive harassment sufficiently; (2) the harassment was sufficiently severe or pervasive; (3) the plaintiff’s failure to report the harassment promptly was unreasonable, in contravention of the affirmative defense’s mandate; and (4) although the plaintiff reported the harassment to the employer and no action was taken in response, the employer nevertheless can prevail on a motion for summary judgment.

Part II analyzes the tendency of many lower federal courts to hold plaintiffs to a strict standard while granting leniency to employers. This Part also addresses how courts place an unfair burden on sexual harassment victims to make an early prediction of whether asserted); Indest v. Freeman Decorating, Inc., 164 F.3d 258, 264 (5th Cir. 1999) (stating that actionable cases have involved patterns or allegations of extensive, long-lasting, unrepressed, and uninhibited sexual threats or conduct that permeated the plaintiffs’ work environment); McGraw, 1997 WL 799437, at *5 (stating that courts have required plaintiffs to show they have been subjected to continued explicit propositions, sexual epithets, or persistent offensive touchings).

20 See Faragher, 524 U.S. at 807 (stating that an employer will be subject to vicarious liability for an “actionable” hostile environment); Phillips v. Taco Bell Corp., 83 F. Supp. 2d 1029, 1033–34 (E.D. Mo. 2000) (granting summary judgment to employer where victim did not report an isolated incident that occurred three months prior to the majority of the harassment, but did promptly report the subsequent incidents).

21 See infra notes 29–53 and accompanying text.

22 See infra notes 29–54 and accompanying text.

23 See infra notes 55–79 and accompanying text.

24 See infra notes 80–212 and accompanying text.

25 See infra notes 80–212 and accompanying text.

26 See infra notes 213–325 and accompanying text.
ditional harassment will occur in order to preserve a cause of action.\textsuperscript{27} Part II concludes by proposing a standard that would balance the obligations of employees and employers in sexual harassment hostile work environment cases.\textsuperscript{28}

I. HOW THE “SEVERE OR PERVERSIVE” REQUIREMENT AND THE EMPLOYER’S AFFIRMATIVE DEFENSE TRAP

Plaintiffs in a Catch-22

The dilemma that plaintiffs face is most apparent in cases in which an employee initiallyexperiences low-level or isolated harassment that subsequently develops into more severe harassment.\textsuperscript{29} Such a victim’s predicament can be demonstrated by the following hypothetical. In an initial act of harassment, a female employee’s supervisor makes a lewd comment or touches her inappropriately.\textsuperscript{30} No further incidents occur for a few weeks or even months.\textsuperscript{31} Then, after an extended incident-free period, her supervisor resumes the inappropriate conduct. This time, the harassing behavior is more severe or frequent. The victim may decide that this conduct is too offensive to tolerate and report the harassment to her employer. As will be demonstrated, however, at this point she may already have forfeited any

\textsuperscript{27} See infra notes 213–325 and accompanying text.
\textsuperscript{28} See infra notes 326–353 and accompanying text.
\textsuperscript{29} See infra notes 30–54 and accompanying text.
\textsuperscript{30} Isolated or trivial incidents will almost categorically fail to constitute an actionable hostile work environment. See Faragher v. City of Boca Raton, 524 U.S. 775, 788 (1998) (stating that isolated incidents, unless extremely severe, will not create an actionable environment); Alfano v. Costello, 294 F.3d 365, 374 (2d Cir. 2002) (stating that a plaintiff must demonstrate either that a single incident was extraordinarily severe or that a series of incidents was sufficiently continuous and concerted); Indest v. Freeman Decorating, Inc., 164 F.3d 258, 264 (5th Cir. 1999) (stating that actionable cases have involved patterns or allegations of extensive, long-lasting, unredressed, and uninhibited sexual threats or conduct that permeated the plaintiffs’ work environment); McGraw v. Wyeth-Ayerst Labs., Inc., No. CIV. A. 96-5780, 1997 WL 799437, at *5 (E.D. Pa. Dec. 30, 1997) (stating that courts have required plaintiffs to show they have been subjected to continued explicit propositions, sexual epithets, or persistent offensive touchings).
\textsuperscript{31} Many courts will evaluate the initial incident in conjunction with any future incidents even if they are separated by a significant time interval. See, e.g., Conatzer v. Med. Prof’l Bldg. Servs. Corp., 95 F. App’x 276, 278 (10th Cir. 2004) (finding unreasonable delay in reporting harassment where victim did not report an incident where supervisor leaned up against her and rubbed against her side, but did report the incident seventeen days later promptly following a second, more serious incident); Phillips v. Taco Bell Corp., 83 F. Supp. 2d 1029, 1033–34 (E.D. Mo. 2000) (finding delay unreasonable where there were three months between the initial and subsequent incidents, even though all harassment was promptly reported following the subsequent incidents).
potential legal action because she failed to report the initial incident promptly.32

Following the initial incident, the victim must decide how to respond. She must determine whether the conduct was severe or pervasive, as well as predict whether additional harassment will occur.33 This preliminary response to the single incident is crucial to the victim’s future legal standing.34 Thus, the victim’s options are to report the conduct to her employer after the incident, attempt to resolve the situation on her own, or simply look past the incident, as courts may, as an isolated incident, in an effort to avoid the controversy likely to arise if she does report it.35

There are five common scenarios that follow from this situation:

1. Even if the victim reports the incident to her employer, at this point, the employer has no apparent legal obligation to take action in response.36 An employer will only be subject to vicarious liability for an actionable hostile work environment.37 To be actionable, the harassment must be severe or pervasive.38

2. If the harassment stops after this incident, the employer will not be liable regardless of its response.39 Assuming the initial

---

32 See infra notes 33–47 and accompanying text.
33 See Faragher, 524 U.S. at 807 (stating that an employer will be subject to vicarious liability for an “actionable” hostile environment); Phillips, 83 F. Supp. 2d at 1033–34 (granting summary judgment to employer where victim did not report an isolated incident that occurred three months prior to the majority of the harassment, but did promptly report the subsequent incidents).
34 See Conatzer, 95 F. App’x at 278, 281 (finding delay unreasonable where victim did not report an incident where supervisor leaned up against her and rubbed against her side, but did report the incident seventeen days later promptly following a second incident where he placed her in a headlock with his thighs); Phillips, 83 F. Supp. 2d at 1033–34.
35 See Faragher, 524 U.S. at 788; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; McGraw, 1997 WL 799437, at *5.
36 See Faragher, 524 U.S. at 807; Saidu-Kamara v. Parkway Corp., 155 F. Supp. 2d 436, 438, 440 (E.D. Pa. 2001) (granting summary judgment to employer where plaintiff’s supervisor, among other things, touched her breasts and buttocks, and employer took no action even though plaintiff repeatedly reported these incidents).
37 Faragher, 524 U.S. at 807; see Saidu-Kamara, 155 F. Supp. 2d at 438, 440.
38 Faragher, 524 U.S. at 807; see Saidu-Kamara, 155 F. Supp. 2d at 438, 440.
39 See Faragher, 524 U.S. at 807; Brooks v. City of San Mateo, 229 F.3d 917, 921, 926 (9th Cir. 2000) (finding work environment non-actionable where co-worker touched plaintiff’s stomach and then breast under her sweater); Simmons v. Mobile Infirmary Med. Ctr., 391 F. Supp. 2d 1124, 1133 (S.D. Ala. 2005) (finding work environment non-actionable where supervisor touched plaintiff’s breast four to five times, put his hands on her hips once, and brushed her leg once); Schaber-Goa v. Dep’t of Rehab. & Corr., No. 3:03CV7524, 2005 WL 1223891, at *5 (N.D. Ohio May 23, 2005) (finding work environment non-actionable
incident was insufficiently severe to be actionable, the employer is insulated from vicarious liability because the environment was never actionable.40

3. If the harassment continues, but never becomes severe or pervasive, the employer will not be liable regardless of its response.41 As in the second scenario, the employer cannot be liable because the victim cannot make out a prima facie case.42

4. If the harassment continues and eventually becomes severe or pervasive, the employer will not be liable if the victim failed to report the conduct promptly following the first incident.43 This is true even if she reported the conduct once it became severe or pervasive.44 The employer will succeed on the affirmative defense because the victim failed to report the harassment promptly—that is, promptly following the first incident, no matter how trivial or long ago.45

5. If the harassment continues and eventually becomes severe or pervasive, the employer may be liable if the employer reported the conduct within a reasonable time of the first incident and within a reasonable time following subsequent incidents.46 The employer, however, will only be liable if it

where plaintiff’s co-worker touched her crotch and grabbed her breast because the two incidents were “situationally isolated” and, therefore, not “sufficiently pervasive”); Saidu-Kamara, 155 F. Supp. 2d at 438, 440 (finding work environment non-actionable where plaintiff’s supervisor touched her breasts and buttocks).

40 See Faragher, 524 U.S. at 807; Saidu-Kamara, 155 F. Supp. 2d at 438, 440.

41 See Faragher, 524 U.S. at 807; Brooks, 229 F.3d at 921, 926; Simmons, 391 F. Supp. 2d at 1133; Schaber-Goa, 2005 WL 1223891, at *5; Saidu-Kamara, 155 F. Supp. 2d at 438, 440.

42 See Faragher, 524 U.S. at 807; Saidu-Kamara, 155 F. Supp. 2d at 438, 440.


44 See Phillips, 83 F. Supp. 2d at 1033–34.

45 See Conatzer, 95 F. App’x at 278 (finding delay unreasonable where victim did not report an incident where supervisor leaned up against her and rubbed against her side); Phillips, 83 F. Supp. 2d at 1033–34 (finding delay unreasonable where victim did not report incident that occurred three months prior to the majority of the harassment).

46 See Wyatt v. Hunt Plywood Co., 297 F.3d 405, 413–14 (5th Cir. 2002) (affirming partial summary judgment to employer where plaintiff complained promptly of supervisor’s conduct to a superior, but when the superior then began to harass plaintiff, she failed to complain promptly); Phillips, 83 F. Supp. 2d at 1033–34 (granting summary judgment to employer where victim did not report an isolated incident that occurred three months
failed to exercise reasonable care to prevent and promptly correct the harassment after it was reported.  

In order to establish a prima facie case, the harassment must either consist of an extremely severe incident or continue long enough to become severe or pervasive. Once the conduct becomes severe or pervasive, the employer’s affirmative defense still may serve to bar the plaintiff’s prima facie case. One isolated incident is not likely to be actionable, so a plaintiff could not assert a claim of severe or pervasive harassment. If, however, further incidents occur, sufficient to make the harassment actionable, then the plaintiff still may not prevail on her claim because the employer may succeed on its affirmative defense. Such an employee finds herself in a difficult situation. If she complains, she will only have a cause of action if the conduct is severe or pervasive. But if she does not complain, she will have no cause of action if the conduct eventually becomes severe or pervasive.

A. Supreme Court Decisions: Creation of the Hostile Work Environment Cause of Action and the Employer’s Affirmative Defense

In 1986, in Meritor Savings Bank v. Vinson, the U.S. Supreme Court first recognized hostile work environment sexual harassment as a type prior to the majority of the harassment, but did report promptly following the subsequent incidents).

---

47 See Cadena v. Pacesetter Corp., 224 F.3d 1203, 1209 (10th Cir. 2000) (affirming finding that employer could not satisfy affirmative defense where evidence showed less than reasonable efforts to remedy the harassment and the investigation was inadequate, if not a “complete sham”); White v. N.H. Dep’t of Corr., 221 F.3d 254, 261–62 (1st Cir. 2000) (affirming finding that employer could not satisfy affirmative defense where it did not handle the investigation properly or timely, and allowed the conduct to continue); Crenshaw v. Delray Farms, Inc., 968 F. Supp. 1300, 1306–07 (N.D. Ill. 1997) (finding no evidence that employer should have been aware after the first of three incidents that a warning and reprimand would not sufficiently deter harasser, when it took three separate warnings to end the harassment).

48 See, e.g., Faragher, 524 U.S. at 788; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264.

49 See Matvia v. Bald Head Island Mgmt., Inc., 259 F.3d 261, 269–70 (4th Cir. 2001) (rejecting plaintiff’s argument that she needed time to collect evidence against her supervisor so company officials would believe her, and rejecting argument that it was proper to refrain from reporting her supervisor so she could determine whether he was a “predator” or merely an “interested man” who could be politely rebuffed); Phillips, 83 F. Supp. 2d at 1033–34.

50 See, e.g., Faragher, 524 U.S. at 788; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264.

51 See Matvia, 259 F.3d at 269–70; Phillips, 83 F. Supp. 2d at 1033–34.

52 See Alfano, 294 F.3d at 374; Phillips, 83 F. Supp. 2d at 1033–34.

53 See Alfano, 294 F.3d at 374; Phillips, 83 F. Supp. 2d at 1033–34.

54 See Alfano, 294 F.3d at 374; Phillips, 83 F. Supp. 2d at 1033–34.
of discrimination based on sex.\textsuperscript{55} The Court found that the phrase “terms, conditions, or privileges of employment” in Title VII of the Civil Rights Act of 1964 “evinces a congressional intent to strike at the entire spectrum of disparate treatment of men and women in employment.”\textsuperscript{56} To be actionable, the Court held, the harassment must be sufficiently severe or pervasive to alter the conditions of the victim’s employment and create an abusive working environment.\textsuperscript{57}

The Court next addressed a hostile work environment claim in 1993, in \textit{Harris v. Forklift Systems, Inc.}\textsuperscript{58} The Court stated that an employer violates Title VII when the workplace permeates with discriminatory intimidation, ridicule, and insult, that is sufficiently severe or pervasive to alter the conditions of the victim’s employment and create an abusive working environment.\textsuperscript{59} The Court held that the harassment must create an objectively hostile or abusive work environment—one that a reasonable person would find hostile or abusive, although the victim need not suffer physical or psychological harm.\textsuperscript{60} The Court instructed lower courts to look at the totality of the circumstances in determining whether a hostile work environment exists.\textsuperscript{61} The Court provided a list of factors to consider, including the frequency of the discriminatory conduct; its severity; whether it is physically threatening or humiliating, or a mere offensive utterance; and whether it unreasonably interferes with an employee’s work performance.\textsuperscript{62} The Court further indicated that no single factor is required to create a hostile work environment.\textsuperscript{63}

Following \textit{Meritor} and \textit{Harris}, there was a general consensus among lower federal courts that with regard to harassment by co-workers, an employer would be vicariously liable if the employer knew or should have known of the harassment and failed to take prompt and effective remedial action.\textsuperscript{64} The standard for employer liability with respect to the conduct of supervisory employees, however, was

\textsuperscript{55} See 477 U.S. 57, 66–67 (1986).
\textsuperscript{56} Id. at 64 (quoting L.A. Dep’t of Water & Power v. Manhart, 435 U.S. 702, 707 (1978)).
\textsuperscript{57} Id. at 67.
\textsuperscript{58} See \textit{Harris v. Forklift Sys., Inc.}, 510 U.S. 17, 21–23 (1993).
\textsuperscript{59} Id. at 21.
\textsuperscript{60} Id. at 21–22.
\textsuperscript{61} Id. at 23.
\textsuperscript{62} Id.
\textsuperscript{63} \textit{Harris}, 510 U.S. at 23.
\textsuperscript{64} See, \textit{e.g.}, Burlington Indus., Inc. v. Ellerth, 524 U.S. 742, 759 (1998); Rabidue v. Osceola Refining Co., 805 F.2d 611, 621 (6th Cir. 1986).
Employers were not automatically liable for the conduct of supervisory employees. However, the mere existence of a grievance procedure and a policy against discrimination, coupled with the plaintiff’s failure to invoke that procedure, did not alone insulate an employer.

In 1998, in the companion cases of Faragher v. City of Boca Raton and Burlington Industries, Inc. v. Ellerth, the U.S. Supreme Court addressed the issue of employer liability for the conduct of supervisory employees. For any harassment not resulting in a tangible employment action—such as firing, demotion, or unfavorable reassignment—the plaintiff must prove that the conduct was sufficiently severe or pervasive to create an actionable environment. And when the victim suffers no tangible employment action an employer may raise an affirmative defense to liability for sexual harassment by supervisory employees.

To assert the affirmative defense, an employer must prove by a preponderance of the evidence that (1) the employer exercised reasonable care to prevent and promptly correct any sexually harassing behavior, and (2) the plaintiff unreasonably failed to take advantage of any preventive or corrective opportunities provided by the employer. The Court stated that, although the existence of an anti-harassment policy with a complaint procedure is not necessary in every instance as a matter of law, the need for such mechanisms may appropriately be considered when analyzing the steps taken by the employer to prevent and correct sexual harassment. And although proof that an employee failed to exert reasonable care to avoid harm is not limited to showing an unreasonable failure to use any complaint procedure provided by the employer, a demonstration of such failure will normally suffice to satisfy the employer’s burden under the second prong.

---

65 See Faragher, 524 U.S. at 785.
66 Meritor, 477 U.S. at 72.
67 Id.
68 Faragher, 524 U.S. at 775–811; Ellerth, 524 U.S. at 742–74.
69 Faragher, 524 U.S. at 786; see Ellerth, 524 U.S. at 761 (defining a tangible employment action as “a significant change in employment status, such as hiring, firing, failing to promote, reassignment with significantly different responsibilities, or a decision causing a significant change in benefits”).
70 Faragher, 524 U.S. at 807; Ellerth, 524 U.S. at 765.
71 Faragher, 524 U.S. at 807; Ellerth, 524 U.S. at 765.
72 Faragher, 524 U.S. at 807; Ellerth, 524 U.S. at 765.
73 Faragher, 524 U.S. at 807–08; Ellerth, 524 U.S. at 765.
cause it had failed to disseminate its policy to employees and the policy
did not ensure that the harassing supervisor could be bypassed when
employees registered complaints.74

The Faragher Court also discussed what types of conduct create a
hostile work environment.75 The Court stated that the standards for
judging hostility ensure that Title VII does not become a general civility
code for the workplace and that, properly applied, these standards will
filter out complaints attacking the ordinary tribulations of the work-
place, such as the sporadic use of abusive language, gender-related
jokes, and occasional teasing.76 Therefore, simple teasing, offhand
comments, and isolated incidents—unless extremely serious—will not
create an actionable environment.77 In a later case, the Court added
that hostile work environment claims are not based on discrete acts, but
rather, on incidents that occur over a series of days or perhaps years.78
Such claims are based on the cumulative effect of individual acts.79

B. Lower Federal Court Decisions: Defining the Actionable “Severe or
Pervasive” Threshold and When the Affirmative Defense Is Satisfied

1. Cases Finding Conduct Insufficiently Severe or Pervasive to
Constitute a Hostile Work Environment

Despite the seemingly reasonable threshold alluded to by the Su-
preme Court in Harris—severe or pervasive harassment by a reasonable
person’s standards—many lower federal courts began requiring an ar-
guably higher threshold than what a reasonable person would find hos-
tile or abusive.80 In 1997, in Crenshaw v. Delray Farms, Inc., the U.S. Dis-
trict Court for the Northern District of Illinois held the harassing
conduct at issue to be insufficiently severe or pervasive to be action-
able.81 In this case, a co-worker grabbed the plaintiff’s breast.82 A week
later, he told her he “needed somebody small like her to pick up in the
air and have sex with.”83 Shortly thereafter, he grabbed and squeezed

74 Faragher, 524 U.S. at 808.
75 See id. at 788.
76 Id.
77 Id.
79 Id.
80 See Harris, 510 U.S. at 21; infra notes 81–117 and accompanying text.
81 See 968 F. Supp. at 1306.
82 Id. at 1302.
83 Id.
her buttocks for several seconds and called her an “ignorant ass bitch.”\textsuperscript{84} Two months later, another co-worker came up behind her, rubbed his penis on her back and said “Oh, mamasita.”\textsuperscript{85} Two months after this incident, another co-worker asked the plaintiff if she cheated on her husband and told her that he would pay her to “be with him.”\textsuperscript{86} A few days later, the same co-worker grabbed her and attempted to kiss her.\textsuperscript{87} The next week, the plaintiff was in the women’s bathroom with the door locked when the same co-worker popped the lock open and entered.\textsuperscript{88} He attempted to grab the plaintiff, but she escaped.\textsuperscript{89} Two months after these incidents, another co-worker told the plaintiff that her pants “made his groins growl” and that he wanted to take her to a hotel.\textsuperscript{90} He told the plaintiff that he wanted to “eat [her] out” and “do all types of tricks with [her].”\textsuperscript{91} Finally, a month later, the plaintiff was surrounded by two supervisors and two managers all of whom yelled at her and threatened her job.\textsuperscript{92}

The court stated that the plaintiff must establish not only that \textit{she} was adversely affected by the conduct, but also that a \textit{reasonable} person would have been adversely affected.\textsuperscript{93} The court concluded that isolated instances of inappropriate conduct do not constitute sexual harassment and found that the offensive conduct at issue did not constitute a hostile work environment because it was not persistent.\textsuperscript{94} The court also found that the employer appropriately dealt with the perpetrators, even though the plaintiff was harassed by at least six different co-workers and supervisors on several occasions.\textsuperscript{95} In one instance, it took two warnings for a co-worker to stop harassing her, and in another case, it took three warnings to end the harassment.\textsuperscript{96}

In 1996, in \textit{Hannigan-Haas v. Bankers Life & Casualty Co.}, the U.S. District Court for the Northern District of Illinois found that a serious physical assault was insufficient as a matter of law to create a hostile

\textsuperscript{84} Id.
\textsuperscript{85} Id.
\textsuperscript{86} Crenshaw, 968 F. Supp. at 1302.
\textsuperscript{87} Id.
\textsuperscript{88} Id. at 1303.
\textsuperscript{89} Id.
\textsuperscript{90} Id.
\textsuperscript{91} Crenshaw, 968 F. Supp. at 1303.
\textsuperscript{92} Id.
\textsuperscript{93} Id. at 1305–06.
\textsuperscript{94} Id. at 1306.
\textsuperscript{95} Id. at 1301–03.
\textsuperscript{96} Crenshaw, 968 F. Supp. at 1306.
work environment because it was only one incident. \footnote{97}{See No. 95 C 7408, 1996 WL 650419, at *5 (N.D. Ill. Nov. 6, 1996).} One of the plaintiff’s supervisors, a senior vice president of the company, approached her and asked her to accompany him into an office. \footnote{98}{Id. at *1.} Once there, he closed the door and, unbeknownst to the plaintiff, locked it. \footnote{99}{Id.} When she attempted to leave the room, he stood up, pushed her against a wall, leaned his chest against hers, and said “open your mouth” as he attempted to kiss her with an open mouth. \footnote{100}{Hannigan-Haas, 1996 WL 650419, at *1.} She turned her head to avoid the kiss. \footnote{101}{Id.} He then tried to touch her breasts and to place his hands under her pantyhose. \footnote{102}{Id.} The plaintiff was able to break free and ran from the room. \footnote{103}{Id. at *5.} Despite the serious nature of the assault, the court found that it was insufficiently severe or pervasive to be actionable because it was only a single incident. \footnote{104}{See 51 F. Supp. 2d 858, 864 (N.D. Ohio 1999).}

In 1999, in \textit{Blough v. Hawkins Market, Inc.}, another case involving serious physical contact, the U.S. District Court for the Northern District of Ohio found the challenged conduct insufficiently severe or pervasive to establish a hostile work environment. \footnote{105}{Id. at 862.} The plaintiff’s co-worker patted her on the buttocks and, several months later, grabbed her crotch. \footnote{106}{Id.} Another co-worker attempted to kiss her and apparently engaged in self-stimulation while she was looking. \footnote{107}{Id.} The court concluded that a hostile environment requires more than a few isolated incidents of offensive conduct. \footnote{108}{Id.}

In 1995, in \textit{Blankenship v. Parke Care Centers, Inc.}, the U.S. District Court for the Southern District of Ohio found the following harassment to be insufficiently severe or pervasive to be actionable. \footnote{109}{See 913 F. Supp. 1045, 1055 (S.D. Ohio 1995).} Two seventeen-year-old female employees were harassed by a forty-year-old male co-worker. \footnote{110}{Id.} During a six-day period, the first plaintiff endured the co-worker blowing kisses at her, licking his lips, and smiling at her seductively or perversely on several occasions; making obscene gestures toward his crotch on one occasion; touching her chest above her
breast with his finger on one occasion; and making vulgar sexual remarks in her presence about another female co-worker on one occasion.\textsuperscript{111} The second plaintiff endured similar harassment.\textsuperscript{112} The co-worker once suggested that he was “falling in love” with her, tickled her on another occasion, and kissed her on the cheek or hugged her on four occasions.\textsuperscript{113} On yet another occasion, he approached her from behind and lifted her breasts.\textsuperscript{114} He also repeatedly asked her to go on a date.\textsuperscript{115}

The court reasoned that although the facts suggested that these particular plaintiffs subjectively perceived this conduct to affect the conditions of their employment, the likely reaction of a “hypothetical reasonable person” was less clear.\textsuperscript{116} The court went on to conclude that neither of the plaintiffs’ contentions definitively met the “severe or pervasive” standard, and thus neither was actionable.\textsuperscript{117}

2. Cases Finding That Conduct Was Sufficiently Severe or Pervasive to Create a Hostile Work Environment

Two cases from the U.S. Court of Appeals for the Seventh Circuit illustrate the magnitude of harassment required by some courts to establish that conduct was sufficiently severe or pervasive to be actionable.\textsuperscript{118} In 2004, in \textit{McPherson v. City of Waukegan}, the Seventh Circuit held that the two most severe incidents of the following conduct were sufficient to create a hostile work environment.\textsuperscript{119} The plaintiff stated that her supervisor asked her what color bra she was wearing a “handful” of times.\textsuperscript{120} Once, when the plaintiff called in sick, her supervisor responded by suggestively asking if he could “make a house call.”\textsuperscript{121} He asked the same question two or three more times during the course of her employment.\textsuperscript{122} On another occasion, at a time when they were unobserved by others, he asked her what color bra she was wearing and then pulled back her tank top to see for him-

\textsuperscript{111} Id. at 1053.
\textsuperscript{112} Id. at 1053–54.
\textsuperscript{113} Id. at 1054.
\textsuperscript{114} \textit{Blankenship}, 913 F. Supp. at 1054.
\textsuperscript{115} Id.
\textsuperscript{116} Id. at 1055.
\textsuperscript{117} Id.
\textsuperscript{118} See infra notes 119–158 and accompanying text.
\textsuperscript{119} See 379 F.3d 430, 439 (7th Cir. 2004).
\textsuperscript{120} Id. at 434.
\textsuperscript{121} Id. at 435.
\textsuperscript{122} Id.
self.\textsuperscript{123} About a month later, he called the plaintiff into his office.\textsuperscript{124} As they were discussing a work document, he slid his hand under her shirt and felt her breasts.\textsuperscript{125} Five days later, he called her into his office again and asked her to shut the door.\textsuperscript{126} When she turned to leave, he approached her and slid his hand under her shirt, touching her breasts.\textsuperscript{127} The plaintiff asked him to stop, but he pushed her toward the wall behind his office door.\textsuperscript{128} He then put his hand down her pants and inserted his finger into her vagina.\textsuperscript{129}

The court found that all of the conduct prior to the two physical assaults was merely boorish behavior that was not severe or pervasive.\textsuperscript{130} The court concluded that the earlier behavior was “lamentably inappropriate,” but insufficient due to its limited nature and infrequency.\textsuperscript{131} The court did find, however, that the two instances of physical assault were sufficient to constitute a hostile work environment.\textsuperscript{132}

Similarly, in 1999, in \textit{Wilson v. Chrysler Corp.}, the Seventh Circuit held the following harassment to create an actionable work environment.\textsuperscript{133} The plaintiff worked on the assembly line at an automobile plant.\textsuperscript{134} Over the course of approximately eight years, she endured repeated acts of harassment involving improper touching, verbal abuse, and the display of offensive objects, cartoons, and pictures.\textsuperscript{135} Several of the plaintiff’s complaints involved the use or display of fake rubber penises.\textsuperscript{136} At least twice a week, for the duration of her employment, these fake penises were sent down the assembly line past her workstation.\textsuperscript{137} On one occasion, a co-worker brandished a fake penis between his legs and yelled at the plaintiff, “Look what I got for you; bet you can’t handle this.”\textsuperscript{138} Three other times, fake penises were left on a car in her work area.\textsuperscript{139} She also related two instances of offensive touch-

\begin{itemize}
  \item \textsuperscript{123} \textit{Id.}
  \item \textsuperscript{124} \textit{McPherson, 379 F.3d at 435.}
  \item \textsuperscript{125} \textit{Id.}
  \item \textsuperscript{126} \textit{Id.}
  \item \textsuperscript{127} \textit{Id.}
  \item \textsuperscript{128} \textit{Id.}
  \item \textsuperscript{129} \textit{McPherson, 379 F.3d at 435.}
  \item \textsuperscript{130} \textit{Id. at 439.}
  \item \textsuperscript{131} \textit{Id.}
  \item \textsuperscript{132} \textit{Id.}
  \item \textsuperscript{133} \textit{See 172 F.3d 500, 511 (7th Cir. 1999).}
  \item \textsuperscript{134} \textit{Id. at 507.}
  \item \textsuperscript{135} \textit{Id.}
  \item \textsuperscript{136} \textit{Id.}
  \item \textsuperscript{137} \textit{Id.}
  \item \textsuperscript{138} \textit{Wilson, 172 F.3d at 507.}
  \item \textsuperscript{139} \textit{Id.}
\end{itemize}
A co-worker would regularly stand close to her, look at her breasts, and say “umm.” He would deliberately position his hand so that it would come into contact with her breasts when she bent down. Another co-worker flicked his finger on her breast. She also cited several humiliating exchanges involving verbal abuse. One co-worker addressed her by saying “Hi, ho,” meaning “whore.” Another accused her of “sucking a white boy's dick on the roof.” Another commented, “I can make your pussy bloom.” Yet another co-worker told her that white men only spoke to her because they “wanted her tutu.”

More frequently, the plaintiff was taunted with offensive literature. A co-worker handed her a paper that pontificated on why beer is better than women. Another co-worker showed her a picture of a nude woman and stated, “Too bad you don’t have a body like this.” On another occasion, a photograph of a nude man was taped to the hood of a car on which she was working. She also found an offensive cartoon taped to her work station. Sexually explicit cartoons were handed to her or displayed in her work area. Finally, while on medical leave, she received a lewd greeting card that had been signed by approximately thirty-eight employees at the plant including a supervisor and three foremen.

The court noted the multifaceted nature of the harassment, its frequent and at times routine character, and the participation of so many members of the workforce. Considering the cumulative effect of the harassment, the court noted that the plaintiff depicted a workplace in which harassment rose almost to the level of an institutional

\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
\[\text{id.}\]
The court concluded that the plaintiff had stated an actionable claim for an objectively hostile work environment.158

3. Cases Finding the Plaintiff’s Failure to Report Unreasonable

In addition to holding plaintiffs to a high threshold for satisfying the “severe or pervasive” standard, many courts also dismiss plaintiffs’ cases for failure to report trivial or isolated incidents through the employer’s affirmative defense.159 In 2000, in Phillips v. Taco Bell Corp., the U.S. District Court for the Eastern District of Missouri granted the employer’s motion for summary judgment in a case in which the plaintiff was initially subjected to one isolated incident of offensive conduct by her supervisor.160 No other incidents of harassment occurred until three months later, when she was subjected to four incidents in a seven-day period.161 The initial incident involved the supervisor squeezing the plaintiff’s breasts.162 Three months later, he grabbed the front of her pants; the next day, he grabbed her hand and placed it on the front of his pants; four days later, he rubbed his hands across her buttocks; and the next day, he rubbed his hands across and down her back.163 Two days after this final incident, the plaintiff reported the offensive conduct to her employer.164

The court found that, by not reporting the offensive conduct until two days after the final incident, the plaintiff had unreasonably failed to take advantage of any preventive or corrective opportunities provided by her employer.165 The court considered all five incidents together and found that the plaintiff had waited over three months to report the harassment.166 The court concluded that, although the five incidents were sufficiently severe or pervasive, the employer had successfully proven its affirmative defense to liability and was entitled to summary judgment.167

Similarly, in 2004, in Conatzer v. Medical Professional Building Services Corp., the Tenth Circuit affirmed a grant of summary judgment

157 Id.
158 Wilson, 172 F.3d at 511.
159 See infra notes 160–191 and accompanying text.
160 83 F. Supp. 2d at 1033–34.
161 Id.
162 Id.
163 Id.
164 Id. at 1033.
165 Phillips, 83 F. Supp. 2d at 1034.
166 Id.
167 Id. at 1033–34.
for the employer in a case in which the plaintiff alleged that her supervisor had sexually harassed her on two occasions. The first incident involved physical contact between the plaintiff and her supervisor, described by the shift supervisor as the plaintiff’s supervisor stepping up to the plaintiff, leaning against her, and rubbing against the side of her chest. Two weeks later, as the plaintiff bent over to pick something up, her supervisor briefly placed her in a headlock with his thighs. The plaintiff registered a complaint four or five days after the second incident, or seventeen days after the first incident.

The court concluded that, absent an adequate explanation for this delay, the plaintiff had acted unreasonably. Furthermore, the court rejected the argument that because the shift supervisor witnessed the first incident, the employer was obligated to respond prior to the plaintiff’s formal complaint. The court also stated that there was no indication that the shift supervisor observed anything other than an isolated incident.

Similarly, in 2001, in Matvia v. Bald Head Island Management, Inc., the Fourth Circuit affirmed an employer’s motion for summary judgment. Beginning in September 1997, the plaintiff’s supervisor approached her, said he needed a hug, and proceeded to hug her. He told the plaintiff, after she had just dyed her hair brown, that he would now have to fantasize about a brunette rather than a blond. He informed her that he no longer had sexual relations with his wife. He placed a pornographic picture on her desk. He told her she looked good enough to eat. He frequently placed his arm around her and massaged her shoulder. He also repeatedly told her that he loved her and had a crush on her. In December 1997, he

\[\text{References:}\]

168 95 F. App’x at 278, 281.
169 Id. at 278.
170 Id.
171 Id.
172 Id. at 281.
173 Conatzer, 95 F. App’x at 281.
174 Id.
175 259 F.3d at 270.
176 Id. at 265.
177 Id.
178 Id.
179 Id.
180 Matvia, 259 F.3d at 265.
181 Id.
182 Id.
told her that he dreamt that she sued him for sexual harassment and warned her that if she did bring suit, she would be in big trouble. The court rejected the plaintiff’s argument that she needed time to collect evidence against her supervisor so that company officials would believe her. The court also rejected her argument that it was proper for her to refrain from reporting her supervisor so she could determine whether he was a “predator” or merely an “interested man” who could be politely rebuffed. The court stated that case law makes no distinction between “predators” and “interested men,” and the label given to the harasser is immaterial. The court concluded that the gravity and frequency of the incidents made it clear that the supervisor was not merely an interested man who could be politely rebuffed. Although the plaintiff explained that she did not report the earlier conduct because she was not sure if she was being sexually harassed and whether the company would believe her, the court affirmed the employer’s motion for summary judgment. The court, with the benefit of hindsight, concluded that the plaintiff should have found the earlier conduct sexually harassing and reported it.

4. Cases Dismissing Plaintiff’s Claim Despite Complaint Made to Employer and Employer’s Refusal to Take Action in Response

In addition to granting employers’ motions for summary judgment when the plaintiff failed to report or delayed in reporting harassment, some courts granted these motions even when the plaintiff did complain and the employer ignored it. In 2002, in Wyatt v. Hunt Plywood Co., the Fifth Circuit affirmed the employer’s motion for summary judgment for the majority of the following harassment. In

183 Id.
184 Id.
185 Matvia, 259 F.3d at 270. The opinion does not state how the employer came to learn of the harassment. See id. at 261–73.
186 Id. at 269.
187 Id.
188 Id.
189 Id. at 270.
190 Matvia, 259 F.3d at 269–70.
191 See id.
192 See infra notes 193–212 and accompanying text.
193 See 297 F.3d 405, 414 (5th Cir. 2002).
this case, the plaintiff was sexually harassed both by her immediate supervisor and the supervisor to whom she complained. The plaintiff’s immediate supervisor began harassing her sexually, referring to her in vulgar terms and continually asking her to have sex with him. She promptly complained of this harassment to her supervisor’s superior, but despite her complaints, the harassment persisted. Instead of remedying the problem, eventually the superior himself subjected the plaintiff to sexual advances and harassment. The plaintiff conceded, however, that she never reported this conduct to anyone higher up in the management chain. The original supervisor’s harassment reached its peak when he snuck up behind her and pulled down her sweat pants in the plain view of other employees. She immediately complained to the superior again, who agreed to write an incident report. In an attempt to downplay the harassment, however, he omitted the incident where the plaintiff’s pants were pulled down.

The court found that once it became clear not only that the superior was ineffective in dealing with the harassment, but also that he himself was a sexual harasser, the plaintiff’s failure to report either harasser’s behavior to a different supervisor was unreasonable. Furthermore, the court held that the plaintiff’s reliance on the superior’s admonitions not to “go over his head” did not excuse her failure to disclose the harassment to a higher authority. The court found the employer’s sexual harassment policy and its implementation of the policy “more than adequate,” and did not consider that the most serious incident was intentionally omitted from the formal report in an effort to downplay the harassment.

The only portion of the harassment for which the employer could not assert the affirmative defense was the brief period that began with the plaintiff’s first report to the superior and ended when he

---

194 Id. at 407.
195 Id.
196 Id.
197 Id.
198 Wyatt, 297 F.3d at 407.
199 Id.
200 Id.
201 Id. at 407–08.
202 Id. at 413.
203 Wyatt, 297 F.3d at 413.
204 Id.
himself began harassing her.\textsuperscript{205} The court affirmed the employer’s motion for summary judgment regarding all other incidents of harassment.\textsuperscript{206}

In 2001, in \textit{Saidu-Kamara v. Parkway Corp.}, the U.S. District Court for the Eastern District of Pennsylvania granted an employer’s motion for summary judgment in a case in which the plaintiff alleged that she was subjected to various forms of discrimination and harassment throughout her employment.\textsuperscript{207} Her supervisor asked her out on dates on several occasions, directed sexual innuendos towards her, and, on at least one occasion, touched her breasts and buttocks.\textsuperscript{208} The plaintiff repeatedly reported these incidents to the facility manager, but no action was taken.\textsuperscript{209} The court stated that, considering the totality of the circumstances, the plaintiff had failed to demonstrate sufficiently severe or pervasive discrimination.\textsuperscript{210} Therefore, the court concluded, the plaintiff had not satisfied a necessary element of her hostile work environment claim and it granted the employer’s motion for summary judgment.\textsuperscript{211} In reaching its conclusion, the court did not address the fact that the plaintiff had repeatedly reported these incidents and that no action had been taken in response.\textsuperscript{212}

\section*{II. Analysis of Lower Federal Court Practices and a Proposal for a More Equitable and Consistent Standard}

As the foregoing cases illustrate, the legal standard applied to hostile work environment sexual harassment cases is inconsistent and self-contradictory.\textsuperscript{213} As demonstrated, the “severe or pervasive” stan-

\textsuperscript{205} Id. at 414.
\textsuperscript{206} Id.
\textsuperscript{208} Id.
\textsuperscript{209} Id.
\textsuperscript{210} Id. at 440.
\textsuperscript{211} Id.
\textsuperscript{212} \textit{See} \textit{Saidu-Kamara}, 155 F. Supp. 2d at 439–40.
\textsuperscript{213} \textit{See} \textit{Faragher v. City of Boca Raton}, 524 U.S. 775, 788 (1998) (stating that isolated incidents, unless extremely severe, will not create an actionable environment); \textit{Alfano v. Costello}, 294 F.3d 365, 374 (2d Cir. 2002) (stating that a plaintiff must demonstrate either that a single incident was extraordinarily severe, or that a series of incidents was sufficiently continuous and concerted); \textit{Matvia v. Bald Head Island Mgmt., Inc.}, 259 F.3d 261, 269–70 (4th Cir. 2001) (rejecting plaintiff’s argument that she needed time to collect evidence against her supervisor so company officials would believe her, and rejecting argument that it was proper to refrain from reporting her supervisor so she could determine whether he was a “predator” or merely an “interested man” who could be politely rebuffed); \textit{Phillips v. Taco Bell Corp.}, 83 F. Supp. 2d 1029, 1033–34 (E.D. Mo. 2000) (granting summary judgment to employer where victim did not report incident that occurred three months prior
standard is a very high threshold in many jurisdictions. A plaintiff can theoretically meet this threshold by showing one extremely severe incident or a pattern of extensive, continuous, and long-lasting harassment. Because most plaintiffs fail to show that one incident was sufficiently severe to satisfy the standard, they ordinarily must present a case involving several incidents occurring within a concentrated period of time.

A plaintiff who is able to establish a prima facie case must then get past the obstacle of the employer’s affirmative defense. The paradox, however, is that as the plaintiff’s prima facie case grows stronger, the probability that the employer will prevail on its affirmative defense also increases. Thus, by the time the harassment becomes severe or pervasive, the plaintiff has likely waited too long to complain. If the plaintiff was subjected to sufficiently severe or pervasive harassment, then the court looks to whether she filed formal complaints following each incident. If she has not, the employer to the majority of the harassment, but did report promptly following the subsequent incidents); supra notes 80–212 and accompanying text.

See supra notes 80–158 and accompanying text.

See Faragher, 524 U.S. at 788; Alfano, 294 F.3d at 374; Indest v. Freeman Decorating, Inc., 164 F.3d 258, 264 (5th Cir. 1999) (stating that actionable cases have involved patterns or allegations of extensive, long-lasting, unredressed, and uninhibited sexual threats or conduct that permeated the plaintiffs’ work environment).

See, e.g., Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; McGraw v. Wyeth-Ayerst Labs., Inc., No. CIV. A. 96-5780, 1997 WL 799437, at *5 (E.D. Pa. Dec. 30, 1997) (stating that courts have required plaintiff to show she has been subjected to continued explicit propositions or sexual epithets or persistent offensive touchings).


See Faragher, 524 U.S. at 788 (stating that isolated incidents, unless extremely severe, will not create an actionable environment); Conatzer v. Med. Prof’l Bldg. Servs. Corp., 95 F. App’x 276, 278 (10th Cir. 2004) (finding delay unreasonable where victim did not report an isolated incident where supervisor leaned up against her and rubbed against her side, but did report the incident seventeen days later promptly following a second, more serious incident); Alfano, 294 F.3d at 374 (stating that a plaintiff must demonstrate either that a single incident was extraordinarily severe, or that a series of incidents was sufficiently continuous and concerted); Indest, 164 F.3d at 264 (stating that actionable cases have involved patterns or allegations of extensive, long-lasting, unredressed, and uninhibited sexual threats or conduct that permeated the plaintiffs’ work environment); Phillips, 83 F. Supp. 2d at 1033–34 (finding an unreasonable delay where victim did not report an isolated incident that occurred three months prior to the majority of the harassment, but did report promptly following the subsequent incidents).

See Faragher, 524 U.S. at 788; Conatzer, 95 F. App’x at 278; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; Phillips, 83 F. Supp. 2d at 1033–34.

See Conatzer, 95 F. App’x at 278 (finding delay unreasonable where victim did not report initial incident, but did report promptly following a second incident); Wyatt v. Hunt Plywood Co., 297 F.3d 405, 413–14 (5th Cir. 2002) (affirming partial summary judgment
will likely be able to assert the affirmative defense on the grounds that the plaintiff failed to report the harassment promptly. Essentially, a victim of sexual harassment can be stripped of her ability to assert a cause of action even before she has an actionable claim.

A. Lower Federal Courts Interpret the Same Conduct Differently When Evaluating the Plaintiff’s Prima Facie Case and the Employer’s Affirmative Defense

Many lower federal courts have set the “severe or pervasive” threshold for establishing a hostile work environment far too high. The Supreme Court stated that to meet the standard, the harassing conduct must “create an objectively hostile or abusive work environment—an environment that a reasonable person would find hostile or abusive.” A reasonable person would likely feel sexually harassed in some of the “nonactionable” cases set out above, many of which involved lewd comments and physical touching of intimate body parts. When analyzing a plaintiff’s prima facie case, however, many courts tend to downplay the severity of the harassment and the impact it would have on a reasonable person. Conversely, when analyzing the employer’s affirmative defense, courts characterize trivial harassment as significant, concluding that

---

221 See Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033–34.
222 See Faragher, 524 U.S. at 788; Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033–34.
223 See supra notes 80–158 and accompanying text.
225 See supra notes 80–117 and accompanying text.
226 See DeAngelis v. El Paso Mun. Police Officers Assoc., 51 F.3d 591, 593 (5th Cir. 1995) (stating that the actionable standard requires “conduct so egregious as to alter the conditions of employment and destroy [women’s] equal opportunity in the workplace”); Baskerville v. Culligan Int’l Co., 50 F.3d 428, 430–31 (7th Cir. 1995) (stating that the concept of sexual harassment is designed to protect women from the kind of behavior that can make the workplace “hellish” for women, and concluding that “only a woman of Victorian delicacy—a woman mysteriously aloof from contemporary American popular culture in all its sex-saturated vulgarity”—would find the complained-of behavior distressing); Hosey v. McDonald’s Corp., Civ. No. AW-95-196, 1996 WL 414057, at *1–2 (D. Md. May 17, 1996) (describing several unwelcome sexual advances, obscene comments, and approximately ten incidents of offensive touching as “teenagers . . . asking each other for dates”); supra notes 80–117 and accompanying text. The court in DeAngelis reasoned that “a less onerous standard of liability would attempt to insulate women from everyday insults as if they remained models of Victorian reticence,” and that “[a] lesser standard of liability would mandate not equality but preference for women.” 51 F.3d at 593.
the plaintiff acted unreasonably by not taking the conduct more seriously and filing a complaint with her employer.\textsuperscript{227} This not only makes it more difficult for a plaintiff to establish severe or pervasive harassment, but it also makes it more difficult to defeat the employer’s affirmative defense.\textsuperscript{228}

Furthermore, this approach forces an employee who is subjected to one isolated incident of trivial harassment to treat the incident as if it were much more serious.\textsuperscript{229} Although such an incident may be a minor factor in the courts’ prima facie case analysis, it could play a decisive role in the affirmative defense analysis if a court finds that this incident was the first episode in a pattern of harassment that the plaintiff unreasonably failed to report promptly.\textsuperscript{230} This is true even if the incident was of negligible significance or was the only offensive act to occur for months.\textsuperscript{231}

Hence, when many courts look at a plaintiff’s prima facie case they ordinarily declare that incidents of harassment occurring weeks apart were too isolated to be actionable.\textsuperscript{232} These courts are willing to conclude that multiple incidents were merely “isolated,” and therefore non-actionable, even when the conduct involved was reasonably severe.\textsuperscript{233} When evaluating the employer’s affirmative defense, however, the courts do not discuss isolated incidents or whether every incident mentioned in the plaintiff’s complaint should be part of the

\textsuperscript{227} See supra notes 160–191 and accompanying text.

\textsuperscript{228} See supra notes 80–117, 160–191 and accompanying text.

\textsuperscript{229} See Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033–34; supra notes 160–191 and accompanying text.

\textsuperscript{230} See Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033–34; supra notes 80–117, 160–191 and accompanying text.

\textsuperscript{231} See Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033–34; supra notes 80–117, 160–191 and accompanying text.

\textsuperscript{232} See, e.g., Schaber-Goa v. Dep’t of Rehab. & Corr., No. 3:03CV7524, 2005 WL 1223891, at *5 (N.D. Ohio May 23, 2005) (finding work environment non-actionable where plaintiff’s co-worker touched her crotch and grabbed her breast because the two incidents were “situationally isolated” and, therefore, not “sufficiently pervasive”); Blough v. Hawkins Mkt., Inc., 51 F. Supp. 2d 858, 862, 864 (N.D. Ohio 1999) (finding work environment non-actionable where one co-worker patted the plaintiff on the buttocks and grabbed her crotch, while another co-worker attempted to kiss her and apparently engaged in self-stimulation while she was looking, because the court concluded these were merely isolated incidents of offensive conduct); Crenshaw v. Delray Farms, Inc., 968 F. Supp. 1300, 1301–03, 1306 (N.D. Ill. 1997) (finding work environment non-actionable where plaintiff was harassed by at least six different co-workers and supervisors on at least ten different occasions because these were merely isolated incidents that were not persistent); supra notes 80–117 and accompanying text.

\textsuperscript{233} See, e.g., Schaber-Goa, 2005 WL 1223891, at *5; Blough, 51 F. Supp. 2d at 862, 864; Crenshaw, 968 F. Supp. at 1301–03, 1306; supra notes 80–117 and accompanying text.
same analysis. Rather, in determining whether the plaintiff acted reasonably, many courts look only at the interval between the first incident mentioned in the complaint and the date she reported the harassment.

This creates a substantial problem for a plaintiff who suffered sufficiently severe or pervasive harassment, but was also subjected to a trivial or isolated incident occurring much earlier. Although the initial incident would be too isolated to support a prima facie case, courts routinely use this same incident to find that the plaintiff unreasonably failed to report the harassment promptly and thereby allow the employer to assert the affirmative defense. These courts treat the same incidents and the same time interval differently when analyzing the prima facie case and the affirmative defense. The usual result of this practice is a grant of summary judgment for the employer.

If the plaintiff was harassed by multiple co-workers or supervisors, courts ordinarily require that the victim report each succeeding incident of harassment and each succeeding harasser. Employers are only required to respond to the incidents of harassment that are reported. They are not required to end the harassment, either by the

234 See, e.g., Conatzer, 95 F. App’x at 281; Phillips, 83 F. Supp. 2d at 1034; supra notes 160–191 and accompanying text.
235 See, e.g., Conatzer, 95 F. App’x at 281; Phillips, 83 F. Supp. 2d at 1034; supra notes 160–191 and accompanying text.
236 See, e.g., Conatzer, 95 F. App’x at 281; Phillips, 83 F. Supp. 2d at 1034; supra notes 80–117, 160–191 and accompanying text.
238 See, e.g., Conatzer, 95 F. App’x at 281; Schaber-Goa, 2005 WL 1223891, at *5; Phillips, 83 F. Supp. 2d at 1034; Blough, 51 F. Supp. 2d at 862, 864; Crenshaw, 968 F. Supp. at 1301–03, 1306; supra notes 80–117, 160–191 and accompanying text.
239 See, e.g., Conatzer, 95 F. App’x at 281 (granting employer’s motion for summary judgment because plaintiff failed to report harassment promptly); Schaber-Goa, 2005 WL 1223891, at *5 (granting employer’s motion for summary judgment because incidents were isolated); Phillips, 83 F. Supp. 2d at 1034 (granting employer’s motion for summary judgment because plaintiff failed to report harassment promptly); Blough, 51 F. Supp. 2d at 864 (granting employer’s motion for summary judgment because incidents were isolated); Crenshaw, 968 F. Supp. at 1306 (same); supra notes 80–117, 160–191 and accompanying text.
240 See Wyatt, 297 F.3d at 413; Blough, 51 F. Supp. 2d at 862, 864; Crenshaw, 968 F. Supp. at 1301–03, 1306; supra notes 80–117, 160–191 and accompanying text.
harasser who is reported or by subsequent harassers. This low expectation for employers does a disservice to the Supreme Court’s language in Faragher v. City of Boca Raton and Burlington Industries, Inc. v. Ellerth that employers must exercise “reasonable care to prevent and correct promptly any sexually harassing behavior.”

B. Plaintiffs Are Forced to Make an Early Prediction of Whether an Incident of Harassment Is an Isolated Incident or the First in a Chain of Harassment

Although the Supreme Court stated that lower federal courts should look to the cumulative effect of the harassment to judge its severity and pervasiveness, it gave no guidance to victims as to what types of conduct they must report to preserve a future cause of action. The Court stated that hostile work environment claims are not based on discrete acts, but rather on incidents that occur over a series of days, months, or perhaps years. Although courts have the luxury of viewing the whole record of harassment, victims cannot predict what the cumulative effect of the harassment will be after only experiencing the first of what could develop into a series of incidents. Furthermore, many sexual harassment victims do not possess the legal expertise to know what types of harassment are serious enough to warrant or require a formal complaint.

There is a general consensus among lower federal courts that only extremely severe harassment creates a cause of action based on one incident. Therefore, the majority of claims are based on the accumulation of incidents. Victims of harassment, however, are not able to judge what the cumulative effect of the harassment will be. After being subjected to one incident of harassment, a victim must decide how to react based on that single incident. Presumably, if a

---

242 See Llewellyn v. Celanese Corp., 693 F. Supp. 369, 377 (W.D.N.C. 1988) (stating that employers are only required to take remedial action “reasonably calculated to end the harassment”).

243 See Faragher, 524 U.S. at 807; Ellerth, 524 U.S. at 765; supra notes 160–191 and accompanying text.


245 Id.


248 See, e.g., Faragher, 524 U.S. at 788; Alfano, 294 F.3d at 374; supra notes 80–158 and accompanying text.

249 See, e.g., Morgan, 536 U.S. at 115; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; supra notes 80–158 and accompanying text.

250 See Morgan, 536 U.S. at 115; supra notes 80–212 and accompanying text.

251 See Faragher, 524 U.S. at 788; supra notes 80–212 and accompanying text.
woman knew that she would be harassed further or more severely in the future, she would have a much stronger reaction to the single incident.\textsuperscript{252} If, however, she knew that she would not be subjected to any future incidents, she may be more inclined to ignore the isolated incident.\textsuperscript{253}

Filing a formal complaint not only causes an inconvenience for the victim, but it may create more aggravation than the offensive incident itself.\textsuperscript{254} After complaining of sexual harassment, women have been ostracized, threatened, and even terminated from their employment.\textsuperscript{255} Aside from fear of possible retaliation, many other concerns may influence a victim’s decision not to report harassment.\textsuperscript{256} Victims of sexual harassment are often young women who are being harassed by older men or by direct supervisors who have authority over them.\textsuperscript{257} They may be new employees or may simply have a general hesitation to create tension or acrimony with co-workers and supervisors, or they may be ashamed or embarrassed about the event.\textsuperscript{258} Some victims might feel uneasy with confrontational situations, making it difficult to face the harasser or to complain about the conduct.\textsuperscript{259} These concerns, among others, may simply outweigh the seriousness of the initial conduct, leading a victim of harassment to conclude that she would rather not file a formal complaint.\textsuperscript{260} Faced with the possibility of retaliation or ostracism by co-workers for report-

\textsuperscript{252} See Morgan, 536 U.S. at 115; \textit{supra} notes 80–212 and accompanying text.
\textsuperscript{253} See Faragher, 524 U.S. at 788; \textit{supra} notes 80–212 and accompanying text.
\textsuperscript{254} See, e.g., Noviello v. City of Boston, 398 F.3d 76, 82, 98 (1st Cir. 2005) (finding that after reporting incident of sexual harassment, plaintiff was retaliated against and ostracized by her co-workers); Hanna v. Boys & Girls Home & Family Servs., Inc., 212 F. Supp. 2d 1049, 1069 (N.D. Iowa 2002) (finding a factual issue whether plaintiff was terminated in retaliation for complaining of sexual harassment); Pereira v. Schlage Elecs., 902 F. Supp. 1095, 1099, 1100–01 (N.D. Cal. 1995) (describing how after plaintiff complained of harassment, co-workers threatened to kill her, burn down her house, and kidnap her and leave her where she would be raped and killed, as well as finding a factual issue whether plaintiff was terminated in retaliation for complaining of sexual harassment).
\textsuperscript{255} See, e.g., Noviello, 398 F.3d at 82, 98; Hanna, 212 F. Supp. 2d at 1069; Pereira, 902 F. Supp. at 1099, 1100–01.
\textsuperscript{256} See Noviello, 398 F.3d at 82, 98; Hanna, 212 F. Supp. 2d at 1069; Pereira, 902 F. Supp. at 1099, 1100–01.
\textsuperscript{257} See \textit{supra} notes 80–212 and accompanying text.
\textsuperscript{258} See \textit{supra} notes 80–212 and accompanying text.
\textsuperscript{259} See \textit{supra} notes 80–212 and accompanying text.
\textsuperscript{260} See \textit{supra} notes 80–212 and accompanying text.
ing harassment, it is no surprise that many victims choose to ignore trivial or isolated incidents. As the standard is applied in many lower federal courts, however, the victim must report the first incident or she will be barred from asserting a cause of action in the future if the harassment continues or escalates. Such a requirement is difficult to meet without the benefit of being able to predict what will happen in the future. For example, in Conatzer v. Medical Professional Building Services Corp., the Tenth Circuit barred the plaintiff’s claim due to the employer’s affirmative defense. The court found it unreasonable that the plaintiff did not report the first of two incidents, described as her supervisor stepping up to her, leaning against her, and rubbing against the side of her chest. It hardly seems unreasonable that a woman would choose not to report one such ambiguous occurrence, especially when there was no indication at that time that any additional offenses would occur. The court concluded that the reasonable and expected response to such an incident would be to file a formal sexual harassment complaint with the employer. Under the circumstances, this may have been a fairly drastic response to what at the time seemed an innocuous and ambiguous act.

The second incident occurred two weeks later. When the plaintiff bent over to pick something up, her supervisor briefly placed her in a headlock with his thighs. First, this incident was much more offensive than the previous one. For a woman’s supervisor to place her head between his legs is much more objectionable than his merely rubbing against the side of her chest. Second, this incident was more significant because it was the second time in two weeks that

261 See Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033–34; supra notes 80–212 and accompanying text.

262 See Conatzer, 95 F. App’x at 278 (finding delay unreasonable where victim did not report an incident where supervisor leaned up against her and rubbed against her side); Phillips, 83 F. Supp. 2d at 1033–34; supra notes 80–117, 160–191 and accompanying text.

263 See, e.g., Faragher, 524 U.S. at 807–08; Ellerth, 524 U.S. at 765; Conatzer, 95 F. App’x at 281; Phillips, 83 F. Supp. 2d at 1034; supra notes 80–117, 160–191 and accompanying text.

264 95 F. App’x at 281.

265 Id. at 278, 281.

266 See Conatzer, 95 F. App’x at 281; supra notes 168–174 and accompanying text.

267 See Conatzer, 95 F. App’x at 281.

268 See id. at 278, 281.

269 Id. at 278.

270 Id.

271 See id.

272 See Conatzer, 95 F. App’x at 278.
her supervisor had physically touched her in an inappropriate manner. In light of the first incident, the victim was then able to evaluate whether this was developing into a pattern of harassment rather than a trivial and isolated incident. Although the first incident may have annoyed her, it was reasonably perceived as inconsequential enough to ignore. It should at least have been reasonable for the victim to decide that it did not mandate a formal complaint. In conjunction with the second incident, however, the initial conduct took on greater significance. The victim now could reasonably conclude that the first incident was more than an isolated occurrence of annoying behavior. In fact, the victim in Conatzer appeared to reach this conclusion because she reported both incidents promptly following the second incident.

Sexual harassment victims as potential plaintiffs must benefit from a view of the whole picture, as do the courts and employers. Considering the first incident described in Conatzer in isolation, it would have been a bold, and likely fruitless, step for the victim to file a complaint against her supervisor. Once she realized, however, that the supervisor’s conduct would persist, she filed a complaint within four or five days. When the court in Conatzer was presented with the facts of this case, it had the full record to evaluate. To the court, the plaintiff acted unreasonably because she failed to report the first incident after it happened. This was, of course, with the knowledge that the second incident would later occur.

This analysis forces an employee to file a formal complaint every time any ambiguous conduct occurs. Thus, it contradicts the Supreme Court’s admonition that Title VII is not meant to be a general

---

273 See id.
274 See id.
275 See id.
276 See id.
277 See Conatzer, 95 F. App’x at 278.
278 See id.
279 See id.
280 See Morgan, 536 U.S. at 115; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; supra notes 80–117, 160–191 and accompanying text.
281 See Conatzer, 95 F. App’x at 278.
282 See id.
283 See Morgan, 536 U.S. at 115; Conatzer, 95 F. App’x at 278; supra notes 168–174 and accompanying text.
284 See Conatzer, 95 F. App’x at 281.
285 See id.
286 See id. at 278, 281; supra notes 80–117, 160–191 and accompanying text.
civility code for the workplace. Although it seems reasonable not to file a sexual harassment complaint for insignificant occurrences, many courts evaluating the affirmative defense find that such incidents do constitute acts of sexual harassment, at least when followed by further incidents in the future.

The U.S. District Court for the Eastern District of Missouri, in Phillips v. Taco Bell Corp., engaged in a similar analysis, although in this case the initial incident was more serious than in Conatzer. In Phillips, the first incident occurred when the supervisor squeezed the plaintiff’s breasts. Three months later, and within a seven-day period, he subjected her to four more incidents of harassment. Two days after the fourth incident, the plaintiff reported the conduct to her employer. Although the initial incident was serious, the victim had no reason to suspect that the conduct would be repeated. She may have had a multitude of reasons for deciding not to file a formal complaint in response. For three months, it would have appeared that the first incident was nothing more than an isolated occurrence.

Then, three months later, the harassment resurfaced when the supervisor subjected the victim to four incidents of harassment in a seven-day period. She reported the conduct to her employer two days after the last incident. The court looked at the date of the initial incident and the date the plaintiff made the report and concluded that a delay of over three months was unreasonable. In making this calculation, the court presupposed that the initial incident and the four incidents occurring three months later were all part of the same pattern of harassment. What if the first incident had occurred one year earlier? Or five years?

---

289 See Conatzer, 95 F. App’x at 278; Phillips, 83 F. Supp. 2d at 1033; supra notes 160–174 and accompanying text.
290 83 F. Supp. 2d at 1033.
291 Id.
292 Id.
293 See id.
294 See, e.g., Noviello, 398 F.3d at 82, 98; Hanna, 212 F. Supp. 2d at 1069; Pereira, 902 F. Supp. at 1099, 1100–01; supra notes 160–167 and accompanying text.
295 See Phillips, 83 F. Supp. 2d at 1033.
296 Id.
297 Id.
298 See id. at 1034.
299 See id. at 1033–34.
C. Although Courts Are Unduly Strict When Dealing with Plaintiffs, They Show Leniency to Employers

Lower federal courts hold sexual harassment victims and employers to different standards of “reasonableness.” When analyzing the affirmative defense, courts hold plaintiffs to a strict standard of reasonable conduct while allowing employers to engage in quite unreasonable practices. In fact, many courts imply that the plaintiff has the burden to refute a presumption in favor of the employer when evaluating the employer’s affirmative defense.

In Wyatt v. Hunt Plywood Co., the court effectively found the plaintiff at fault for the employer’s unsatisfactory response. After the plaintiff’s immediate supervisor began harassing her, she promptly complained to the supervisor’s superior, but despite her complaints, the harassment persisted. Instead of remediating the problem, the superior himself eventually subjected the plaintiff to sexual advances and harassment. The court found that the plaintiff’s failure to report either harasser’s behavior to an additional supervisor was unreasonable. Even though the plaintiff reported the first supervisor’s conduct to an appropriate person within the company, the court found that she needed to report it again because the person to whom she complained was ineffective in remediating the problem. The court concluded that any harassment occurring after the victim realized the superior was ineffective in dealing with the complaint could not survive a motion for summary judgment because the plaintiff had

300 See supra notes 80–117, 160–212 and accompanying text.
301 See supra notes 160–212 and accompanying text.
302 See, e.g., Faragher, 524 U.S. at 807–08; Ellerth, 524 U.S. at 765 (stating that the existence of an anti-harassment policy with a complaint procedure is not necessary in every instance as a matter of law, and proof that an employee failed to use any complaint procedure provided by the employer will normally suffice to satisfy the employer’s burden to show that an employee failed to fulfill the obligation of reasonable care to avoid harm); Wyatt, 297 F.3d at 407–08, 413 (finding that once it became clear that the supervisor was ineffective in dealing with the harassment, the plaintiff’s failure to report the harassment to a different supervisor was unreasonable); supra notes 160–212 and accompanying text. The Wyatt court also found that the employer’s sexual harassment policy and its implementation were more than adequate, even though the plaintiff had to report the harassment multiple times, was sexually harassed by the supervisor to whom she complained, and the employer intentionally omitted the most serious incident to downplay the harassment when a formal report was finally completed. 297 F.3d at 407–08, 413.
303 See 297 F.3d at 413.
304 See id.
305 Id.
306 Id.
307 Id.
acted unreasonably. Instead of holding the employer liable for creating an ineffective complaint procedure, the court found that the victim had a responsibility to keep reporting the harassment until she could find someone who would take her complaint seriously. Although the victim promptly reported the harassment, only to be sexually harassed by the person to whom she had complained, the court found that she had acted unreasonably, not the employer.

Conversely, courts hold employers to a low standard of reasonableness, even when they act quite unreasonably. In *Saidu-Kamara v. Parkway Corp.*, the plaintiff experienced various forms of harassment throughout her employment. Although she repeatedly reported these incidents, her employer took no action in response. The court stated that the plaintiff had failed to demonstrate sufficiently severe or pervasive harassment, and granted the employer’s motion for summary judgment. In a case such as this, a victim could only force the employer to act if the harassment finally became severe or pervasive. Only then would the court potentially grant relief. Otherwise, she must either accept the harassment and the fact that her employer has no duty to end it, or quit her job. Allowing employers to ignore reports of harassment like this sets a troubling precedent. The courts are conveying the message that employers have no obligation to respond to complaints of harassment unless the harassment is of the magnitude found in the actionable cases set out above—that is, a high level of severity or pervasiveness, and, in some

308 Wyatt, 297 F.3d at 413.
309 See id.
310 See id.
311 See *Saidu-Kamara*, 155 F. Supp. 2d at 438; *supra* notes 193–212 and accompanying text.
312 155 F. Supp. 2d at 438.
313 Id.
314 Id. at 440.
315 See *Faragher*, 524 U.S. at 807; *Saidu-Kamara*, 155 F. Supp. 2d at 440; *supra* notes 80–212 and accompanying text.
316 See *Faragher*, 524 U.S. at 807; *Saidu-Kamara*, 155 F. Supp. 2d at 440; *supra* notes 80–212 and accompanying text.
317 See *Faragher*, 524 U.S. at 807; *Saidu-Kamara*, 155 F. Supp. 2d at 440; *supra* notes 80–212 and accompanying text.
318 See *Faragher*, 524 U.S. at 807; *Saidu-Kamara*, 155 F. Supp. 2d at 440; *supra* notes 193–212 and accompanying text.
cases, more abusive than physical groping of intimate body parts and sexual assaults.\textsuperscript{319}

Admittedly, it is not desirable to have courts require employers to respond to complaints of truly trivial harassment or, in the words of the Supreme Court, “the ordinary tribulations of the workplace, such as the sporadic use of abusive language, gender-related jokes, and occasional teasing.”\textsuperscript{320} It should be just as undesirable to require victims to report these incidents.\textsuperscript{321} If employers are only required to respond to actionable harassment, employees should be held to the same standard.\textsuperscript{322} An employee must report every instance of harassment to preserve a cause of action, while the employer need not respond.\textsuperscript{323} The legal obligations of employees and employers should be placed on equal footing.\textsuperscript{324} If an employer is only required to respond to an actionable environment, then an employee should only be required to report actionable harassment.\textsuperscript{325}

D. Leveling the Field: Placing Equal Obligations on Employees and Employers

Courts should require hostile work environment victims to report sexual harassment only once it becomes severe or pervasive.\textsuperscript{326} This

\textsuperscript{319} See Faragher, 524 U.S. at 807; McPherson v. City of Waukegan, 379 F.3d 430, 439 (7th Cir. 2004); Wilson v. Chrysler Corp., 172 F.3d 500, 511 (7th Cir. 1999); Saidu-Kamara, 155 F. Supp. 2d at 440; supra notes 80–212 and accompanying text.
\textsuperscript{320} See Faragher, 524 U.S. at 788 (quoting BARBARA LINDEMANN & DAVID KADUE, SEXUAL HARASSMENT IN EMPLOYMENT LAW 175 (1992)); supra notes 55–212 and accompanying text.
\textsuperscript{321} See Faragher, 524 U.S. at 788; supra notes 55–212 and accompanying text.
\textsuperscript{322} See Faragher, 524 U.S. at 807; Conatzer, 95 F. App’x at 281 (finding that the harassment was comprised of isolated incidents and, therefore, was non-actionable, but also finding that the plaintiff unreasonably failed to report the conduct); Slay v. Glickman, 137 F. Supp. 2d 743, 751–52 (S.D. Miss. 2001) (finding that plaintiff unreasonably failed to report the harassment, but also finding that a reasonable person would not conclude that the complained-of harassment would constitute a hostile work environment); supra notes 55–212 and accompanying text.
\textsuperscript{323} See Faragher, 524 U.S. at 807; Conatzer, 95 F. App’x at 281; Saidu-Kamara, 155 F. Supp. 2d at 438, 440; Slay, 137 F. Supp. 2d at 751–52; supra notes 55–212 and accompanying text.
\textsuperscript{324} See Faragher, 524 U.S. at 807; Conatzer, 95 F. App’x at 281; Saidu-Kamara, 155 F. Supp. 2d at 438, 440; Slay, 137 F. Supp. 2d at 751–52; supra notes 55–212 and accompanying text.
\textsuperscript{325} See Faragher, 524 U.S. at 807; Conatzer, 95 F. App’x at 281; Saidu-Kamara, 155 F. Supp. 2d at 438, 440; Slay, 137 F. Supp. 2d at 751–52; supra notes 55–212 and accompanying text.
\textsuperscript{326} See Faragher, 524 U.S. at 807–08 (stating that demonstration that the plaintiff unreasonably failed to use any complaint procedure provided by the employer will normally suffice to satisfy the employer’s burden, but that an employer will be subject to vicarious liability for an “actionable” hostile environment); Ellerth, 524 U.S. at 765 (stating that demonstration that the plaintiff unreasonably failed to use any complaint procedure provided by the employer will normally suffice to satisfy the employer’s burden); Madray v.
rule serves three purposes.\textsuperscript{327} First, the victim is not required to report harassment if it is not truly serious enough to upset her.\textsuperscript{328} She would only need to file a report when a severe incident occurs or minor incidents go on so long as to become sufficiently pervasive.\textsuperscript{329} Second, this standard would be truer to the requirements of the affirmative defense.\textsuperscript{330} The employer would still avoid liability if it took appropriate action to prevent and correct promptly any sexually harassing behavior of which it was aware.\textsuperscript{331} Under this proposed standard, the employer could better judge the necessity for a response because the reported behavior would be more serious.\textsuperscript{332} The employer would still only be required to act once it had notice of the harassment.\textsuperscript{333} Third, this standard would place the plaintiff’s prima facie case and the employer’s affirmative defense on equal footing.\textsuperscript{334} Under this approach, the plaintiff would have an actionable claim, and the employer would have an affirmative defense.\textsuperscript{335}

Publix Supermarkets, Inc., 208 F.3d 1290, 1300 (11th Cir. 2000) (noting that, once an employer has promulgated an effective anti-harassment policy, it is then incumbent upon the employees to utilize the procedural mechanisms established); \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{327} See \textit{infra} notes 328–335 and accompanying text.

\textsuperscript{328} See Faragher, 524 U.S. at 807–08; Ellerth, 524 U.S. at 765; Madray, 208 F.3d at 1300; \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{329} See Faragher, 524 U.S. at 807–08; Ellerth, 524 U.S. at 765; Alfano, 294 F.3d at 374 (stating that a plaintiff must demonstrate either that a single incident was extraordinarily severe, or that a series of incidents was sufficiently continuous and concerted); Madray, 208 F.3d at 1300; Indest, 164 F.3d at 264 (stating that actionable cases have involved patterns or allegations of extensive, long-lasting, unredressed, and uninhibited sexual threats or conduct that permeated the plaintiffs’ work environment); \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{330} See Faragher, 524 U.S. at 807 (concluding that to assert its affirmative defense an employer must prove that (1) it exercised reasonable care to prevent and promptly correct any sexually harassing behavior, and (2) the plaintiff employee unreasonably failed to take advantage of any preventive or corrective opportunities provided by the employer or to otherwise avoid harm); Ellerth, 524 U.S. at 765; \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{331} See Faragher, 524 U.S. at 807; Ellerth, 524 U.S. at 765; \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{332} See Faragher, 524 U.S. at 807–08 (stating that an employer will be subject to vicarious liability for an “actionable” hostile work environment); Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{333} See Faragher, 524 U.S. at 807–08; Ellerth, 524 U.S. at 765; Madray, 208 F.3d at 1300; \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{334} See Faragher, 524 U.S. at 788; Conatzer, 95 F. App’x at 278; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; Phillips, 83 F. Supp. 2d at 1033–34; \textit{supra} notes 55–212 and accompanying text.

\textsuperscript{335} See Faragher, 524 U.S. at 788; Conatzer, 95 F. App’x at 278; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; Phillips, 83 F. Supp. 2d at 1033–34; \textit{supra} notes 55–212 and accompanying text.
This standard is much more consistent than the current system. Presently, the courts are faced with plaintiffs who cannot present a prima facie case and also who cannot defeat the employer’s affirmative defense. An employer is not legally required to do anything until the harassment becomes actionable. If an employer is only required to respond to actionable harassment, then an employee should only be required to report actionable harassment.

This proposed standard would also make the courts’ analysis of a plaintiff’s prima facie case and the employer’s affirmative defense more consistent with each other. Under the present standards, many courts conclude that a plaintiff suffered only isolated or trivial harassment—that is, non-actionable—while also concluding that the plaintiff was unreasonable for failing to report it. When evaluating the plaintiff’s prima facie case, such courts find that the harassment was insignificant, some even opining that the victim was overly sensitive for perceiving such conduct as offensive. When evaluating the affirmative defense, however, the same incident of harassment could be the decisive factor in absolving the employer of liability because the court may conclude that any reasonable person would have promptly reported the incident. Under the proposed standard, the courts’ reasoning would be more consistent and equitable. If the court finds that the harassment was sufficiently severe or pervasive to be actionable, then it could logically conclude that the plaintiff was

336 See Faragher, 524 U.S. at 788; Conatzer, 95 F. App’x at 278; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; Phillips, 83 F. Supp. 2d at 1033–34; supra notes 55–212 and accompanying text.
337 See, e.g., Faragher, 524 U.S. at 788; Conatzer, 95 F. App’x at 278; Alfano, 294 F.3d at 374; Indest, 164 F.3d at 264; Phillips, 83 F. Supp. 2d at 1033–34; supra notes 80–117, 160–191 and accompanying text.
338 See Faragher, 524 U.S. at 788; supra notes 193–212 and accompanying text.
339 See Conatzer, 95 F. App’x at 278 (finding that the plaintiff unreasonably failed to report harassment even though it was no more than isolated incidents, and therefore, non-actionable); DeAngelis, 51 F.3d at 593; Baskerville, 50 F.3d at 430–31; Slay, 137 F. Supp. 2d at 751–52 (finding that the plaintiff unreasonably failed to report the harassment while also finding that a reasonable person would not consider the complained-of harassment to constitute a hostile work environment); Phillips, 83 F. Supp. 2d at 1033–34; Hosey, 1996 WL 414057, at *1–2; supra notes 80–117, 160–191 and accompanying text.
341 See DeAngelis, 51 F.3d at 593; Baskerville, 50 F.3d at 430–31; supra notes 80–117 and accompanying text.
343 See supra notes 160–191 and accompanying text.
344 See supra notes 80–117, 160–191 and accompanying text.
unreasonable in failing to report it. If the harassment was not severe or pervasive, however, then it is irrational to require the victim to report what is non-actionable. Under the proposed standard, plaintiffs and employers are treated equally.

Although victims of sexual harassment should be encouraged to report any incidents that they deem serious or that could lead to more severe incidents, they should only have a legal obligation to report actionable, and thus “severe or pervasive,” harassment. It is inappropriate to require victims to report every trivial incident of annoying behavior simply to ensure they can retain a cause of action in the future. This inconsistent legal standard is unduly burdensome for victims of sexual harassment. It is not good for workplace morale or workplace relations. And it creates additional administrative costs for employers because they must investigate complaints that employees may not otherwise have chosen to file unless required to do so. Therefore, lower federal courts should abandon the current standards for hostile work environment sexual harassment cases in favor of an approach that is more consistent, logical, and evenhanded.

Conclusion

Hostile work environment sexual harassment has become a recurrent problem in this country. The standards developed by the U.S. Supreme Court for handling this issue have been ineffective because, for the most part, these standards unduly favor employers over victims. The leniency that lower federal courts show towards employers has led them to engage in liability-avoidance techniques rather than harassment-prevention practices. Conversely, the strict standards to which plaintiff-employees are held make it extremely difficult to establish a prima facie case. Even if a plaintiff establishes an actionable claim, she may still find it challenging to prevail because many courts effectively require a plaintiff to refute a presumption in favor of the employer on the employer’s own affirmative defense.

348 See supra notes 55–212 and accompanying text.
349 See supra notes 55–212 and accompanying text.
350 See supra notes 55–212 and accompanying text.
351 See supra notes 55–212 and accompanying text.
352 See supra notes 55–212 and accompanying text.
353 See supra notes 55–212 and accompanying text.
First, lower federal courts require an unreasonably high level of severity or pervasiveness to establish an actionable claim. Federal courts have essentially required that the conduct either be extremely severe or be extensive, continuous, and long-lasting. Second, courts give plaintiffs a narrow window in which to report harassment. Without regard to the many reasons victims of sexual harassment may wish not to report isolated or trivial conduct, courts have found it unreasonable for victims to delay even a few weeks before filing a formal report with the employer. Third, even in cases where harassment is promptly reported, courts show leniency to employers in evaluating the reasonableness of their response. If the complained-of harassment was not severe or pervasive, the courts do not require an employer to make any response at all, and in cases where the harassment was actionable, employers are not required to end the harassment. These practices have tipped the scales heavily in favor of employers.

Given the disparity in resources available to most employers in comparison to employees, it is unfair to give them an additional advantage in court. It is unreasonable to expect a victim of sexual harassment to know which types of conduct are actionable and which are not. Likewise, it is unlikely that a victim is aware that if she waits more than a couple of weeks to report an incident, she will lose her cause of action, if she ever had one. Because victims are usually unaware of the legal standards, courts should show them leniency, rather than showing it to employers that commonly deal with the legal issues surrounding sexual harassment.

If victims of sexual harassment were only required to report actionable conduct, it would be more reasonable to hold them accountable for failing to bring it to the employer’s attention. It is understandable that victims are more likely to report serious incidents than trivial ones. Furthermore, only requiring victims to report actionable harassment places their legal claims on the same footing as the employer’s affirmative defense. The plaintiff will win or lose on her prima facie case based on whether the harassment was sufficiently severe or pervasive. The employer will win or lose on its affirmative defense based on whether the victim reported the harassment and the reasonableness of the employer’s response. This is more consistent than the current standard, under which courts may find that the plaintiff failed to establish a prima facie case and also failed to get past the employer’s affirmative defense by failing to report what was not actionable.

A standard that provides that sexual harassment must either be extremely severe or continue long enough to be pervasive, but then
also provides that a victim must promptly report the conduct following the first incident—whether actionable or not—is contradictory. Under the proposed standard, by contrast, the victim must report the conduct only if (1) the conduct is extremely severe and therefore actionable, or (2) the conduct continues long enough to be pervasive and is therefore actionable. This standard is both more consistent with the principles behind the hostile work environment cause of action and more equitable.

Evan D. H. White