## TABLE OF CONTENTS

### ARTICLES

<table>
<thead>
<tr>
<th>Article</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Roots of Removal</td>
<td>Debra Lyn Bassett</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Rex R. Perschbacher</td>
<td></td>
</tr>
<tr>
<td>Questioning Law Enforcement:</td>
<td>Shirin Sinnar</td>
<td>41</td>
</tr>
<tr>
<td>The First Amendment and Counterterrorism Interviews</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperative NRDA &amp; New Governance:</td>
<td>Michael B. Runnels</td>
<td>107</td>
</tr>
<tr>
<td>Getting to Restoration in the Hudson River, the Gulf of Mexico, and</td>
<td>Andrea Giampetro-Meyer</td>
<td></td>
</tr>
<tr>
<td>Beyond</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### THE ABRAHAM L. POMERANTZ LECTURE

<table>
<thead>
<tr>
<th>Lecture</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don’t Blink: Snap Decisions and Securities Regulation</td>
<td>Frank Partnoy</td>
<td>151</td>
</tr>
</tbody>
</table>

### TRIBUTE

<table>
<thead>
<tr>
<th>Tribute</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Trager: Jurist</td>
<td>Jeffrey Brandon Morris</td>
<td>181</td>
</tr>
</tbody>
</table>

### NOTES

<table>
<thead>
<tr>
<th>Note</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolving the Conflict Between the Stolen Valor Act of 2005 and the First Amendment</td>
<td>Stephanie L. Gal</td>
<td>223</td>
</tr>
<tr>
<td>Cutting the Baby in Half: An Economic Critique of Indivisible Resource Partition</td>
<td>Zachary D. Kuperman</td>
<td>263</td>
</tr>
</tbody>
</table>
Corporate Corruption & the New Gold Mine: How the Dodd-Frank Act Overincentivizes Whistleblowing  
Jenny Lee  303

An Unconstitutional Response to *Citizens United*: Mandating Shareholder Approval of Corporate Political Expenditures  
Robert B. Sobelman  341

Protecting a Right to Access Internet Content: The Feasibility of Judicial Enforcement in a Non-neutral Network  
Philip F. Weiss  383
ARTICLES

The Roots of Removal*

Debra Lyn Bassett†
Rex R. Perschbacher‡

Academic observers of federal litigation generally describe the field by reference to its constitutional and statutory foundations. Also at play within this landscape are powerful normative policy elements recognized by scholars and practitioners—at least implicitly—as essential to an adequate description of the litigation choices available to the participants. One of these policy features is the oft-repeated maxim that the plaintiff is the master of the claim. Although this basic premise quietly dominates both academic discussions and practice realities, a number of factors operate to impose very real limitations on that principle. These limitations, in turn, shape how we approach federal litigation and include, specifically, how we approach federal court jurisdiction. One of these limitations on a plaintiff’s power that implicates federal jurisdiction is removal—and removal provides an instructive example of the exceptionally rich environment where policy elements interact with constitutional and statutory features. Removal is a means of moving a state court lawsuit into federal court, and approximately thirty thousand civil cases are transferred in this manner annually. Through removal, under certain circumstances, a defendant is able to defeat the plaintiff’s choice of forum. Thus, removal inherently raises questions about what limitations should be placed on a plaintiff’s choice of forum. These questions have been answered in different ways depending on the specific issue and the timing, which includes both the historical context and the litigation point in time. Three particular aspects of removal law illustrate the dramatic way in which these differences unfold. As a general matter, when removal occurs at the very outset of a lawsuit, the procedures are quite straightforward. However, removal instituted after the initial

* © 2011 Debra Lyn Bassett & Rex R. Perschbacher. All rights reserved.
† Justice Marshall F. McComb Professor of Law, Southwestern Law School.
‡ Professor of Law and Daniel J. Dykstra Chair in Law, University of California, Davis.
thirty-day removal period raises more issues and has the potential to become more complicated. Three of the most difficult issues in removal law arise in such a subsequently instituted removal context—the first-(versus last-) served defendant rule, the voluntary-involuntary rule, and the one-year limitation for removing diversity cases. Courts and commentators typically have discussed these issues separately, without realizing that these issues share an underlying commonality that yields a surprisingly effective analytical framework: the inherent tension between deferring to the plaintiff's choice of forum versus the defendant's right of removal.

INTRODUCTION

The law of removal is a study in contradictions. The United States Constitution expressly authorizes arising-under and diversity jurisdiction, without mentioning removal. Yet removal is regularly classified as one of the bases for federal court subject-matter jurisdiction. Because removal is a procedure rather than a true form of jurisdiction, the Constitution makes no direct mention of removing cases from state court to federal court. However, the U.S. legal landscape has included removal since the creation of federal courts; the First Congress enacted removal procedures in the first Judiciary Act of 1789. This gives removal a unique place in federal court jurisprudence—a statutory regime of quasi-constitutional character.

At the same time, removal runs directly contrary to one of the most deeply embedded, yet implicit, maxims of United States adversarial procedure: the plaintiff is the master of his or her claim. Despite the lack of scholarly commentary, those who are involved in the practice of law in this country, and those of us teaching it, each accept the baseline norm that among the choices available by law, plaintiffs have the initial choice of judicial system (federal or state depending upon the limits of subject-matter jurisdiction), the parties who will join as plaintiffs, the parties to be named as defendants (assuming personal jurisdiction is available for court process to reach them), and the place of trial (venue). This plaintiff-choice system has been acknowledged by no less than the U.S. Supreme Court. See Piper Aircraft Co. v. Reyno, 454 U.S. 235, 242 (1982); Hoffman v. Blaski, 363 U.S. 335, 344 (1960). Moreover, plaintiffs are able to take advantage of any jurisdiction in which the action can be brought and where the statute of limitations against the plaintiffs' claim has not run, even if only one such state remains. See Keeton v. Hustler Magazine, Inc., 465 U.S. 770, 779-80 (1984).
these two concepts—that of the defendant’s right to remove versus the plaintiff as master of the claim—continues to be seen in court decisions characterizing the defendant’s ability to remove as inferior to the plaintiff’s choice of forum.\(^5\)

Despite the continuing debate over the historical purpose of diversity jurisdiction\(^6\) and the paucity of historical documentation,\(^7\) the theories as to diversity’s purpose originate in the concept of local bias or prejudice.\(^8\) Diversity offers a rich context in the conflict between removal, a defendant’s tool, and the plaintiff’s traditional role as master of the claim.

Strict application of the rule that gives plaintiffs absolute mastery of the litigation would allow plaintiffs, but not defendants, the right to choose to invoke federal diversity jurisdiction in qualifying cases and avoid the dangers of local bias, or nevertheless to select the state forum and its attendant risks of local bias. Defendants would simply have to live with the plaintiff’s forum (and other) choices. Removal provides a significant counterbalance. Rather than according the plaintiff exclusive control over the choice between a federal forum versus a state forum, as would be consistent with the “plaintiff as master of the claim” maxim, the Judiciary Act of 1789 provided for removal, which expressly permits a defendant to defeat the plaintiff’s forum choice. The original draft bill authorized

\(^5\) See, e.g., Marathon Oil Co. v. Ruhrgas, 145 F.3d 211, 219 n.11 (5th Cir. 1998) (“The defendant’s right to remove and the plaintiff’s right to choose the forum are not equal . . . .” (quoting 16 JAMES W. MOORE ET AL., MOORE’S FEDERAL PRACTICE § 107.05 (3d ed. 1997)), rev’d, 526 U.S. 574 (1999); Auchinleck v. Town of LaGrange, 167 F. Supp. 2d 1066, 1069 (E.D. Wis. 2001) (“The plaintiff’s right to choose his forum is superior to the defendant’s right of removal.”).

\(^6\) Compare Erie R.R. v. Tompkins, 304 U.S. 64, 74 (1938) (“Diversity of citizenship jurisdiction was conferred in order to prevent apprehended discrimination in state courts against those not citizens of the state.”), with Henry J. Friendly, The Historic Basis of Diversity Jurisdiction, 41 HARV. L. REV. 483, 496-97 (1928) (“The desire to protect creditors against legislation favorable to debtors was a principal reason for the grant of diversity jurisdiction.”); see also 13E CHARLES ALAN WRIGHT ET AL., FEDERAL PRACTICE AND PROCEDURE § 3601, at 13, 15 (3d ed. 2009) (noting “the traditional explanation, and the one most often cited by federal judges and legal scholars, of the purpose of the constitutional provision for diversity of citizenship jurisdiction and its immediate congressional implementation—the fear that state courts would be prejudiced against out-of-state litigants”). But see id. § 3601, at 15-16 (“Several historians have suggested that the real fear was not of the state courts, but of the state legislatures . . . . The fear of state legislatures may have arisen less from interstate hostility than from a desire to protect commercial interests from class bias.”).

\(^7\) See Friendly, supra note 6, at 484-85 (noting that diversity jurisdiction “had not bulked large” in the eyes of the Constitution’s framers, “[n]or are the records of the Convention fruitful to a student of the diversity clause”). See generally Debra Lyn Bassett, The Hidden Bias in Diversity Jurisdiction, 81 WASH. U. L.Q. 119, 122-36 (2003) (providing historical background of diversity jurisdiction).

\(^8\) Bassett, supra note 7, at 119-32.
removal of any lawsuit for which diversity jurisdiction existed; the final version authorized removal only when the plaintiff filed suit in her home state against an out-of-state defendant.10

The Supreme Court weighed in early on removal and supported removal in no uncertain terms, expressly rejecting any contention that only the plaintiff's forum choice was entitled to protection:

The [C]onstitution of the United States was designed for the common and equal benefit of all the people of the United States. The judicial power was granted for the same benign and salutary purposes. It was not to be exercised exclusively for the benefit of parties who might be plaintiffs, and would elect the national forum, but also for the protection of defendants who might be entitled to try their rights, or assert their privileges, before the same forum. . . . [A]s the plaintiff may always elect the state court, the defendant may be deprived of all the security which the constitution intended in aid of his rights. Such a state of things can, in no respect, be considered as giving equal rights. To obviate this difficulty, we are referred to the power which it is admitted congress possess to remove suits from state courts to the national courts . . . .11

The view that a defendant's right of removal has equal stature and the same constitutional dimension as a plaintiff's right to select the forum12 runs contrary to the plaintiff as master of the claim maxim, because removal's very purpose lies in defeating the plaintiff's choice of forum. However, this should not be seen as surprising, given the number of limitations on the plaintiff as master of the claim principle, both within and without the removal context.13

---

10 Id. at 91; Judiciary Act of 1789, ch. 20, § 12, 1 Stat. 73, 79-80 (authorizing the removal of any action “commenced in any state court against an alien, or by a citizen of the state in which the suit is brought against a citizen of another state, and [where] the matter in dispute exceeds the aforesaid sum or value of five hundred dollars, exclusive of costs”).
13 See, e.g., 28 U.S.C. § 1404(a)-(b) (2006) (authorizing motions for change of venue, through which a party may transfer an action to a different federal judicial district from that where the plaintiff originally filed the suit); Grable & Sons Metal Prods., Inc. v. Darue Eng’g & Mfg., 545 U.S. 308, 314 (2005) (despite plaintiff's election to sue in state court on a state law–based claim, defendant permitted to remove on the basis of arising under jurisdiction because the case raised a contested and substantial federal question
In yet another element of removal’s contradictory nature, the popular maxim that the courts must construe the removal statutes strictly does not alter this interpretation. A strict construction of the removal statutes does not suggest a bias against removal, nor does it restore the plaintiff as master of the claim to a superior position. Supreme Court case law indicates that the strict construction approach has nothing to do with subjugating defendants or elevating plaintiffs but instead has everything to do with basic concepts of federalism, and thus actions should be removed from state court jurisdiction only to the extent authorized by federal statute.

When removal comes within the statutory authority of 28 U.S.C. § 1441, the general removal statute, by definition the suit is one in which the federal and state courts have concurrent jurisdiction, so the federal and state courts have overlapping authority. This concurrent power means that removal does not offend the dignity of state judiciaries unless, again, the plaintiff as master of the claim is superior to a defendant’s right to remove—and thus removal serves to circumvent the state court’s superior claim to adjudicate the case. This position cannot prevail in a legal regime that allows removal in order to defeat the plaintiff’s choice of forum under

that the federal court could hear “without disturbing any congressionally approved balance of federal and state judicial responsibilities”); Beneficial Nat’l Bank v. Anderson, 539 U.S. 1 (2003) (despite plaintiff’s election to sue in state court on a state law based claim, federal preemption of state law resulted in recharacterization of plaintiff’s claim); Federated Dep’t Stores, Inc. v. Moiré, 452 U.S. 394, 397 n.2 (1981) (stating that although the plaintiff had pleaded his claim solely in terms of state law, “at least some of the claims had a sufficient federal character to support removal”).

14 Shamrock Oil & Gas Corp. v. Sheets, 313 U.S. 100, 108 (1941) (stating, in the removal context, that “[t]he power reserved to the states under the Constitution to provide for the determination of controversies in their courts, may be restricted only by the action of Congress in conformity to the Judicial Articles of the Constitution. ‘Due regard for the rightful independence of state governments, which should actuate federal courts, requires that they scrupulously confine their own jurisdiction to the precise limits which the statute has defined’” (citations omitted)). Shamrock Oil has been cited for the proposition that a plaintiff’s right to select the forum is superior to the defendant’s right to removal. See, e.g., Auchinleck v. Town of LaGrange, 167 F. Supp. 2d 1066, 1069 (E.D. Wis. 2001) (citing Shamrock Oil as authority in stating that “[t]he plaintiff’s right to choose his forum is superior to the defendant’s right of removal”); see also In re World Trade Ctr. Disaster Site Litig., 270 F. Supp. 2d 357, 366 (S.D.N.Y. 2003) (citing Shamrock Oil for “the right of plaintiffs to choose the forum in which to bring suit”). But Shamrock Oil said no such thing. Indeed, Shamrock Oil merely held that a plaintiff cannot remove a state court lawsuit to federal court on the basis of a counterclaim. See Shamrock Oil, 313 U.S. at 106-07.

15 See ALISON L. LACROIX, THE IDEOLOGICAL ORIGINS OF AMERICAN FEDERALISM 189 (2010) (noting that concurrent powers between the federal government and the states refers “to a structure in which multiple levels of government within a single polity possess[] overlapping authority to regulate, legislate, or adjudicate”).
specified circumstances. As expressed by one commentator, “Stating that a plaintiff has a superior ‘right’ to select a forum is merely an unsupported claim, not a self-evident fact.”

The concurrent jurisdiction of the federal and state courts where removal is properly invoked, removal’s history, and the statutory authorization of removal all serve to reduce the plaintiff’s power. “Master of the claim” becomes the minimalist axiom that the plaintiff chooses the initial court in which to file the claim subject, whenever federal jurisdiction is available, to the defendant’s right to rely on the removal statutes’ authority. The defendant’s right to remove does, and should, rightfully defeat the plaintiff’s selected forum.

Removal is a popular procedure, transferring approximately thirty thousand cases annually out of state courts and into federal courts. When a plaintiff files a civil lawsuit in state court, federal statutes authorize the defendant to remove the suit from the state court to federal court under certain specified circumstances and pursuant to specified procedures. The basic removal provisions, especially for lawsuits that involve a single defendant, are relatively straightforward and unremarkable. Only a defendant can remove, and a defendant can only effect removal from a state court to a federal court. The federal court to which the lawsuit is removed must be the federal court for the district and division encompassing the state court. As is true for any lawsuit that seeks to proceed in federal court, the action must

---

40 See 28 U.S.C. § 1441(a) (specifically referring to defendants); Chi., R.I. & P.R. Co. v. Stude, 346 U.S. 574, 580 (1954) (“The plaintiff under 28 U.S.C. § 1441(a) . . . cannot remove.”); see also Or. Egg Producers v. Andrew, 458 F.2d 382, 383 (9th Cir. 1972) (“A plaintiff who commences his action in a state court cannot effectuate removal to a federal court even if he could have originated the action in a federal court and even if a counterclaim is thereafter filed that states a claim cognizable in federal court.”). See generally 14C Charles Alan Wright et al., Federal Practice and Procedure § 3730, at 429 (4th ed. 2009).
42 Id.
have a basis for federal subject-matter jurisdiction. A defendant has thirty days from the date of service to effect removal by filing a notice of removal in the appropriate federal court and serving copies on the other parties and the clerk of the state court. If diversity provides the basis for federal subject-matter jurisdiction, then the defendant may not remove the lawsuit if he is a citizen of the state where the plaintiff filed the suit.

Removal becomes more complicated, and its contradictory nature becomes more apparent in a manner that has bedeviled the courts and commentators, when it does not occur within the initial thirty-day period. Indeed, three different removal issues have the potential to come into play in a subsequently instituted removal situation: the first- (versus last-) served defendant rule; the voluntary-involuntary rule; and the one-year limit for removing suits on the basis of diversity. Encompassing a contradictory variety of underlying principles and policies, these issues—one expressly created by statute and the other two judicially created—have resulted in a disjointed and inconsistent approach to removal that has obscured the underlying tension intrinsic to the removal concept: a tension between according deference to the plaintiff’s choice of forum and honoring the defendant’s right to remove. An examination of these three subsequently instituted removal issues illustrates the dramatic way that this underlying tension unfolds.

This article analyzes these three issues as they arise in the subsequently instituted removal context and identifies an overarching and unifying framework. Part I presents and analyzes the first-served, last-served, and intermediate rules. Part II presents and analyzes the voluntary-involuntary rule. Part III presents and analyzes the one-year limit on diversity-based removal. Finally, Part IV identifies the common factors that motivate these removal issues, analyzes the reach of deference to the plaintiff’s forum choice and of the defendant’s right to remove, and proposes an overarching framework for analyzing these three removal issues.

24 Id. § 1446(b), (d).
25 Id. § 1441(b). But see id. § 1453 (eliminating this restriction for certain class actions pursuant to the Class Action Fairness Act of 2005).
26 See infra notes 30-98 and accompanying text.
27 See infra notes 99-137 and accompanying text.
28 See infra notes 138-54 and accompanying text.
29 See infra notes 155-69 and accompanying text.
I. THE FIRST-SERVED, LAST-SERVED, AND INTERMEDIATE RULES

Title 28 of the United States Code, at section 1446, requires a defendant to file a notice of removal within thirty days of service, and case law interpreting section 1441 requires that all defendants joined and served in the action must consent to removal. When a complaint names multiple defendants, the potential exists that—whether due to the plaintiff’s intentional staggering of service or unanticipated service difficulties—all defendants will not be served on the same day. The removal statutes are silent as to how to reconcile the thirty-day time limit with differences in the timing of service of multiple defendants. Accordingly, the federal courts created the first-served, last-served, and intermediate rules as interpretations of how courts should implement the statutory removal timing restrictions. The significance of these rules stems from their connection to still another judicially created rule—the rule of unanimity. The rule of unanimity requires all defendants joined and served to consent to removal subject to some limited exceptions. Because all defendants must join in the removal petition and the removal statute imposes a thirty-day removal window, courts have struggled to determine how to reconcile these provisions in multiple-defendant situations when the defendants were served on different dates.

32. See Lindsay E. Hale, Triggering Removal Under 28 U.S.C. § 1446: The Eleventh Circuit’s Adoption of the Last-Served Defendant Rule in Bailey v. Janssen Pharmaceutica, Inc., 32 Am. J. Trial Advoc. 363, 364 (2008) (noting that as a result of the lack of clarification within 28 U.S.C. § 1446(b) regarding how to calculate the thirty-day removal period when there are multiple defendants who were served at different times, the federal courts have created their own interpretations, including the first-served and last-served defendant rules); see also infra notes 36-38.
33. See Haiber, supra note 17, at 648-49 (noting that the rule of unanimity “is not found in the text for the removal statutes,” but has “long [been] required” by the federal courts).
34. See, e.g., Beardsley v. Torrey, 2 F. Cas. 1188, 1189 (C.C.D. Pa. 1822) (No. 1190) (“[I]t is not competent to one defendant to remove the cause without the consent of his co-defendants.”).
35. See Mathews, 826 F. Supp. at 1318-19 (listing exceptions to the rule of unanimity, including nominal parties and “where federal jurisdiction of a party is based on a separate and independent jurisdictional grant”). In addition, certain class actions need not satisfy the rule of unanimity pursuant to the Class Action Fairness Act of 2005, 28 U.S.C. § 1453.
The circuit courts are divided as to the practical ramifications of such staggered service. The Fifth Circuit Court of Appeals has adopted the first-served defendant rule, whereby the time for removal expires thirty days after the first defendant is served without regard to when other defendants are served.\footnote{See Brown v. Demco, Inc., 792 F.2d 478, 481-82 (5th Cir. 1986). See generally 14C Wright et al., supra note 20, § 3731, at 586.} Courts of Appeals for the Sixth, Eighth, Ninth, and Eleventh Circuits have adopted the alternative and contradictory last-served defendant rule, whereby the time for removal is calculated from the date of service upon the defendant who attempts removal, regardless of when the first defendant was served.\footnote{See Destfino v. Reiswig, 630 F.3d 952, 955-56 (9th Cir. 2011); Bailey v. Janssen Pharmaceuticals, Inc., 536 F.3d 1202 (11th Cir. 2008); Marano Enters. of Kan. v. Z-Teca Rests., L.P., 254 F.3d 753 (8th Cir. 2001); Brierly v. Alusuisse Flexible Packaging, Inc., 184 F.3d 527 (6th Cir. 1999). See generally 14C Wright et al., supra note 20, § 3731, at 597-99.} The Fourth Circuit Court of Appeals has adopted a third view, the so-called intermediate rule, which requires the filing of the notice of removal within the first-served defendant's thirty-day window but permits later-joined defendants to join that original removal notice within thirty days of the date of their own, subsequent service.\footnote{Barbour v. Int'l Union, 640 F.3d 599, 607 (4th Cir. 2011).} The circuit courts, in choosing among these three alternatives, have grounded their approaches upon those principles and policies to which the circuit gives priority.

A. The Articulated Principles and Policies Motivating the First-Served Defendant Rule

The first-served defendant rule attempts to reconcile three general, undisputed removal principles: (1) the statutory provision for a thirty-day window within which a defendant must effect removal; (2) the axiom that courts should interpret the removal statutes strictly; and (3) the rule of unanimity. Under the first-served defendant rule, the time for removal expires thirty days after the first defendant is served. This interpretation is in accord with both the thirty-day window and a strict statutory construction. The first-served defendant rule also comports with the rule of unanimity, because if the first-served defendant prefers to remain in state court and thus does not remove the action, then this indicates a lack of unanimity among all the defendants, whenever served, over whether to remove the
lawsuit. The Fifth Circuit, in adopting the first-served defendant rule, expressly relied on these three principles.\textsuperscript{39}

In Brown v. Demco, Inc.,\textsuperscript{40} the plaintiff sued various defendants in Louisiana state court after suffering an employment-related injury. The plaintiff was a Louisiana domiciliary, no defendant was a citizen of Louisiana, and the plaintiff sought more than $2 million in damages, so federal diversity jurisdiction was clearly available. Although the plaintiff served all the defendants promptly, no defendant removed the action to federal court. Four years later, the plaintiff amended his complaint to add another defendant. Within thirty days of service, the newly added defendant filed a notice of removal in which the other original defendants all concurred. The Fifth Circuit concluded that removal was improper because it fell outside the thirty days allotted to the first-served defendant to remove.

[The first-served defendant rule] follows logically from the unanimity requirement, the thirty-day time limit, and the fact that a defendant may waive removal by proceeding in state court. Moreover, by restricting removal to instances in which the statute clearly permits it, the rule is consistent with the trend to limit removal jurisdiction and with the axiom that the removal statutes are to be strictly construed against removal.\textsuperscript{41}

With respect to the effect of the first-served rule upon the ability of later-served defendants to remove, the court stated that it did not perceive any unfairness to later-served defendants\textsuperscript{42} and observed that an alternative approach would introduce delay and uncertainty.\textsuperscript{43}

Although Brown involved distinctive facts due to the four-year interim before the addition of the final defendant, the Fifth Circuit reached the same conclusion in a subsequent case involving a much smaller disparity in the timing of service. In Getty Oil Corp. v. Insurance Co. of North America,\textsuperscript{44} the plaintiff served one defendant on September 3, another defendant on September 5, and the third defendant on September 24.\textsuperscript{45} The

\textsuperscript{39} Brown, 792 F.2d at 478, 481-82.
\textsuperscript{40} Id.
\textsuperscript{41} Id. at 482 (citation omitted).
\textsuperscript{42} However, of course, under the first-served defendant rule, a belatedly served defendant has no opportunity to remove in such an instance despite the statutory removal authority.
\textsuperscript{43} Brown, 792 F.2d at 482.
\textsuperscript{44} 841 F.2d 1254 (5th Cir. 1988).
\textsuperscript{45} Id. at 1256.
first defendant filed a notice of removal on September 26, within the thirty-day window and after the plaintiff had served all defendants. However, the last-served defendant did not consent to removal until October 24—within thirty days of its own service, but not within thirty days of service upon the first-served defendant. The Fifth Circuit found removal improper:

It follows that since all served defendants must join in the petition, and since the petition must be submitted within thirty days of service on the first defendant, all served defendants must join in the petition no later than thirty days from the day on which the first defendant was served. This rule . . . promotes unanimity among the defendants without placing undue hardships on subsequently served defendants. Indeed, if a removal petition is filed by a served defendant and another defendant is served after the case is thus removed, the latter defendant may still either accept the removal or exercise its right to choose the state forum by making a motion to remand.

The last-served defendant rule offers an alternative approach, although motivated by exactly the same principles and policies.

B. The Articulated Principles and Policies Motivating the Last-Served Defendant Rule

The last-served defendant rule attempts to reconcile the same three removal principles as the first-served rule, albeit with a different emphasis and conclusion: (1) the statute provides a thirty-day window within which a defendant must effect removal; (2) the axiom that courts should interpret the removal statutes strictly; and (3) the rule of unanimity. More recent cases have also asserted that the U.S. Supreme Court’s decision in Murphy Brothers, Inc. v. Michetti Pipe Stringing, Inc., supports the last-served defendant rule.

In Brierly v. Alusuisse Flexible Packaging, Inc., the plaintiff sued two defendants in state court. The first-served defendant removed the suit to federal court, but the federal court remanded the case when the defendant could not prove the citizenship of the codefendant and thus could not
demonstrate the existence of complete diversity.\footnote{Id. at 530.} After remand, the plaintiff served the second defendant, who filed a notice of removal within thirty days of that service; the first defendant filed a notice of consent to the removal on the same day.\footnote{Id. at 530-31.} The Sixth Circuit found removal proper.

In reaching its conclusion, the Sixth Circuit stated that each defendant must be accorded thirty days in which to remove, because an alternative construction (i.e., the first-served rule) “would require us to insert ‘first’ before ‘defendant’ into the language of the statute.”\footnote{Id. at 533 (citation omitted). The Ninth Circuit has adopted a similar construction. See Destfino v. Reiswig, 630 F.3d 952, 955 (9th Cir. 2011) (“The removal statute speaks of ‘the defendant’—not ‘first defendant’ or ‘initial defendant’—and its most straightforward meaning is that each defendant has thirty days to remove after being brought into the case.”).\footnote{Brierly, 184 F.3d at 534.} The court viewed the rule of unanimity as permitting the earlier-served defendant to consent to the last-served defendant’s removal, even though the first defendant had failed in its removal attempt, because “holding otherwise would vitiate the removal application of the later-served defendants and thereby nullify our holding that later-served defendants are entitled to 30 days to remove the case to district court.”\footnote{54 Id. at 1204 n.1.}

In Bailey v. Janssen Pharmaceutica, Inc.,\footnote{55 536 F.3d 1202, 1204 (11th Cir. 2008).\footnote{Id. The notice of removal was within the last-served defendant’s thirty-day window because thirty days from receipt of service was July 22, 2006, a Saturday. Thus, pursuant to Fed. R. Civ. P. 6(a), the last day for filing was Monday, July 24, 2006. Id. at 1204 n.1.\footnote{Id. at 1205.} The Ninth Circuit has also endorsed this rationale. See Destfino, 630 F.3d at 955 (stating that the last-served defendant rule is “grounded in statutory construction, equity and common sense” and that the approach “treats all defendants equally, regardless of when they happen to be served.”).}} an Eleventh Circuit decision, the plaintiff sued three defendants in state court and served the defendants on three different dates: May 12, May 15, and June 22. All three defendants were represented by the same lawyer, who filed a notice of removal on July 24.\footnote{Id. at 1206. The Ninth Circuit has also endorsed this rationale. See Destfino, 630 F.3d at 955 (stating that the last-served defendant rule is “grounded in statutory construction, equity and common sense” and that the approach “treats all defendants equally, regardless of when they happen to be served.”).} The Eleventh Circuit noted the circuit split with respect to the first- and last-served defendant rules,\footnote{Id. at 1205.} but concluded that “both common sense and considerations of equity favor the last-served defendant rule. The first-served rule has been criticized by other courts as being inequitable to later-served defendants who, through no fault of their own, might, by virtue of the first-served rule, lose their statutory right to seek removal.”\footnote{Id. at 1206. The Ninth Circuit has also endorsed this rationale. See Destfino, 630 F.3d at 955 (stating that the last-served defendant rule is “grounded in statutory construction, equity and common sense” and that the approach “treats all defendants equally, regardless of when they happen to be served.”).}
The Eleventh Circuit further noted that the last-served rule was consistent with both the rule of unanimity and a strict construction of the removal statutes:

[The last-served rule] is not inconsistent with the rule of unanimity. Earlier-served defendants may choose to join in a later-served defendant's motion or not, therefore preserving the rule that a notice of removal must have the unanimous consent of the defendants. The unanimity rule alone does not command that a first-served defendant's failure to seek removal necessarily waives an unserved defendant's right to seek removal; it only requires that the later-served defendant receive the consent of all then-served defendants at the time he files his notice of removal. . . . [W]e do not find that a strict construction of the removal statute necessarily compels us to endorse the first-served defendant rule. . . .

In addition, the Eleventh Circuit emphasized that the Supreme Court's Murphy Brothers decision supported the last-served defendant rule and stated that absent the Murphy Brothers decision, "the issue of which rule to endorse would be a closer call." In light of Murphy Brothers' significance to the Eleventh Circuit, and its mention by both the Eighth and Ninth Circuits, some brief exploration of that case is warranted at this juncture.

The issue in Murphy Brothers involved which of two events triggered the thirty-day removal window: the date that the defendant was formally served with process, or a previous date when the defendant had been faxed a courtesy copy of the complaint. The answer might have seemed obvious until one reviewed the relevant statutory language, which states that "[t]he notice of removal of a civil action or proceeding shall be filed within thirty days after the receipt by the defendant, through service or otherwise, of a copy of the initial pleading."

---

59 Bailey, 536 F.3d at 1207. The Ninth Circuit has endorsed this rationale as well. See Destfino, 630 F.3d at 956 (observing that "the fact that a defendant hasn't taken the initiative to seek removal doesn't necessarily mean he will object when another defendant does").


61 Bailey, 536 F.3d at 1208.

62 See Destfino, 630 F.3d at 956 (citing Murphy Bros. as exemplifying the Supreme Court's relaxation of the traditional axiom that the removal statutes are to be strictly construed); see also Marano Enters. of Kan. v. Z-Teca Rests., L.P., 254 F.3d 753, 756 (8th Cir. 2001) (concluding that Murphy Brothers indicated that "if faced with the issue before us today, the [Supreme] Court would allow each defendant thirty days after receiving notice within which to file a notice of removal, regardless of when—or if—previously served defendants had filed such notices").
setting forth the claim for relief upon which such action or proceeding is based. . . . \textsuperscript{63}

On the basis of the reference to “or otherwise,” the Eleventh Circuit had held that the defendant’s receipt of the faxed courtesy copy of the complaint started the thirty-day removal window.\textsuperscript{64} The Supreme Court reversed and stated that the removal provisions are subject to the “bedrock principle” that “[a]n individual or entity named as a defendant is not obliged to engage in litigation unless notified of the action, and brought under a court’s authority, by formal process.”\textsuperscript{65} The Court went on to explain that the statutory language reflected Congress’s attempt to create a uniform rule that would accommodate the vagaries of state provisions—in particular, under the practices of some states, service of process was considered to commence the action and service could precede the filing of the complaint, which created the potential that the removal window could close before the defendant had seen the complaint.\textsuperscript{66} However, the majority observed, “Nothing in the legislative history . . . so much as hints that Congress, in making changes to accommodate atypical state commencement and complaint filing procedures, intended to dispense with the historic function of service of process as the official trigger for responsive action by an individual or entity named defendant.”\textsuperscript{67} Further, the majority noted, fax machines did not exist at the time that Congress drafted this provision, so it could not have anticipated this specific scenario.\textsuperscript{68}

Murphy Brothers involved a single defendant rather than multiple defendants, and thus did not discuss the first- or

\textsuperscript{64} See Murphy Bros., 526 U.S. at 349.
\textsuperscript{65} Id. at 347.
\textsuperscript{66} Id. at 351. The Senate Report explained the problem and the statutory accommodation as follows:

In some States suits are begun by the service of a summons or other process without the necessity of filing any pleading until later . . . . [T]his places the defendant in the position of having to take steps to remove a suit to Federal court before he knows what the suit is about. As said section is herein proposed to be rewritten, a defendant is not required to file his petition for removal until 20 [now 30] days after he has received (or it has been made available to him) a copy of the initial pleading filed by the plaintiff setting forth the claim upon which the suit is based and the relief prayed for. It is believed that this will meet the varying conditions of practice in all the States.

\textsuperscript{67} Murphy Bros., 526 U.S. at 352-53.
\textsuperscript{68} Id. at 353 n.5.
last-served defendant rules. Moreover, by emphasizing service of process as critical to triggering the thirty-day removal window, the Court expressed no opinion in support or opposition to the competing first- and last-served defendant rules, both of which rely upon timing rules dating from service of process and nothing else. Nevertheless, the Eighth, Ninth, and Eleventh Circuit Courts of Appeals have relied on the case in endorsing the last-served defendant rule. All three circuits have characterized Murphy Brothers as representing a shift away from the traditional strict construction of the removal statutes and relied on language in Murphy Brothers in stating that defendants “are not required to take action . . . until they are properly served, ‘regardless of when—or if—previously served defendants had filed such notices.’” Therefore, according to these three circuits, the first-served defendant rule adopts a construction that would obligate a defendant to seek removal before receiving formal process.

This leads us to the third view addressing removal in the context of the staggered service of multiple defendants, referred to by its proponent circuit as “the intermediate rule.”

C. The Articulated Principles and Policies Motivating the Intermediate Rule

In what is now a familiar theme, the intermediate rule once again attempts to reconcile (1) the statutory thirty-day window within which a defendant must effect removal; (2) the axiom that courts should strictly interpret the removal statutes; and (3) the rule of unanimity. The intermediate rule, as described by the U.S. Court of Appeals for the Fourth Circuit, “requires a notice of removal to be filed within the first-served defendant’s thirty-day window, but gives later-served defendants thirty days from the date they were served to join the notice of removal.”

---

69 See supra note 62 and accompanying text.
70 See Bailey v. Janssen Pharmaceutica, Inc., 536 F.3d 1202, 1208 (11th Cir. 2008) (quoting Marano Enters. of Kan. v. Z-Teca Rests., L.P., 254 F.3d 753, 756 and citing Murphy Bros., 526 U.S. 344); see also Destfino v. Reiswig, 630 F.3d 952, 956 (9th Cir. 2011). Murphy Bros. takes no side in the first- versus last-served defendant debate. It simply rules out as a triggering date something other than service of process.
71 See Amy G. Doehring, Eleventh Circuit Adopts Last-Served Defendant Rule for Removal, A.B.A. LITIG. NEWS, Oct. 9, 2008, at 2. However, such a defendant might—for a variety of reasons, including the plaintiff’s default—never be served and thus play no role in removal.
The Fourth Circuit's en banc decision in Barbour v. International Union is the official source of the intermediate rule and builds upon language in a previous decision from the same circuit. Despite labeling this position as an intermediate rule, which suggests that it adopts a compromise position between the first-served and last-served defendant rules, the intermediate rule does not operate as a compromise measure—it is, in effect, consistent with the first-served defendant rule.

In Barbour, Chrysler Corporation retirees sued the International Union and two local unions—Local 1183 and Local 1212—in state court for alleged negligence and negligent misrepresentation in advising the plaintiffs to retire, which caused the plaintiffs to lose eligibility for a retirement incentive package known to the defendants but not publicly announced until two weeks after the plaintiffs retired. The three defendants filed a joint notice of removal on April 28, more than thirty days after service upon the first defendant, within thirty days of service upon the second defendant, and before service upon the third defendant. The en banc Fourth Circuit concluded that "because [the first-served defendant] did not seek removal within its thirty-day window, the plain language of [section] 1446(b) dictates that it forfeited its right to removal." Thus, under the so-called intermediate rule, the time for removal expires thirty days after the first defendant is served, without regard to when other defendants are served—the same result required by the first-served defendant rule.

Although Barbour makes much of according later-served defendants a full thirty days under the intermediate rule in which to decide whether to join or challenge the existing removal notice, the distinction between the intermediate rule and the first-served defendant rule in this regard is more subtle than Barbour suggests. Under the rule of unanimity, the

73 Id.
75 Barbour, 640 F.3d at 602-03.
76 Id. at 604.
77 Id. at 611.
78 Id. at 607 ("Like the First-Served Defendant Rule, the McKinney Intermediate Rule requires a notice of removal to be filed within the first-served defendant's thirty-day window, but gives later-served defendants thirty days from the date they were served to join the notice of removal."); see also id. at 611 ("If the first-served defendant files a notice of removal, later-served defendants have ample time—thirty days from the date that each such defendant is served—to decide whether to join the notice of removal . . . .").
defendants who have been joined and served must all consent to removal. Later-served defendants, however, are not bound by the removal notice—they may move to remand and thereby defeat removal. Accordingly, the distinction between the intermediate rule and the first-served defendant rule is seen when subsequently served defendants are served after the first defendant, but before the expiration of the first-served defendant's thirty-day window. For example, if Defendant #1 is served on Day 1 and Defendant #2 is served on Day 10, under the first-served defendant rule Defendant #2 has twenty days to decide whether to join the notice of removal, whereas under the intermediate rule he would have thirty days. This difference is hardly a compromise between the first-served and last-served defendant rules. The intermediate rule offers a relatively minor difference of statutory interpretation by giving every subsequently served defendant a full thirty days to evaluate whether to join an existing removal notice, but adds nothing to the statute's motivating principles and policies.\footnote{See id. at 605 (“Removal statutes . . . must be strictly construed, inasmuch as the removal of cases from state to federal court raises significant federalism concerns.”); id. at 613 (stating that the intermediate rule is consistent with “constru[ing] removal statutes narrowly and that doubts concerning removal should be resolved in favor of state court jurisdiction”); id. at 611 (noting that 28 U.S.C. § 1446(b) provides a thirty-day removal window and “[i]f you do not seek removal within the thirty-day window, you have forfeited your right to remove”); id. (rule of unanimity); id. at 614 (“All three of the rules before the court are consistent with the rule of unanimity, because each of them requires all of the defendants at some point in time to unanimously agree to removal.”).}

In sum, due to section 1446(b)'s silence on the issue, and despite ostensibly relying on the same principles and policies, the circuits are divided in their application of the statute in the multiple-defendant context. Indeed, even some of the challenges in reading the statute are the same: one of the justifications sometimes proffered in favor of the last-served rule—that the alternative first-served approach would require reading “first served” into the statute's text\footnote{See, e.g., Brierly v. Alusuisse Flexible Packaging, Inc., 184 F.3d 527, 533 (6th Cir. 1999).}—is itself subject to the same challenge (i.e., the last-served defendant approach similarly requires reading “last served” into the statute).

Choosing among these approaches has been made more difficult because the courts' decisions have failed to recognize the practical and conceptual dilemma that underlies all the analyses: the tension between plaintiff control and defendant-initiated removal.
D. The Unseen Tension Behind the Rules: The Power to Select the Forum

Although the majority of federal courts addressing the issue follows the first-served defendant rule, some courts and commentators have referred to the alternative last-served rule as the current “trend.” All but a few court decisions fail to acknowledge the unexpressed rationale or concern underlying the choice between the first-served, last-served, and intermediate rules: limiting the defendant’s opportunity to remove (the first-served and intermediate rules) versus maximizing the defendant’s opportunity to remove (the last-served rule). In selecting which rule to follow, courts must strike a balance between a plaintiff’s right to select the forum versus a defendant’s right to remove to federal court, and this ultimate choice drives judicial policy. Often—and certainly more often than is acknowledged—the outcome may depend on the court’s normative approach as more pro-plaintiff or more pro-defendant.

1. Plaintiff’s Control of the Forum

In Brown v. Demco, Inc., the U.S. Court of Appeals for the Fifth Circuit elected to follow the first-served defendant rule and characterized it as “[t]he general rule.” Although Brown was cast in an unusual posture—the matter had been proceeding in state court for four years and removal was sought only after the plaintiff added a new defendant—concerns that removal would unfairly benefit the defendants and cause an unfair detriment to the plaintiff

---


82 See Destfino v. Reiswig, 630 F.3d 952, 956 (9th Cir. 2011) (“The trend in recent case law favors the later-served defendant rule.”); Bailey v. Janssen Pharmaceutica, Inc., 536 F.3d 1202, 1205 (11th Cir. 2008) (observing that “the trend in recent case law favors the last-served defendant rule”); Gen. Pump & Well, Inc. v. Laibe Supply Corp., No. CV607-30, 2007 WL 3238721, at *2 (S.D. Ga. Oct. 31, 2007) (“More recently, . . . the trend in the case law has been toward the later-served rule.”); Hagins, supra note 81, at 426 (“There is a trend away from the first-served defendant rule.”); Hale, supra note 32, at 381 (noting that a “concern often cited as supportive of the last-served rule is the need to prevent a tactical advantage by the plaintiff in manipulating the removal statute in order to prevent removal to federal court”); Matthew C. Lucas, Diversity Jurisdiction Removal in Florida, 77 FLA. B.J. 54, 57 (2003) (noting that the last-served defendant rule “appears to be gaining acceptance in the courts”).

83 See supra notes 39-43 and accompanying text.

84 Brown, 792 F.2d at 478.
clearly influenced the court. The court noted “the axiom that the removal statutes are to be strictly construed against removal,” and stated, in a particularly revealing passage,

Here all of the appellees but [one] not only let the thirty-day period elapse, but also defended this action in state court for four years.... To permit the defendants in this case to obtain removal after they have tested state-court waters for four years would give them a second opportunity to forum-shop and further delay the progress of the suit. The unfairness of this to the plaintiff outweighs the unfairness, if any, to the last-joined defendant. The forum for a suit ought to be settled at some time early in the litigation.\(^{86}\)

The Brown Court’s analysis is worth a deeper review. First, the court relied on the strict construction rule, referring to construing the statutes “strictly... against removal,” thus indicating its preference for protecting the plaintiff’s choice of forum.\(^{87}\) Second, the court followed the strict construction rule with the characterization that the last-joined defendant’s attempt to remove presented an opportunity to “forum-shop and further delay the progress of the suit.”\(^{88}\) Third, after characterizing the defendants in an unfavorable manner (i.e., as seeking to delay the proceedings and employ a procedural route to a more favorable outcome), the court said such forum shopping and delay were unfair to the plaintiff, who should have assurances as to the forum “early in the litigation.”\(^{89}\) In Brown, the plaintiffs did not benefit from section 1446(b)’s one-year limitation on diversity-based removal; Congress did not enact the one-year provision until two years after the Brown decision.\(^{90}\) However, the absence of any outer time limit for removal at the time of the Brown decision arguably should have resulted in greater protection of the defendant’s right to remove, not less. The deference accorded to the plaintiff’s choice of forum is thus a powerful policy that has the ability not only to skew certain outcomes in a plaintiff’s favor, but to actually overcome the defendant’s statutory right to removal.

\(^{86}\) Id. at 482.
\(^{87}\) Id.
\(^{88}\) Id.
\(^{89}\) Id.
\(^{90}\) Brown v. Demco, Inc. was decided in 1986. See id. Congress added the one-year limitation to 28 U.S.C. § 1446(b) in 1988. See id. Moreover, see Barnes v. Westinghouse Elec. Corp., 962 F.2d 513, 515 n.7 (5th Cir. 1992) (noting that the one-year limitation was enacted on November 19, 1988).
2. Defendant’s Right to Removal and Potential Plaintiff Manipulation

The potential for manipulation by plaintiffs appears to play a role in many of the decisions that have chosen the last-served rule. As explained below, the first-served and intermediate rules encompass the possibility that a plaintiff suing multiple defendants might strategically use service of process to defeat removal. In contrast, the last-served rule eliminates this potential for manipulation by leaving open the removal option for later-served defendants.

One of the cases expressing concern about plaintiff manipulation most directly was White v. White, a federal district court decision in which circuit law bound the district court to follow the first-served defendant rule. However, due to concerns about plaintiff manipulation, the district court concluded that “exceptional circumstances” permitted removal. The district court stated that the plaintiff set a “removal trap” through “first serving an unsophisticated defendant who is least likely to attempt removal. Then the trap is sprung by not serving the more sophisticated defendants who are likely to attempt removal until 30 days has elapsed. Snap, removal is barred . . . .” Although White is one of the few cases to find the “exceptional circumstances” exception to the first-served defendant rule satisfied, its rationale fits justifications for the last-served defendant rule. As one commentator observed, White’s broad use of “forum manipulation” as an exceptional circumstance “hints that any plaintiff failing to serve every known defendant promptly at the same time runs the risk of failing in a motion to remand a removed case back to state court.”

The plaintiff manipulation concern has two alternative potential sources. On the one hand, when a court frames the issue as plaintiff manipulation, the court may view deference to

91 See, e.g., McKinney v. Bd. of Trs. of Mayland Cmty. Coll., 955 F.2d 924, 928 (4th Cir. 1992); White v. White, 32 F. Supp. 2d 890, 893 (W.D. La. 1998), abrogated in part by Murphy Bros., Inc. v. Michetti Pipe Stringing, Inc., 526 U.S. 344 (1999); see also Hale, supra note 32, at 381 (“A concern often cited as supportive of the last-served rule is the need to prevent a tactical advantage by the plaintiff in manipulating the removal statute in order to prevent removal to federal court.”).
92 32 F. Supp. 2d at 893.
93 Id. at 892-93.
94 Id. at 893.
95 Id.
96 Hagins, supra note 81, at 426 (expressing disbelief that “such a broad exception” would be considered “an exceptional one”).
the plaintiff’s choice of forum as desirable, but may view this particular plaintiff as undeserving because she engaged in manipulative behavior. On the other hand, courts may use broad-brush characterizations of plaintiff manipulation that reflect a more generalized distrust of plaintiffs with corresponding greater sympathy for defendants. In the specific context of adopting the last-served defendant rule, several federal courts have employed rationales reflecting a concern that the first-served rule was simply too pro-plaintiff.

II. THE VOLUNTARY-IN Voluntary Rule

Section 1446(b)—the underlying source of the first- and last-served defendant rules—is also the source of the voluntary-involuntary rule. Section 1446(b) provides, as relevant to this discussion,

If the case stated by the initial pleading is not removable, a notice of removal may be filed within thirty days after receipt by the defendant, through service or otherwise, of a copy of an amended pleading, motion, order or other paper from which it may first be ascertained that the case is one which is or has become removable. . . .

Although this statutory language suggests that the defendant has an entitlement to remove upon the specified receipt of a paper reflecting that a case initially nonremovable has become removable, there is a judicially created precondition: the case must have become removable due to the

---

97 More generally in the field of federal subject-matter jurisdiction, courts have relied on statutory authority to defeat manipulative efforts to invoke federal court jurisdiction, see 28 U.S.C. § 1359 (2006), but often have tolerated manipulative efforts to defeat federal court jurisdiction. See Gentie v. Lamb-Weston, Inc., 302 F. Supp. 161, 163 (D. Me. 1969) (noting that 28 U.S.C. § 1359 prohibits improper or collusive joinder to create federal jurisdiction but no similar statute bars collusive action to defeat federal jurisdiction); see also id. at 165-66 (noting that “many, though not all, federal courts have sustained the use of assignments to defeat diversity”). There is no pretense of balanced treatment in this area.

98 See, e.g., Destfino v. Reiswig, 630 F.3d 952, 956 (9th Cir. 2011) (stating that the first-served defendant rule could “encourage plaintiffs to engage in unfair manipulation by delaying service on defendants most likely to remove”); Collings v. E-Z Serve Convenience Stores, Inc., 936 F. Supp. 892, 894 (N.D. Fla. 1996) (stating that each defendant must be allowed thirty days in which to remove because to hold otherwise “opens the way for the plaintiff to deliberately avoid removal by delayed service upon a defendant anticipated to seek removal”); see also McKinney v. Bd. of Trs. of Maryland Cmty. Coll., 955 F.2d 924, 928 (4th Cir. 1992) (stating that under the first-served defendant rule, “the rights of defendants generally could be rather easily overcome by tactical maneuvering by plaintiffs”).

plaintiff's voluntary action. Unlike the first- versus last-served defendant rules, there is no circuit split here—every circuit addressing the issue has followed the voluntary-involuntary rule, although there are some differences among the circuits in the specifics of applying the rule. Accordingly, in a lawsuit filed in state court and based on state law, if the plaintiff voluntarily dismisses a nondiverse defendant, the remaining defendant(s) may remove the suit based on diversity. But if the court dismisses that same nondiverse defendant without the plaintiff's consent, then the remaining defendant(s) may not remove because the plaintiff's voluntary action did not accomplish the dismissal. The death of a nondiverse defendant is the sole exception to the voluntary-involuntary rule and permits removal.

The voluntary-involuntary rule—characterized by one commentator as having “a questionable pedigree” and “suspect” justifications—is a particularly interesting contradiction in removal jurisprudence. The rule's origins ostensibly come from two U.S. Supreme Court decisions a century ago. In Powers v. Chesapeake & Ohio Railway Co., the Supreme Court held that the case became removable after the plaintiff voluntarily dismissed the nondiverse defendants. Subsequently, in

---

100 See generally John B. Oakley, Prospectus for the American Law Institute's Federal Judicial Code Revision Project, 31 U.C. DAVIS L. REV. 855, 998-99 (1998) (“Under the voluntary-involuntary rule, when a court dismisses a removal-defeating claim with the plaintiff's consent, the case becomes removable. However, when such a dismissal is without the plaintiff's consent, the case does not become removable despite the change in its structure.”).


102 See Heather R. Barber, Removal and Remand, 37 LOY. L.A. L. REV. 1555, 1583 (2004) (explaining that the Second Circuit defines “voluntary” more broadly than the other circuits to include situations “where the removability of the case is the result of a decision of the court,” and where the plaintiff elects not to appeal the court's decision); Jeff Fisher, Everybody Plays the Fool, Sometimes; There's No Exception to the Rule Procedural Misjoinder Is Not an Exception to the Voluntary-Involuntary Rule, 60 BAYLOR L. REV. 993, 999-1000 (2008) (noting that the circuit courts “sometimes disagree about what constitutes a voluntary act.”).


105 169 U.S. 92 (1898).

106 Id. at 102.
Whitcomb v. Smithson,\(^{107}\) the Supreme Court determined that the case did not become removable when the trial court dismissed the nondiverse defendant without the plaintiff's assent.\(^{108}\) As summarized by the Eleventh Circuit,

> [T]he long-standing, judicially created “voluntary-involuntary” rule . . . is a rule developed in diversity cases “that if the resident defendant was dismissed from the case by the voluntary act of the plaintiff, the case became removable, but if the dismissal was the result of either the defendant’s or the court’s action against the wish of the plaintiff, the case could not be removed.”\(^{109}\)

Although the court in the excerpt above seems to suggest that the voluntary-involuntary rule is limited to diversity cases, courts have applied the rule to both diversity and arising-under cases.\(^{110}\)

Congress amended the removal statutes in 1948 and again in 1949.\(^{111}\) Some courts and commentators have argued that the voluntary-involuntary rule did not survive the amendments because amended section 1446(b) states that “an amended pleading, motion, order or other paper” could render a case removable;\(^{112}\) the reference to an “order” as rendering a case removable, without more, seemed to suggest that voluntariness (or involuntariness) played no role.\(^{113}\)

\(^{107}\) 175 U.S. 635 (1900).

\(^{108}\) Id. at 638.

\(^{109}\) Insinga v. LaBella, 845 F.2d 249, 252 (11th Cir. 1988) (citing Weems v. Louis Dreyfus Corp., 380 F.2d 545, 546 (5th Cir. 1967)).

\(^{110}\) See People v. Keating, 986 F.2d 346, 348 (9th Cir. 1993) (“Here, this case was transformed into an action ‘arising under’ federal law not by the voluntary act of the plaintiff, but instead by action of a defendant. Since a voluntary act by the plaintiff has not rendered the case removable, it must remain in state court.”).


\(^{112}\) 28 U.S.C. § 1446(b) (2006); see also Lyon v. Ill. Cent. R.R., 228 F. Supp. 810, 811 (S.D. Miss. 1964) (“There is nothing in [amended section 1446(b)] from which it can be properly inferred that Congress intended that a removal could be effected only in the event the plaintiff voluntarily did something which removed the local defendant from the case.”); Weems, 380 F.2d at 546-47 (noting that “[t]he effect of [the 1949] amendment has been variously interpreted,” and that “[i]t is contended . . . that the [voluntary-involuntary] rule did not survive an amendment to the Judicial Code in 1949”).

\(^{113}\) See Weems, 380 F.2d at 547-49; see also Joan Steinman, Postremoval Changes in the Party Structure of Diversity Cases: The Old Law, the New Law, and Rule 19, 38 U. KAN. L. REV. 863, 872 n.25 (1990) (“C)ommentators had observed that it was not entirely clear whether the voluntary-involuntary distinction had survived the 1949 amendments to § 1446.”); Underwood, supra note 104, at 1100, 1106 (stating that “there is nothing in the language of the statute to suggest any intent to require federal courts to continue utilizing the voluntary/involuntary rule—the bare language of the statute at least hinting at the inverse,” and urging that the rule be abandoned).
The Supreme Court has explained the basic history leading up to the 1948 and 1949 amendments to section 1446(b); the concerns that motivated the amendments did not involve the voluntary-involuntary rule, but rather centered on the concern that the defendant have access to the complaint before the removal period commenced.\textsuperscript{114} Despite the potential argument that the statutory amendments eliminated the voluntary-involuntary rule, the federal circuit courts ultimately rejected this contention.\textsuperscript{115} One court explained,

\begin{quote}
We will not buck the trend, nor will we rehash the legislative history. Suffice it to say that when Congress referred to “a case which is or has become removable” in section 1446(b), Congress apparently intended to incorporate the existing definition of “removable,” a definition that included the voluntary/involuntary rule.\textsuperscript{116}
\end{quote}

This offers a plausible, but not a mandated, interpretation of the 1948 and 1949 amendments. The congressional report, which courts have cited for the proposition that Congress

\textsuperscript{114} Murphy Bros., 526 U.S. at 351-52.

Prior to 1948, a defendant could remove a case any time before the expiration of her time to respond to the complaint under state law. Because the time limits for responding to the complaint varied from State to State, however, the period for removal correspondingly varied. To reduce the disparity, Congress in 1948 enacted the original version of § 1446(b), which provided that “[t]he petition for removal of a civil action or proceeding may be filed within twenty days after commencement of the action or service of process, whichever is later.” . . . Congress soon recognized, however, that § 1446(b), as first framed, did not “give adequate time and operate uniformly” in all States. In States such as New York, most notably, service of the summons commenced the action, and such service could precede the filing of the complaint. Under § 1446(b) as originally enacted, the period for removal in such a State could have expired before the defendant obtained access to the complaint. To ensure that the defendant would have access to the complaint before commencement of the removal period, Congress in 1949 enacted the current version of § 1446(b): “The petition for removal of a civil action or proceeding shall be filed within twenty days [now thirty days] after the receipt by the defendant, through service or otherwise, of a copy of the initial pleading setting forth the claim for relief upon which such action or proceeding is based.”

\textsuperscript{115} See Poulos v. Naas Foods, Inc., 959 F.2d 69, 71-72 (7th Cir. 1992) (noting that although defendants argued that section 1446(b) had eliminated the rule, “e)very court of appeals that has addressed the voluntary/involuntary rule has held that it survived the enactment of section 1446(b)” (citation omitted)).

\textsuperscript{116} Id. at 72.
intended to retain the rule, is more ambiguous than the courts have suggested.  

A closer look at the cited report illustrates that the courts have lifted one particular quotation out of context. In fact, although we have set out the relevant portion of the report in full in the footnote below, the simple addition of the sentence preceding and the sentence following the lifted quotation make the context apparent:

The second paragraph of the amendment to [section 1446,] subsection (b) is intended to make clear that the right of removal may be exercised at a later stage of the case if the initial pleading does not state a removable case but its removability is subsequently disclosed. This is declaratory of the existing rule laid down by the decisions. (See for example, Powers v. Chesapeake etc., Ry. Co., 169 U.S. 92.)

---

117 Id. (citing a quote from the Senate Report that the amendment is “declaratory of the existing rule laid down by the decisions”).

118 The relevant description of the bill in full states:

Subsection (b) of section 1446 of title 28, U.S.C., as revised, has been found to create difficulty in those States, such as New York, where suit is commenced by the service of a summons and the plaintiff's initial pleading is not required to be served or filed until later.

The first paragraph of the amendment to subsection (b) corrects this situation by providing that the petition for removal need not be filed until 20 days after the defendant has received a copy of the plaintiff’s initial pleading.

This provision, however, without more, would create further difficulty in those States, such as Kentucky, where suit is commenced by the filing of the plaintiff’s initial pleading and the issuance and service of summons without any requirement that a copy of the pleading be served upon or otherwise furnished to the defendant. Accordingly the first paragraph of the amendment provides that in such cases the petition for removal shall be filed within 20 days after the service of the summons.

The first paragraph of the amendment conforms to the amendment of rule 81(c) of the Federal Rules of Civil Procedure, relating to removed actions, adopted by the Supreme Court on December 29, 1948, and reported by the Court to the present session of Congress.

The second paragraph of the amendment to subsection (b) is intended to make clear that the right of removal may be exercised at a later stage of the case if the initial pleading does not state a removable case but its removability is subsequently disclosed. This is declaratory of the existing rule laid down by the decisions. (See for example, Powers v. Chesapeake etc., Ry. Co., 169 U.S. 92.)

In addition, this amendment clarifies the intent of section 1446(e) of title 28, U.S.C., to indicate that notice need not be given simultaneously with the filing, but may be given promptly thereafter.


119 Id.
The federal courts have concluded that the reference to “removable” in the quotation above was intended to incorporate all existing case law refinements. This is certainly a plausible construction—in these amendments, Congress was focused primarily on addressing one particular issue (ensuring that defendants would have access to a copy of the complaint before the removal period commenced), and secondarily on clarifying that removal due to changed circumstances could occur late in the case proceedings. However, it is at least equally plausible that due to these same foci, Congress would have expressly included the voluntary-involuntary rule if it intended to preserve it.120 The lack of any statutory reference to the voluntary-involuntary rule—especially in an era of “plain meaning” statutory construction—suggests that courts should exercise extreme caution in continuing to import the rule in the absence of any specific statutory language. In particular, Powers, cited in the congressional report, had permitted the defendant to remove after the plaintiff dismissed claims against the nondiverse defendants despite the fact that these dismissals occurred, and thus removal was sought “when [the case] was called for trial.”121 Thus, Powers serves as an example of authorizing a defendant to remove on the basis of subsequent removability at a very late point in the proceedings, namely the eve of trial, but it is far less clear that Congress intended this reference to affirm the continued viability of the voluntary-involuntary rule. Lacking any clear statutory direction, the continued viability of the voluntary-involuntary distinction—if any justification remains—must rest on the articulated principles and policies

---


Extrinsic materials have a role in statutory interpretation only to the extent they shed a reliable light on the enacting Legislature’s understanding of otherwise ambiguous terms. ... Judicial reliance on legislative materials like committee reports, which are not themselves subject to the requirements of Article I, may give unrepresentative committee members—or, worse yet, unelected staffers and lobbyists—both the power and the incentive to attempt strategic manipulations of legislative history to secure results they were unable to achieve through the statutory text.

Id. at 568.

that courts have used to justify the rule. The next section addresses these policy pillars.

A. The Articulated Principles and Policies Motivating the Voluntary-Involuntary Rule

The voluntary-involuntary rule appears to have two primary directing purposes: (1) promoting judicial economy and (2) deferring to the plaintiff's choice of forum. The judicial-economy rationale comes entirely from circuit court decisions; the Supreme Court has never proffered a judicial-economy rationale for the rule. This judicial-economy rationale appears to stem from finality concerns. If a court dismisses a nondiverse defendant from the action through an involuntary dismissal and the dismissal is appealed, the potential exists that the appellate court could set aside the dismissal, which would destroy complete diversity.

---

122 See, e.g., Poulos, 959 F.2d at 72.

123 See Great N. Ry. Co. v. Alexander, 246 U.S. 276, 282 (1918) ("The obvious principle . . . is that . . . the plaintiff may by the allegations of his complaint determine the status with respect to removability of a case, arising under a law of the United States, when it is commenced, and that this power to determine the removability of his case continues with the plaintiff throughout the litigation, so that whether such a case nonremovable when commenced shall afterwards become removable depends not upon what the defendant may allege or prove or what the court may, after hearing upon the merits, in invitum, order, but solely upon the form which the plaintiff by his voluntary action shall give to the pleadings in the case as it progresses towards a conclusion.").

124 Jenkins v. Nat'l Union Fire Ins. Co., 650 F. Supp. 609, 613-14 (N.D. Ga. 1986) ("Study of the Supreme Court cases that developed the rule discloses that the voluntary-involuntary rule is not based upon an appealability/finality rationale . . . ."); Archibald, supra note 104, at 1386 ("Predictably, courts continuing to adhere solely to the Supreme Court's stated rationale for the rule have rejected these cases and, accordingly, have deemphasized the role of federal courts in allocating cases between state and federal forums.").


[If] the non-diverse party has been involuntary dismissed by order of the state judge, the plaintiff may choose to appeal the dismissal. Although complete diversity may temporarily exist between the parties, suggesting that removal is proper, diversity jurisdiction may ultimately be destroyed if the state appellate court reverses the dismissal of the non-diverse party. Therefore, some cases are not removable despite complete diversity between the parties.

Id.; see also Am. Car & Foundry Co. v. Kettelhake, 236 U.S. 311, 316 (1915) ("[W]here there is a joint cause of action against defendants resident of the same state with the plaintiff and a nonresident defendant, it must appear, to make the case a removable one as to a nonresident defendant because of dismissal as to resident defendants, that the discontinuance as to such defendants was voluntary on the part of the plaintiff, and that such action has taken the resident defendants out of the case, so as to leave a controversy wholly between the plaintiff and the nonresident defendant.").
In addition to the finality/appealability concern, “[t]here also appears to be a policy favoring a plaintiff’s right, absent fraudulent joinder, to determine the removability of his case.”\textsuperscript{126} The Supreme Court decisions that address the voluntary-involuntary rule cite only this second purpose.\textsuperscript{127} Courts have analogized the deference accorded to the plaintiff’s forum choice by the voluntary-involuntary rule to the Mottley rule\textsuperscript{128} in arising-under cases, whereby the presence (or absence) of arising-under jurisdiction is determined by the allegations in the plaintiff’s well-pleaded complaint, without regard to the defendant’s pleadings or the defendant’s anticipated defenses.\textsuperscript{129} However, the analogy of the voluntary-involuntary rule to Mottley raises its own issues and ultimately leads right back to the underlying tension between according deference to the plaintiff’s choice of forum and the defendant’s right to remove.

B. Mottley and the Deference Debate

Courts have analogized the voluntary-involuntary rule, founded upon deference to the plaintiff’s choice of forum, to the Mottley “well-pleaded complaint” rule in arising-under cases.\textsuperscript{130} However, the Mottley analogy is less helpful—and less apt—than it appears initially. First, the Mottley rule is not always clear in application. In applying the Mottley rule, a court should disregard the defendant’s pleadings and examine only

\textsuperscript{126} Insinga v. LaBella, 845 F.2d 249, 253 (11th Cir. 1988). Concerns regarding plaintiff manipulation have been partially addressed by declining to apply the voluntary-involuntary rule to situations involving fraudulent joinder. See id. (“absent fraudulent joinder”); see also Great N. Ry. Co., 246 U.S. at 282 (“The obvious principle of [the voluntary-involuntary rule] is that, in the absence of fraudulent purpose to defeat removal, the plaintiff may by the allegations of his complaint determine the status with respect to removability of a case . . . .”). With respect to fraudulent joinder, see generally Fisher, supra note 102, at 1012-15 (arguing that “[p]rocedural misjoinder and fraudulent joinder behave almost identically,” and that neither is technically an “exception” to the voluntary-involuntary rule; rather, the voluntary-involuntary rule simply “should not be applied to those claims”); see also Underwood, supra note 104, at 1018 (“Stated simply, fraudulent joinder is a doctrine that permits federal courts to essentially ignore the inclusion in a lawsuit of a nondiverse party who would otherwise destroy federal diversity jurisdiction when the district court concludes that the party’s joinder is a sham.”).

\textsuperscript{127} See Jenkins, 650 F. Supp. at 613-14 (“What emerges from an examination of the Supreme Court cases on the voluntary-involuntary rule is the conclusion that the rule is not based upon an appealability/finality rationale but upon a policy favoring the plaintiff’s ‘power to determine the removability of his case.’”).


\textsuperscript{129} See Insinga, 845 F.2d at 253 (noting “[t]he common origins of the voluntary-involuntary rule with Mottley and its progeny in federal question cases”).

\textsuperscript{130} See supra notes 123, 126-30 and accompanying text.
the complaint. Next, the court must ascertain whether the complaint's allegations support arising-under jurisdiction as “well pleaded” and that the complaint does not include anticipated federal defenses. This review can be more difficult than one might think, and a plaintiff may draft her complaint in a manner specifically intended to keep the action in state court by scrupulously avoiding the inclusion of any apparent basis for federal subject matter jurisdiction—yet nevertheless find herself in the very federal court that she had sought to avoid. One prominent example of such a circumstance occurred in Grable & Sons Metal Products, Inc. v. Darue Engineering & Manufacturing. In Grable, the plaintiff constructed its lawsuit as a quiet title action filed in Michigan state court, only to find its suit removed to federal court on the basis of arising-under jurisdiction. According to the U.S. Supreme Court, the quiet title action, although a state-law claim, necessarily raised the federal issue of whether the Internal Revenue Service had given Grable the notice required by section 6335 of the Tax Code before seizing Grable's property to satisfy a federal tax delinquency and then subsequently selling the property to Darue. The Court held that this federal issue was of sufficient importance to invoke arising-under jurisdiction. The fact that a sufficiently necessary, albeit latent, federal issue lay within the state claim took away the plaintiff's preferred state forum and substituted a federal one.

Of course, Grable's federal issue, although latent, existed from the very outset of the litigation, whereas situations involving the voluntary-involuntary rule, by definition, arise due to some change occurring subsequent to the filing of the lawsuit. And if this distinction is not enough,
analysis of the Mottley rule arises in a contextually distinct environment from the voluntary-involuntary analysis. The Mottley rule always concerns the four corners of the complaint and thus always has as its vantage point the outset of the litigation, whereas the voluntary-involuntary rule always concerns some later change in the contours of the litigation. Moreover, although the Mottley rule applies only to arising-under jurisdiction, the voluntary-involuntary rule applies to both arising-under and diversity, which creates the potential for a plaintiff to “double dip”—to obtain the benefit of the Mottley rule in determining the existence of arising-under jurisdiction in the first instance, plus the benefit of the voluntary-involuntary rule after subsequent changes. So even assuming that the Mottley analogy is still apt, it appears that the voluntary-involuntary rule as currently applied may accord too much deference to the plaintiff's choice of forum. Just as the Mottley rule cannot insulate a plaintiff from arising-under jurisdiction that actually exists, so too the voluntary-involuntary rule should not generally insulate a plaintiff from removal when federal jurisdiction actually exists.

Although courts have expressly applied the Mottley analogy only to the plaintiff's control rationale, the Grable decision serves as a reminder that forum selection is a two-way, rather than a one-way, street. If plaintiffs were accorded complete control over forum selection, the removal statutes would be rendered pointless. As noted by one commentator, “Removal does not deprive plaintiffs of any ‘right,’ but merely affords defendants an equal opportunity to litigate in federal court....Additionally, removal does not expand federal jurisdiction, but merely allows cases involving federal jurisdiction to be heard in a federal court.”137 Accordingly, the voluntary-involuntary rule would appear to rest on a largely empty analytical basis.

III. THE ONE-YEAR LIMITATION

In 1988, Congress amended section 1446(b) to provide that “a case may not be removed on the basis of [diversity jurisdiction] more than 1 year after commencement of the action.”138 This generates questions over a third issue of belated removal and yet another contradiction within the doctrine—an

137 Haiber, supra note 17, at 611-12.
absolute cutoff that applies to some, but not all, removal circumstances. Perhaps the primary issue with respect to this provision is why defendants seeking to remove on the basis of arising-under jurisdiction have no outer time limit, whereas defendants seeking to remove on the basis of diversity face a one-year time limit. The legislative history to the 1988 amendment suggests that “[t]he amendment addresses problems that arise from a change of parties as an action progresses toward trial in state court. . . . Removal late in the proceedings may result in substantial delay and disruption.”

However, imposing a one-year limitation on diversity removal, but not arising-under removal, indicates that the concern is not the potential disruption of ongoing state proceedings but simply reflects disfavor toward diversity jurisdiction. Despite the seemingly straightforward nature of this provision, two interpretive issues have plagued the courts: first, whether the one-year limitation reflects a jurisdictional bar or merely a procedural defect, and second, whether the one-year limitation applies to all diversity removals or only to those that were not removable originally. Although not as obvious, these issues once again expose the underlying tension between plaintiff’s-deference and defendant’s-right-to-remove.

The legislative documents reveal that in enacting the one-year limitation on diversity-based removal, Congress intended to reduce “the opportunity for removal after substantial progress has been made in state court.” Rather than attempting to define “substantial progress,” Congress

---

140 See Oakley, supra note 100, at 1002 (“This [one-year limitation] rule has been strongly and aptly criticized as a backhanded attack on diversity jurisdiction . . . .”); see also Underwood, supra note 104, at 1105 (“The fact that this concern does not pertain to federal question cases demonstrates an anti-diversity bias on the part of Congress.”).
143 See Foiles, 730 F. Supp. at 110.
adopted a flat one-year limit and thereby created a provision simultaneously underinclusive and overinclusive. The provision is underinclusive because it does not apply to removal based on arising-under jurisdiction even if the state court has made substantial progress in the case; the provision is also overinclusive by preventing diversity-based removal after one year even in those cases where no substantial progress has been made in state court. As a court reviewing one such case observed, “It is very difficult to see how a removal under the facts of this case can interfere with the state court proceedings when none have occurred because of plaintiffs’ decision to withhold service until the one year time limitation has expired.”

With respect to the competing jurisdictional versus procedural interpretations applicable to the one-year limitation, a jurisdictional approach to the one-year limit serves to bar outright any attempt to remove after one year.\textsuperscript{147} Such an approach results when courts apply a strict statutory construction to the removal statutes—a construction that, as we have seen, exalts the plaintiff-deference policy over the defendant’s statutory right to remove. This jurisdictional approach is far from uniform, however, with a number of courts concluding that the one-year limit is procedural and thus potentially subject to equitable considerations.\textsuperscript{148} Indeed, dicta

---


\textsuperscript{148} See Barnes v. Westinghouse Elec. Corp., 962 F.2d 513, 516 (5th Cir. 1992) (one-year limit is procedural, not jurisdictional, and thus can be waived); see also Tedford v. Warner-Lambert Co., 327 F.3d 423, 426 (5th Cir. 2003) (court may consider parties’ conduct in determining whether it is equitable to apply the one-year limit strictly); id. at 426 n.4 (citing cases concluding that the one-year limit was subject to equitable exceptions); Wise v. Gallagher Bassett Servs., Inc., No. Civ. J FM-02-2323, 2002 WL 2001529, at *1 (D. Md. Aug. 27, 2002) (finding that plaintiff had “engaged in ‘forum manipulation’ in an effort to defeat the defendant’s removal right,” and stating, “[i]t:
in the U.S. Supreme Court’s Caterpillar, Inc. v. Lewis decision referring to the one-year provision as “nonjurisdictional”\(^{149}\) suggest that the Court viewed the one-year limit as procedural.\(^{150}\) Perhaps an even more interesting issue is whether the one-year limit applies both to cases initially removable and those not initially removable.

A fuller excerpt from section 1446(b) aids in understanding the debate between applying the one-year limit to all cases or only those that were not initially removable. In its entirety, section 1446(b) provides:

The notice of removal of a civil action or proceeding shall be filed within thirty days after the receipt by the defendant, through service or otherwise, of a copy of the initial pleading setting forth the claim upon which such action or proceeding is based, or within thirty days after the service of summons upon the defendant if such initial pleading has then been filed in court and is not required to be served on the defendant, whichever period is shorter.

If the case stated by the initial pleading is not removable, a notice of removal may be filed within thirty days after receipt by the defendant, through service or otherwise, of a copy of an amended pleading, motion, order or other paper from which it may first be ascertained that the case is one which is or has become removable, except that a case may not be removed on the basis of jurisdiction conferred by section 1332 of this title more than 1 year after commencement of the action.\(^{151}\)

The one-year limitation for diversity-based removals appears as the final clause of the second paragraph above. By appearing in this particular place within the statute, the one-year limitation seems to act as a modifying or qualifying phrase only with respect to the second paragraph of section 1446(b), and some courts have therefore applied it only to cases that initially were nonremovable.\(^{152}\) However, other courts have

\(^{149}\) Caterpillar, Inc. v. Lewis, 519 U.S. 61, 75 n.13 (1996); see also Henderson ex rel. Henderson, 131 S. Ct. at 1202-03 (noting that rules other than those governing a court's subject-matter or personal jurisdiction should be deemed nonjurisdictional); Reed Elsevier, Inc., 130 S. Ct. at 1243-44 (same); Union Pac. R.R. Co., 130 S. Ct. at 596 (same).

\(^{150}\) See Caterpillar, Inc., 519 U.S. at 75 n.13 (referring to the one-year provision as “nonjurisdictional”).


\(^{152}\) See, e.g., Brierly v. Alusuisse Flexible Packaging, Inc., 184 F.3d 527, 534-35 (6th Cir. 1999) (concluding that if Congress had intended the one-year limit to apply to all diversity-based removals it would have stated so more clearly).
applied the one-year limit to all diversity-based removals, thereby interpreting the limitation to apply both to initially removable cases as well as to those that were not initially removable.\footnote{See Rezendes v. Dow Corning Corp., 717 F. Supp. 1435, 1437 (E.D. Cal. 1989).} The courts that have offered this interpretation have emphasized that they must strictly construe the removal statutes against removal—an approach, as we have seen, used to accord deference to the plaintiff’s choice of forum over the defendant’s right to remove.\footnote{To the extent that one might question whether the one-year limitation negatively impacts defendants, one need only look to the Class Action Fairness Act of 2005, which was heavily promoted by defendant interests and resulted in eliminating the one-year rule in certain class actions. See 28 U.S.C. § 1453.}

With the competing policies of the plaintiff’s right to forum selection and the defendant’s right to remove to federal court now more fully revealed, we can now set out a framework that will generate a more consistent approach to the complex set of issues generated by removal involving section 1446(b).

IV. TOWARD A MORE CONSISTENT CONSTRUCTION OF SECTION 1446(b)

This article has examined three contradictory provisions within section 1446(b), two implied and one express, that arise within the removal context: the first-served/last-served/intermediate rules, the voluntary-involuntary rule, and the one-year limitation on diversity-based removal. A key insight into the resolution of these issues, whether by the courts or by Congress, is recognizing that the debates are not simply disputes over removal doctrine, but reflect the underlying tension in American procedure between plaintiff choices and defendant responses. In this part, we will analyze and synthesize the issues raised by these provisions in order to set out a more consistent analytical framework for considering late-arising removal efforts.

A. Underlying Policies

We begin with a review of policies, again both express and implied, that underlie and motivate these provisions. There are at least six such policies: (1) the statutory language itself, in its provision for removal by defendants as a general matter and in its more specific provisions of a thirty-day window for removal and of the one-year limitation on diversity-based removal; (2) the axiom
that courts should interpret the removal statutes strictly; (3) the judicially imposed rule of unanimity; (4) the promotion of judicial economy; (5) deference to the plaintiff's choice of forum; and (6) an apparent disfavor of diversity jurisdiction. These policies reflect at least three inherent contradictions critical to the construction of a more consistent framework.

First, the policies mix deference to the plaintiff's choice of forum with the defendant's statutory entitlement to removal. Second, the policies mix strict statutory construction principles with the addition of judicially created conditions and rules. Third, the policies mix an articulated goal of judicial economy with the potential for removing cases to a new federal forum after they have long lingered in a state forum. Our next step is to ask whether any of these contradictions yields ready answers or contributes to a potential analytical framework.

The mix of strict statutory construction with judicially created conditions and rules is a somewhat common situation without a ready solution, because it does not necessarily require compromise. Instead, courts and legislators could pursue a range of possible options. Courts could strictly construe the removal statutes and prohibit any supplementation with judicially created conditions and rules. Alternatively, they could honor any number of judicially created rules in addition to, or in explication of, the statutory language. In light of the current prevalence of a “plain language” approach to statutory construction, and for the sake of clarity, Congress could amend section 1446(b) to include any desired judicially created rules expressly and mandate that any rules not so included are expressly rejected.

The mix of judicial-economy concerns with the potential for late removal similarly does not yield any ready resolution: some sort of compromise appears required, yet the strength of these competing interests yields a range of choices. This leaves the mix of deference to the plaintiff's forum choice with the defendant's right to remove.

---

155 See, e.g., Louisville & Nashville R.R. v. Mottley, 211 U.S. 149, 153 (1908) (creating “well pleaded complaint” rule in federal arising-under cases, despite the absence of any such express requirement in the federal arising-under statute); Chi., Rock Island, & Pac. Ry. Co. v. Martin, 178 U.S. 245, 248 (1900) (reaffirming “rule of unanimity” when defendants seek to remove a civil action from state to federal court, despite the absence of any such express requirement in the federal removal statute); Strawbridge v. Curtiss, 7 U.S. (3 Cranch) 267, 267 (1806) (creating “complete diversity” rule in federal diversity jurisdiction cases, despite the absence of any such express requirement in the federal diversity statute).
In examining the conflict between preserving the plaintiff's choice of forum and honoring the defendant's right to remove to federal court when the statutory preconditions are satisfied (which also implicates interpreting the removal statutes strictly), compromise is an absolute necessity. If the policy of deferring to the plaintiff's forum choice was not subject to compromise, then the plaintiff's forum choice would become absolute and the removal statutes would serve no purpose—deferring to the plaintiff's choice of forum would, absent compromise, constitute both the beginning and the end of the discussion and would displace any potential for removal. Accordingly, courts must recognize the axiom regarding deference to the plaintiff's choice of forum for what it is—a starting point but not the only point of consideration. The defendant's right of removal is exactly that—a right, so long as the defendant satisfies the statutory prerequisites. A 1907 U.S. Supreme Court decision stated this plainly and directly:

"[T]he Federal courts may, and should, take such action as will defeat attempts to wrongfully deprive parties entitled to sue in Federal courts of the protection of their rights in those tribunals... Federal courts should not sanction devices intended to prevent a removal to a Federal court where one has that right."156

Some cases have expressly articulated a mistaken belief that the plaintiff's choice of forum should trump the defendant's statutory right of removal.157 Those cases stand in stark contrast to other, older case decisions clearly stating that a defendant's right of removal is equal in stature, and of the same constitutional dimension, as a plaintiff's right to select a forum.158 As the Supreme Court has said, to allow plaintiffs to

157 See, e.g., Marathon Oil Co. v. Ruhrgas, 145 F.3d 211, 219 n.11 (5th Cir. 1998) ("The defendant's right to remove and the plaintiff's right to choose the forum are not equal."), rev'd, 526 U.S. 574 (1999); Auchinleck v. Town of LaGrange, 167 F. Supp. 2d 1066, 1069 (E.D. Wis. 2001) ("The plaintiff's right to choose his forum is superior to the defendant's right of removal.").

The constitution of the United States was designed for the common and equal benefit of all the people of the United States. The judicial power was granted for the same benign and salutary purposes. It was not to be exercised exclusively for the benefit of parties who might be plaintiffs, and would elect the national forum, but also for the protection of defendants who might be entitled to try their rights, or assert their privileges, before the same forum.

Id. at 348; see also Tex. & Pac. Ry. Co. v. Cody, 166 U.S. 606, 609 (1897) (referring to "defendant's constitutional right as a citizen of a different State than the plaintiff, to
“always elect the state court” renders the protection of diversity jurisdiction ineffective for defendants and “[s]uch a state of things can, in no respect, be considered as giving equal rights.”

Rather, Congress authorized removal so that defendants would not be “deprived of all the security which the constitution intended in aid of [their] rights.” By virtue of the fact that a defendant cannot automatically thwart the plaintiff’s choice of forum in every instance but instead can remove only under the circumstances prescribed by statute, the removal statutes constitute a congressional compromise between the interests of plaintiffs and defendants. It remains to apply this framework to resolve the issues that arise in the context of statutory construction—and accordingly, the next question becomes the extent to which a strict statutory construction should be modified by judicially created conditions and rules, which brings us full circle. At this point, a return to each of the three identified removal issues will provide the context necessary for our framework.

B. Applying Policies to the Rules

This section applies the policies identified above and illustrates how these policies impact each of the rules explored.

1. The First-Served, Last-Served, and Intermediate Rules

Returning first to the first-served, last-served, and intermediate rules, and assuming that the rule of unanimity is here to stay, the question becomes which of the three approaches strikes the better compromise between deferring to the plaintiff’s choice of forum and the defendant’s right to removal. The conclusion appears straightforward: the last-served defendant rule honors both the plaintiff’s right of forum selection and the defendant’s right of removal. Although some courts have claimed that the first-served rule does no injustice to defendants due to the rule of unanimity, the first-served


156 Hunter’s Lessee, 14 U.S. (1 Wheat.) at 349.

157 Id.

158 See, e.g., Brown v. Demco, Inc., 792 F.2d 478, 482 (5th Cir. 1986).
rule is susceptible to potential manipulation: the plaintiff might serve less sophisticated defendants first in an attempt to preclude removal.\textsuperscript{162} The last-served rule deprives the plaintiff of no valid right or privilege—the plaintiff loses only the ability to manipulate the timing of service so as to potentially reduce the likelihood of removal—whereas the first-served rule potentially deprives later-served defendants of their right to removal. Accordingly, the last-served defendant rule appears to offer the better compromise.\textsuperscript{163}

2. The Voluntary-Involuntary Rule

The voluntary-involuntary rule presents the contradiction between deferring to the plaintiff's choice of forum and the defendant's right to removal in a very direct manner. Remember that deference to the plaintiff's choice of forum is the Supreme Court's sole articulated justification for the voluntary-involuntary rule.\textsuperscript{164} But it is unclear why deference to the plaintiff's choice of forum should trump the defendant's right to remove in the context of a subsequent change in circumstances.\textsuperscript{165}

Although the plaintiff is said to be the master of her claim,\textsuperscript{166} no obvious reason explains why the plaintiff should maintain ongoing control after filing the complaint, especially when such ongoing control implicates concerns that a plaintiff could manipulate amendments and dismissals in such a manner as to defeat the defendant's right to removal. The

\textsuperscript{162} See supra note 91 and accompanying text (providing examples).

\textsuperscript{163} See Howard B. Stravitz, Recocking the Removal Trigger, 53 S.C.L. REV. 185, 202 (2002) (opining that it "is undoubtedly correct that the first-served defendant rule unfairly shifts [the] balance in favor of plaintiffs"); see also McKinney v. Bd. of Trs. of Mayland Cnty. Coll., 955 F.2d 924, 927-28 (4th Cir. 1992) ("Congress created the removal process to protect defendants. It did not extend such protection with one hand, and with the other give plaintiffs a bag of tricks to overcome it." (quoting McKinney v. Bd. of Trs. of Mayland Cnty. Coll., 713 F. Supp. 185, 189 (W.D.N.C. 1989))).

\textsuperscript{164} See Jenkins v. Nat'l Union Fire Ins. Co. of Pa., 650 F. Supp. 609, 613-14 (N.D. Ga. 1986) ("What emerges from an examination of the Supreme Court cases on the voluntary-involuntary rule is the conclusion that the rule is not based upon an appealability/finality rationale but upon a policy favoring the plaintiff's 'power to determine the removability of his case.'" (quoting Great N. Ry. Co. v. Alexander, 246 U.S. 276, 282 (1918)).

\textsuperscript{165} See Underwood, supra note 104, at 1098 ("The voluntary/ involuntary rule is not just antiquated, but lacking any principled bases. It acts as merely another court-created doctrine designed to limit the ability of litigants to utilize the services of the federal tribunals, trampling on principles of federalism.").

\textsuperscript{166} See Holmes Grp., Inc. v. Vornado Air Circulation Sys., Inc., 535 U.S. 826, 831 (2002) (referring to plaintiff as "the master of the complaint" (quoting Caterpillar Inc. v. Williams, 482 U.S. 386, 398-99 (1996))).
plaintiff is entitled to the initial forum choice and permitted to attempt to structure her lawsuit in such a manner as to avoid federal subject-matter jurisdiction, such as by suing under state law rather than federal law, suing nondiverse defendants, or limiting the recovery sought to one below the jurisdictional threshold. However, the extreme deference to the plaintiff’s choice of forum reflected in the voluntary-involuntary rule seems especially incongruous in light of other ways that we permit defendants to alter the litigation, such as by adding parties, asserting counterclaims and cross-claims, and moving for transfers of venue. Extending the plaintiff’s control beyond the initial filing, so that defendants cannot exercise their right of removal in an instance where federal subject-matter jurisdiction plainly exists, unduly defers to the plaintiff’s forum selection at the expense of the defendant’s right of removal.

3. The One-Year Limitation

Finally, we turn to the one-year limitation on diversity-based removal. As an initial matter, no obvious rationale explains the purpose of imposing an outer one-year time limit on the removal of diversity cases, but no outer time limit whatsoever on the removal of arising-under cases. This presents two potential options: eliminating the one-year limitation on diversity-based removal so that there is no time limit for either diversity or arising-under cases, or imposing an outer time limitation on both diversity and arising-under removal.

In answer to concerns about the one-year limitation as tending to encourage plaintiff manipulation (such as waiting until the expiration of the one-year limit before dismissing a nondiverse defendant), the elimination of the one-year limit would remove this concern and put all bases for removal on the same footing. However, the lack of any outer time limit would permit removal after state courts have potentially invested substantial time and resources in the case, which is inconsistent with judicial-economy concerns. An appropriate compromise in this instance might be for Congress to implement an outer time limitation for all removal, regardless of whether the basis for federal subject-matter jurisdiction is diversity or arising-under. This would eliminate the apparent

167 See, e.g., FED. R. CIV. P. 14.
bias against diversity jurisdiction, would address the articulated concern about removal after state courts have made substantial investments in the case, and would resolve the current dispute as to whether the time limit applies to all cases or only to those not initially removable. Further, Congress could specify that the time limitation is subject to equitable considerations. This would eliminate the current dispute about whether the provision is jurisdictional or procedural; it would also serve to clarify that courts will not look favorably upon the perpetrators of strategic manipulation (e.g., failure to investigate by defendants, or delays in serving defendants or amending pleadings by plaintiffs). Importantly, Congress, as the creator of the statutory limitation on diversity-based removal, must make the choice about applicable amendments to current removal provisions.

CONCLUSION

When removal from state to federal court is delayed beyond the initial thirty days after the action is commenced, such subsequently instituted removal potentially implicates three complicating issues that are all rooted in section 1446(b): the first-served/last-served/intermediate rules, the voluntary-involuntary rule, and the one-year limitation on diversity-based removal. These three issues have developed independently, which has masked the potential for a unifying analytical framework. This article has identified the underlying policies, analyzed the inherent contradictions, and proposed resolutions more consistent with the dual and equal goals of honoring the plaintiff's choice of forum and honoring the defendant's right of removal.
INTRODUCTION

Law enforcement interviews are sometimes viewed as one of the least intrusive and least objectionable investigative techniques in the government’s counterterrorism arsenal. In theory, a law enforcement agent’s voluntary request for information from an individual for a counterterrorism investigation, or a border agent’s questioning of a person returning to the United States after traveling abroad, only minimally impinges on privacy and individual rights. Federal Bureau of Investigation (FBI) and Customs and Border Protection (CBP) interviews do not involve extreme interrogation methods, the imposition of criminal sanctions for speech, or the use of covert investigative tools hidden from public view—all policies that have attracted considerable public and scholarly attention.

Yet law enforcement interviews of U.S. Muslims in the terrorism context involve greater coercion and stigma than...
prevailing accounts recognize. Interviews are startlingly common: some estimates suggest that the FBI, for instance, has questioned hundreds of thousands of U.S. Muslims.⁴ FBI and CBP interviews alike have elicited widespread concern among U.S. Muslims as a result of the coercion involved, the content of the questioning, and the basis for interviewee selection. As personal, direct encounters between individuals in the U.S. Muslim community and the U.S. government, interviews are especially likely to inform targeted individuals’ and communities’ sense of “belonging” and inclusion as ethnic and religious minorities in the United States.

Where law enforcement agents select individuals for questioning primarily on the basis of their speech, associations, or other expressive activities protected by the First Amendment, interviews raise special concern. For instance, according to recent congressional testimony from a Muslim civil rights organization, the FBI questioned a computer programmer after he posted “political articles from mainstream news sources on his Facebook page”—approaching him at his workplace in front of colleagues and supervisors and potentially jeopardizing his job.⁵ The FBI contacted another man for questioning after a local newspaper published his nonviolent comments about the political situation in Pakistan.⁶

Tabbaa v. Chertoff was a rare case of such targeting to actually reach the courts. In Tabbaa, the CBP questioned, fingerprinted, photographed, and searched dozens of individuals returning to the United States after attending an Islamic conference in Toronto, applying the extra screening procedures “normally reserved for suspected terrorists.”⁷ The government had no individualized suspicion regarding any of the plaintiffs, all U.S. citizens, but carried out these procedures on those travelers who told border officials that they had attended the Toronto gathering.⁸ The conference drew over thirteen thousand participants and featured prominent Islamic speakers, musical

---

⁵ Tabbaa v. Chertoff, 509 F.3d 89, 92 (2d Cir. 2007).
⁶ Id. at 92, 94.
performances, spiritual reflection, and communal prayer." Border agents questioned plaintiffs about their activities at the conference, the content of the lectures, and the reasons they attended, and the detentions lasted between four and six hours.9

The CBP defended itself in Tabbaa by asserting that it acted based on information that individuals associated with terrorist activities would attend the conference and that the event might serve as a “meeting point” to plan terrorist activities or “exchange ideas and documents.”10 The agency had ordered border agents to ascertain the identities of conference participants, check their status on watch lists, and search luggage to find any evidence of terrorist plans, documents, or weapons.11 The Second Circuit Court of Appeals held that the measures significantly burdened plaintiffs’ freedom of association but ruled that the government’s security interests justified the intrusion.12

Tabbaa was wrongly decided: the court failed to question the notion that it was rational—and fair—to stop every person returning from a diverse gathering of thirteen thousand people on the possibility that she may have met a terrorist, and perfunctorily dismissed alternative methods of investigation.13 But it appropriately recognized the stigmatic harm from the screening measures and applied heightened scrutiny to the practice—engaging in a form of review other courts have declined to apply in First Amendment challenges to law enforcement investigations and surveillance.14

This article argues for heightened scrutiny of law enforcement interviews triggered by protected speech and association, which impose a substantial burden on those rights. Not all interviews based on First Amendment expression are wrong; speech or association may at times be a relevant basis for law enforcement inquiry. But the harms from such interviews call for careful scrutiny to determine whether a sufficient nexus exists between the First Amendment trigger for the scrutiny and an actual threat. Interviews based on First Amendment expression send a message to affected individuals and communities that their expressions of identity and participation in the public sphere are devalued, imposing stigma and chilling

---

8 Id. at 94.
9 Id. at 94, 98, 100.
10 Id. at 93.
11 Id. at 94.
12 Id. at 102-03.
13 Id. at 104.
14 Id. at 102.
expression. And there are historical reasons to question law enforcement interviews focused on First Amendment activities: in an earlier period of heightened fear over domestic and foreign threats, the FBI deliberately used interviews to suppress political speech and association by creating the impression that “there is an FBI agent behind every mailbox.”

There is now a growing literature on the effects of terrorism investigations on expression and association, yet the scholarship on First Amendment freedom of speech and association doctrine related to investigations remains scant. Several scholars briefly cite doctrinal obstacles to Free Speech Clause challenges to surveillance or investigations before turning their attention elsewhere. Moreover, the literature on law enforcement interviews in the terrorism context is almost nonexistent. Interviews are rarely the subject of new legislation, public announcements, or court challenges that attract public and scholarly notice. Even scholars arguing for greater protection of civil liberties often mention interviews only

18 The primary law review article found focusing on FBI or CBP interviews, though not from a First Amendment perspective, is Tracey Maclin, “Voluntary” Interviews and Airport Searches of Middle Eastern Men: The Fourth Amendment in a Time of Terror, 73 Miss. L.J. 471 (2003); see also Hussain, supra note 16, at 927-32 (addressing Tabbaa in presenting a free exercise theory for claims against targeting of religious expression).
to contrast them with more intrusive methods, thereby casting interviews as a relatively harmless investigative practice.¹⁹

This article addresses these gaps in the literature. Part I contextualizes FBI and CBP interviews of U.S. Muslims and describes three concerns they raise: the coercion of FBI and CBP encounters, the content of questioning, and the discriminatory basis for selection of interviewees. This part further argues that neither Fourth Amendment law nor internal regulations provides meaningful restrictions on interviews. Part II contends that law enforcement interviews that involve First Amendment profiling—the selection of an individual for law enforcement attention because of political, religious, or cultural expression or association—impose particularly grave stigmatic costs and chilling effects on individuals and communities. This part also maps out two separate normative concerns that the practice raises: a suppression concern about deliberate government attempts to suppress speech through an investigation and an overbreadth concern about the scope of an investigation triggered by expression, even where there is not an apparent illegitimate purpose. I argue that even without a suppression purpose, an investigation based on First Amendment profiling raises concern both because of the greater risk that hostility to expression influenced the scope of the investigation and because of the serious harms to individuals and communities.

Part III argues that courts ought to apply heightened scrutiny to interviews based on First Amendment profiling, and that existing First Amendment doctrine on free speech and association, while inconclusive, offers the potential for courts to do so. Some plaintiffs challenging FBI and CBP interviews should be able to surmount standing barriers that courts have erected in First Amendment cases, and further demonstrate, on the merits, substantial harm from investigations. Furthermore, lower courts have split as to whether heightened scrutiny is appropriate for reviewing First Amendment challenges to law enforcement investigations, and I counter the objection that narrow-tailoring requirements would impede critical law enforcement interests. Plaintiffs are most likely to succeed in challenges to First Amendment profiling where they can demonstrate tangible harm,

¹⁹ See Solove, supra note 16, at 175-76 (advocating a warrant requirement for government information-gathering implicating First Amendment values, but not in voluntary interviews); Fisher, supra note 16, at 673 (contrasting voluntary interviews with infiltration of organizations).
such as detention, reputational injuries, or economic costs, in addition to stigma and chilling effects.

A few words about the limited nature of my claims are in order. I do not contend that the “average” interview based on First Amendment profiling will trigger judicial scrutiny or establish a violation of the law. It is also beyond question that any challenge to law enforcement terrorism investigations will need to overcome judicial concerns about second-guessing the factual determinations of law enforcement agencies regarding the appropriateness of particular investigative measures. Yet in the terrorism context, any legal challenge to an investigative practice will face significant obstacles. But precisely because legal challenges in this area are difficult, any avenue for judicial review not foreclosed becomes significant. While I do not argue that plaintiffs will usually prevail in challenging First Amendment profiling, I do contend that ostensibly nonintrusive and unobjectionable law enforcement questioning can impose substantial harm on individuals and communities, and that the First Amendment’s guarantees of free speech and association offer the potential to contest these harms in an important segment of cases.

I. LAW ENFORCEMENT INTERVIEWS

Law enforcement agents question individuals for national-security purposes in a wide variety of contexts: immigration officials interview noncitizens applying for citizenship or permanent residency, local police question motorists stopped in traffic whose names trigger a watch-list match, prosecutors question witnesses before grand juries, and investigators interrogate suspects arrested for terrorism offenses. This article focuses on two of the most common forms of interviews in the counterterrorism context affecting immigrants and U.S. citizens alike—FBI interviews of individuals approached at home, work, or in their communities, and CBP interviews of individuals seeking to reenter the United States at airports or land borders after traveling abroad. The prevailing view in much of the legal doctrine and commentary, sometimes explicit but often unstated, is that interviews in either context inflict little harm, at least relative to other investigative methods, because they are minimally intrusive, overt, and involve “mere questioning.” This part argues that while FBI and CBP interviews serve an important national-security function, in practice, these interviews raise
serious concerns related to their coercion, the content of questioning, and the basis for selection of interviewees, and that existing Fourth Amendment doctrine and internal agency guidelines provide insufficient constraints.

A. FBI and CBP Interviews: Context and Concerns

Some scholars and officials have estimated that the FBI has conducted as many as two hundred thousand or half a million interviews of Muslims in the United States—staggering numbers, if accurate, given estimates that adult Muslims in the United States number fewer than two million. In the first three years following the September 11 attacks, the FBI carried out at least four well-publicized national rounds of interviews of Muslims and Arabs. These rounds included two interview campaigns of thousands of Arab male noncitizens based on demographic information suggesting “similarit[ies]” with al Qaeda terrorists, interviews of nearly ten thousand Iraqi immigrants, including U.S. citizens, before the invasion of Iraq; and interviews of Muslims in the months before the 2004 presidential election. But beyond these announced interview campaigns, the FBI continues to interview U.S. Muslims, in waves and individually, in order to investigate specific terrorist threats, gather general intelligence about communities,

20 Louise A. Cainkar, Homeland Insecurity: The Arab American and Muslim Experience After 9/11 113, 125 (2009) (citing statements in 2005 by a retired FBI counterterrorism official and the director of the MIT Center for International Studies). It is not clear what data these estimates relied on, and the FBI official’s reference to “half a million interviews” may have been intended as a rhetorical statement of the large number of interviews rather than an actual estimate.

21 A recent national study by the Pew Research Center estimated that there are 1.4 million adult Muslims in the United States, although estimates of the community’s size vary considerably. Pew Research Ctr., Muslim Americans: Middle Class and Mostly Mainstream 9-10 (2007).


25 See, e.g., Jake Armstrong, FBI in Lodi: Abusive or Just Assertive?, Lodi News-Sentinel, July 23, 2005 (describing questioning and surveillance of Muslims in Lodi, California after arrest of several residents on terrorism and/or immigration charges); Nathaniel Hoffman, Muslims Endure FBI Persistence in Lodi, Contra Costa Times, June 11, 2005, at A01 (same).
follow up on tips of suspicious activity called in from the public, or solicit people to act as undercover informants.

At U.S. international airports and land borders, the CBP questions returning travelers, including U.S. citizens, to intercept terrorists, weapons, and physical contraband as well as to collect intelligence for law enforcement agencies' broader use. Reports of actual interviews make clear that the intelligence collected is not limited to activities with a specific nexus to the border (such as a person's legal status in the United States or suspicious international travels), but includes information gathering on U.S. mosques and organizations within the United States. Thus, the agency uses its authority to search and question travelers at U.S. borders to acquire a range of information that law enforcement could not easily compel within the United States.

All travelers at U.S. borders can expect some scrutiny at the point of entry, including a review of identification and travel documents, and sometimes immigration-status questioning or luggage inspection. For most U.S. citizens, these encounters are brief, but the CBP pulls aside some individuals—including citizens—for protracted questioning or more intrusive searches.

26 See, e.g., Carrie Johnson & Robin Shulman, Probes Test Trust that Authorities Strove to Win from U.S. Muslims, WASH. POST, Oct. 5, 2009, at A03 (quoting retired FBI special agent describing agents' role as to "know everything that's going on" in a mosque or community); see also Alex Ransom, Muslims Feel Targeted by FBI, MERCURY (Dall.), Apr. 11, 2010, at 1.

27 See, e.g., Eric Bailey, FBI Questions High School Student over "PLO" Doodle, L.A. TIMES, Dec. 16, 2005, at B4 (describing questioning of 16-year-old based on allegation that student had doodled the initials "PLO" on a binder and stored pictures of suicide bombers on his cell phone).


31 See infra notes 65-75 and accompanying text.

These selections may occur either at “random” or as a result of factors like travel to particular countries, presence on a watch list, or undisclosed “risk factors” flagged by an automated program. These selection decisions do not require individualized suspicion.

Both FBI and CBP interviews serve indisputably important purposes. Following the September 11 attacks, the FBI shifted its focus from traditional law enforcement to intelligence gathering to detect and interrupt potential threats. With tens of thousands of threats and suspicious activities identified each year, interviews allow the FBI to gather information directly from individuals believed to have some information about a possible threat. In many cases, as the agency has argued, interviews allow the FBI to quickly rule out individuals who do not pose a real threat—preventing further scrutiny of innocent people and focusing scarce investigative resources on actual concerns.

The same need for efficient screening applies at the border. The CBP states that nearly 1.2 million travelers attempt to cross into the United States each day, and the agency intercepts about five hundred people a year out of terrorism or national-security concerns. According to the agency, border screening interviews have enabled it to prevent actual terrorists from entering.

---

35 Id. at 33-35.
36 Id. at 28-29; see also U.S. Customs & Border Prot., supra note 29.
37 See, e.g., Tabbaa v. Chertoff, 509 F.3d 89, 92 (2d Cir. 1997).
40 “Are You Part of a Revolution Trying to Overthrow the Government of the United States?,” PITTSBURGH CITY PAPER, Jan. 15, 2003, at 22 (quoting FBI explanation that interviewing donors to Muslim charities suspected of links to terrorists allows law enforcement to rule out those who innocently gave donations).
from entering; for example, based on suspicions raised in an interview, it denied entry to a Jordanian national who later killed 132 people in a suicide bombing abroad.\textsuperscript{42}

Despite the clear necessity for FBI and CBP interviews, the way in which these interviews are conducted raises three concerns. One concern relates to the coercion and intimidation interviewees face. Although FBI interviews are nominally voluntary, the tactics the FBI used in some interviews reported by the press or community organizations virtually compelled compliance.\textsuperscript{43} According to these accounts, FBI agents often approached people at work, where they could not refuse to cooperate without eliciting suspicion and fear of reprisal from employers already wary of Muslims; some individuals reportedly lost their jobs after workplace visits.\textsuperscript{44} FBI agents reportedly pressured some individuals to submit to questioning immediately, despite their stated desire to obtain a lawyer first.\textsuperscript{45} At other times, agents knocked on people's doors late in the evening or at night, which heightened the interviewees' perception of intimidation.\textsuperscript{46} In some cases, FBI agents misrepresented the purpose of an interview: agents told a person that they were investigating potential hate crimes against Muslims or conducting general community outreach while asking questions that focused on who the person knew and whether the interviewee presented a threat.\textsuperscript{47} And agents

\begin{footnotes}
\footnotetext[42]{Id. at 4.}
\footnotetext[43]{To be sure, this was not true as a universal matter: FBI agents sometimes told interviewees that the questioning was voluntary. See, e.g., James, supra note 23. In addition, some individuals declined interview requests. See, e.g., Armstrong, supra note 25 (reporting that several Muslims refused to appear for interviews). But the only quantitative evidence of individuals declining interview requests, from the government's initial post-9/11 interview program, suggests the numbers are miniscule. See Memorandum on Final Report on Interview Project from Kenneth L. Wainstein, Dir., Exec. Office for U.S. Att'y's, Dept of Justice, to the Att'y Gen. (Feb. 15, 2002) [hereinafter Final Report on Interview Project] (reporting that only 1 in 69 individuals in Oregon, 1 in 59 in Minnesota, and 8 of 313 in Eastern Michigan declined).}
\footnotetext[44]{See, e.g., Barbara Carmen, FBI Agents Stir Old Fears Among Iraqi-Americans, COLUMBUS DISPATCH, Apr. 4, 2003, at D1C (quoting Ohio Muslim leader stating that some people lost their jobs after FBI workplace visits); Tim Townsend, FBI Interviews Prompt Muslim Rights Project, ST. LOUIS POST-DISPATCH, Apr. 1, 2010, at A1 (noting workplace interviews); ACLU Sues, Says FBI Spying on Muslims, CHI. TRIB., Dec. 3, 2004, at 10 (same); CAINKAR, supra note 20, at 170 (same).}
\footnotetext[45]{Josh Richman, ACLU Sues Over Muslim Interviews, ALAMEDA TIMES-STAR, Oct. 23, 2004.}
\footnotetext[46]{See, e.g., ACLU Sues, Says FBI Spying on Muslims, supra note 44; "Are You Part of a Revolution," supra note 40.}
\footnotetext[47]{"Are You Part of a Revolution," supra note 40; COUNCIL ON AM.-ISLAMIC RELATIONS, GREATER L.A. AREA CHAPTER, THE FBI'S USE OF INFORMANTS, RECRUITMENT AND INTIMIDATION WITHIN MUSLIM COMMUNITIES 6 (2009) [hereinafter CAIR CALIFORNIA] (on file with author).}
\end{footnotes}
reportedly told others that if they refused to submit to an interview, the agents would arrest them.\(^48\)

In an indeterminate number of cases, the FBI engaged in even more overt intimidation to compel people to agree to ostensibly voluntary interviews. For instance, after the arrests of a Lodi, California, father and son on terrorism charges,\(^49\) the FBI aggressively sought information from other Pakistani Muslims: agents stationed their cars in front of homes, followed people for days, circled a mosque hosting a “know your rights” presentation where individuals they sought to interview had gathered, called individuals as many as ten times a day, and warned people that they would be “bad mouthed” at work if they did not cooperate.\(^50\) These measures conveyed a broader impression that the FBI would ratchet up pressure on those who declined an interview request.

At U.S. borders, by contrast, the compulsion is explicit: individuals cannot enter (or return to) the United States without satisfying border agents’ demands. Although U.S. citizens have an absolute right to enter the country,\(^51\) legally preventing the CBP from denying entry altogether to citizens who decline to answer questions,\(^52\) the CBP sometimes prolongs the detention of individuals who refuse to answer questions or subjects them to more intense searches as a result.\(^53\) Agents have not only used their power to delay admission to enforce

\(\text{\textsuperscript{48}}\) Carmen, supra note 44; CAIR CALIFORNIA, supra note 47, at 6.

\(\text{\textsuperscript{49}}\) Linda Goldston & Lisa Fernandez, FBI Expanding Terror Probe Tied to Lodi Father, Son, SAN JOSE MERCURY NEWS, June 10, 2005, at A1. The government initially suggested that others in Lodi, beyond those detained, might be linked to al Qaeda, but no other arrests followed. Many observers questioned the initial charges, including a retired FBI agent who sought to testify in defense of the accused. Shane & Bergman, supra note 38, at A1; Mark Arax, The Agent Who Might Have Saved Hamid Hayat, L.A. TIMES, May 28, 2006, at 116; John Simerman & Jessica Guynn, Arrests Illuminate Terror Probe, CONTRA COSTA TIMES, June 9, 2005, at A01.

\(\text{\textsuperscript{50}}\) Armstrong, supra note 25; Hoffman, supra note 25, at A01; Letter from ACLU of N. Cal. & Lawyers’ Comm. for Civil Rights of the S.F. Bay Area to FBI and Other Agencies Requesting Information Under Freedom of Information Act 2-3 (June 16, 2005), available at https://www.aclu.org/FilesPDFs/adu%20%20nc%20foia%20request%20for%20lodi.pdf (describing complaints of Muslim community members related to Lodi terrorism investigation).

\(\text{\textsuperscript{51}}\) See Nguyen v. INS, 533 U.S. 53, 67 (2001) (indicating that U.S. citizenship confers an “absolute right to enter [the nation’s] borders”).

\(\text{\textsuperscript{52}}\) In several cases, the U.S. government is alleged to have prevented U.S. citizens on the “no-fly list” from boarding flights returning to the United States, sometimes for extended periods, though it eventually permitted them to return. See Peter Finn, Detained Va. Teen Set to Return to U.S., WASH. POST, Jan. 21, 2011, at B01.

\(\text{\textsuperscript{53}}\) See ASIAN LAW CAUCUS, supra note 34, at 12 & n.4.
cooperation with questioning but have also, at times, engaged in more overt intimidation.

For instance, Zakariya Reed—a U.S. citizen, Muslim convert, and National Guard veteran—experienced several intimidating encounters at the U.S.-Canada border. Reed perceived one occasion as deliberately intimidating: a border agent asked Reed about a letter to the editor that Reed wrote, which was critical of U.S. support for Israel and the war in Iraq; in the same interview, another agent conspicuously removed and reloaded the clip of his gun in front of Reed. On a separate occasion, five CBP agents stopped and surrounded Reed's car, frisked him, and held him for several hours; during the interview, a CBP agent asked Reed why he had adopted a Muslim name and converted to Islam. Other interviews of U.S. Muslims have involved handcuffing or displays of physical force or statements that at “the border, . . . you have no rights.”

Protracted questioning at the border has often taken place in conjunction with detailed searches of travelers’ electronic media and reading materials. CBP officers conducted detailed searches of Muslims’ laptop computers, cell phones, and other electronic media, asking the travelers to identify family members appearing in pictures stored on a digital camera, questioning them about websites they visited, or examining websites individuals flagged as “favorites.” Border agents perused travelers’ books, lecture notes, and personal

54 See, e.g., Tabbaa v. Chertoff, 509 F.3d 89, 99-100 (2d Cir. 2007) (stating that U.S. citizens were threatened with continued detention unless they cooperated with CBP inspections); MUSLIM ADVOCATES, UNREASONABLE INTRUSIONS: INVESTIGATING THE POLITICS, FAITH, & FINANCES OF AMERICANS RETURNING HOME 21-22 (2009) [hereinafter MUSLIM ADVOCATES] (reporting that CBP told U.S. citizen reluctant to answer questions that the detention would end sooner if he complied).
56 Id.
57 Id.; MUSLIM ADVOCATES, supra note 54, at 28.
59 ASIAN LAW CAUCUS, supra note 34, at 11.
61 ASIAN LAW CAUCUS, supra note 34, at 16.
62 Id. at 19.
63 Id. at 34.
papers, and sometimes photocopied the documents or asked questions about the travelers’ views on the material.64

Beyond the coercion involved in FBI and CBP interviews, a second concern relates to the content of the questions asked. FBI and CBP officers questioned numerous Muslims, including U.S. citizens, about their religious and political beliefs and activities—subjects that U.S. citizens do not ordinarily expect government officers to probe. For example, CBP agents spent three hours questioning one Ph.D. student returning to the United States from a U.S. government-sponsored trip to Yemen on his local mosques, how long he had been Muslim, and the Islamic organizations in which he participated; when the student questioned the relevance of the religious inquiries, the agents told him that the detention “would end sooner if he simply answered the questions.”65 CBP agents asked other returning U.S. citizens about their views on foreign policies and politics,66 the mosques they attended,67 their charitable activities,68 membership in religious organizations,69 attendance at community events,70 participation in political demonstrations,71 support for lawful organizations,72 the religious sect to which they belonged,73 and prayer habits.74 Similar questions were asked in FBI interviews.75

Such questioning, even without coercion or intimidation, can convey powerful messages about the government’s respect for communities, neutrality towards religions, and overall

---

65 MUSLIM ADVOCATES, supra note 54, at 21-22.
66 Id. at 33, 40; ASIAN LAW CAUCUS, supra note 34, at 34; Jack Chang, Men Say Customs Bureau Asked About Faith, Politics, CONTRA COSTA TIMES, May 29, 2003, at A01.
67 MUSLIM ADVOCATES, supra note 54, at 20, 36, 38.
68 Id. at 20.
69 Id. at 22.
70 Id. at 30.
71 Id. at 39.
72 Id. at 30.
73 Matthai Chakko Kuruvila, Muslims Resent Customs Queries: Group Collects Complaints on Faith Questions, SAN JOSE MERCURY NEWS, May 29, 2003, at 1B.
74 MUSLIM ADVOCATES, supra note 54, at 34.
fairness. Unlike covert investigative methods such as electronic surveillance, an interview is a highly personal encounter between an individual and a law enforcement officer who embodies the full force of the law—the power to arrest and imprison, to detain and deport, or to exclude altogether from the country. In that encounter, even a relatively low-level officer represents the authority of the United States. Thus, the exchange that occurs in an interview signals the U.S. government’s beliefs as to what, or whom, it considers threatening.

When an FBI agent asks an Iraqi-American, selected without individualized suspicion, whether he practices Islam—following questions on knowledge of terrorism or weapons of mass destruction—it sends the message that the government considers the practice of Islam itself to be a threat. Similarly, when a CBP agent asks a U.S. resident at the border his views on the war in Iraq, it signals that the government considers one’s position on U.S. foreign policy relevant to his belonging in the United States. As a uniquely expressive investigative method, interviews carry a particular risk of conveying messages, intended or not.

A third concern relates to the basis for interviewee selection—specifically, the concern that either ethnic or religious profiling, or First Amendment activities, led to that selection. While some FBI and CBP interviews are occasioned by specific threat information or the inclusion of a person on a watch list, ethnic criteria or First Amendment activities supply the explicit basis for other interviews. For instance, the FBI openly relied on national origin in interviewing thousands of U.S. Arabs in the months after the September 11 attacks and in the period preceding the invasion of Iraq; the latter interview campaign included U.S. citizens. At other times, an ethnic basis for interview selection was unannounced but strongly indicated by the demographics of those interviewed.

---

76 See James, supra note 23.
77 MUSLIM ADVOCATES, supra note 54, at 41.
78 In fact, inclusion in the Customs and Border Protection watch list raises separate concerns related to the inadequate review process for watch list additions and the insufficient mechanisms for redress. See ASIAN LAW CAUCUS, supra note 34, at 33-40.
79 The government selected interviewees because of demographic and visa similarities to al Qaeda terrorists. See GAO, PROJECT TO INTERVIEW ALIENS, supra note 22, at 7-8.
80 See, e.g., Brune, supra note 23; James, supra note 20.
81 See e.g., Phillip O’Connor, Tactics with Somali Cabdrivers Stir Criticism of FBI, ST. LOUIS POST-DISPATCH, Feb. 3, 2011, at A1 (describing interviews of twenty-five to fifty Somali cabdrivers after the arrest of one for material support to terrorism).
On other occasions still, the FBI’s focus on particular groups resulted from its response to suspicious activity reports called in by the public, even where the reports clearly suggested ethnic or religious biases.\(^{82}\)

The CBP has also based targeting decisions explicitly on national origin, even for U.S. citizens. For instance, past CBP intelligence directives have called for particular scrutiny of naturalized U.S. citizens of Pakistani origin.\(^{83}\) In addition, CBP officers told some travelers that despite their U.S. citizenship, they were targeted because of where they were born;\(^{84}\) officers told others that even if they acquired U.S. citizenship, they would “always be a foreigner.”\(^{85}\)

Finally, for both FBI and CBP interviews, individuals’ First Amendment activities sometimes triggered the selection decision. Part II of this article elaborates on interviews based on First Amendment profiling, the harm such interviews present, and the line separating justifiable from unwarranted scrutiny.

Ultimately, the coercion, content, and selection criteria of interviews affect not just the rights and liberties of the Muslim community, but also potentially the very interest in security that is the professed goal of the interviews themselves. A growing body of literature suggests that for U.S. Muslims, as with other communities, perceptions of the fairness of law enforcement practices affect community members’ trust in, and willingness to assist, law enforcement.\(^{86}\) For instance, a recent study of New York Muslims by Tom Tyler, Stephen Schulhofer, and Aziz Huq found that perceptions of “procedural justice” involving U.S. counterterrorism policies—but not self-described religiosity, cultural differences, or political background—strongly correlate with individuals’ willingness to cooperate

---


\(^{83}\) Asian Law Caucus, supra note 34, at 29-31 (noting 2004 intelligence directive that called for greater scrutiny of naturalized citizens of Pakistani origin); see also Anne E. Kornblut & Spencer S. Hsu, U.S. Changing Way Air Travelers Screened, Wash. Post, Apr. 2, 2010, at A06.

\(^{84}\) Asian Law Caucus, supra note 34, at 25 (reporting interview in which CBP pulled aside a U.S. citizen because she was born in Pakistan).

\(^{85}\) Id. at 24.

\(^{86}\) See, e.g., Harris, supra note 16, at 132-41; Vera Institute Study, supra note 75, at 87, 94-95.
with antiterrorism policing.” As a visible and overt practice affecting U.S. Muslims, FBI and CBP interviews perceived as unfair may impose broader systemic costs in addition to burdening individual rights.

B. Interviews and Fourth Amendment Fictions

The natural place to begin an examination of the constitutionality of interviews might be the Fourth Amendment—the usual standard for measuring the lawfulness of law enforcement detentions. Two legal fictions, however, presumptively exclude interviews from Fourth Amendment protection. First, because individuals are not required to submit to an FBI interview, courts deem these interviews “voluntary.” Second, because border officials may question any traveler who seeks to enter the United States, courts consider CBP interviews “routine.”

As Tracey Maclin has argued in reference to the Justice Department’s initial post-9/11 interview campaign, most FBI interviews would not constitute seizures under the Supreme Court’s interpretation of the Fourth Amendment, despite the fact that those approached would have difficulty refusing the interview request. The Supreme Court has indicated that police questioning generally falls outside the scope of Fourth Amendment scrutiny because individuals in such encounters are free to terminate the questioning. Under current legal norms, only in the extraordinary case where police engage in “patently abusive and intimidating behavior” would a court find that an interview constitutes a seizure. According to Maclin, empirical evidence suggests that most people would not feel free to terminate ostensibly consensual police encounters because the average person interprets even a law enforcement officer’s polite request for cooperation as a legal command. Nonetheless, the constitutional standard for “voluntary”

---

87 Tom Tyler et al., Legitimacy and Deterrence Effects in Counterterrorism Policing: A Study of Muslim Americans, 44 Law & Soc’y Rev. 365, 368-69 (2010).
88 Maclin, supra note 18, at 493-502.
89 Id. at 494; see also David Cole, No Equal Justice: Race and Class in the American Criminal Justice System 16-20 (1999) (describing Supreme Court’s “reasonable person fiction” that an ordinary person would be able to reject police questioning).
90 Maclin, supra note 18, at 500-01.
91 Id. at 507.
questioning “was never intended to measure the reality of police-citizen encounters.”

In the CBP context, the separate fiction that questioning is “routine” scuttles ordinary Fourth Amendment protections. The Supreme Court has long proclaimed the government’s “paramount” authority to police the entry of people and objects across its borders and has declared that border searches are reasonable simply because they occur at the border. Thus, the Court has determined that the Fourth Amendment imposes no requirement of individualized suspicion for brief questioning on one’s immigration status at border checkpoints or for “routine” searches and seizures.

Applying this restrictive Fourth Amendment doctrine, two federal courts deemed the border detentions of Muslim U.S. citizens returning from abroad routine, despite the fact that the detentions were substantially longer, more intrusive, and more stigmatizing than ordinary CBP inspections of returning U.S. citizens. Thus, in Tabbaa v. Chertoff, the Second Circuit Court of Appeals declared that the questioning, pat-down searches, fingerprinting, photographing, and four- to six-hour detentions of Muslim U.S. citizens returning from Canada—without individualized suspicion—were “routine” even though CBP used screening measures “normally reserved for suspected terrorists.” The district court in Rahman v. Chertoff, a case involving U.S. citizens screened at the border because of mistaken association with a terrorist watch list, dismissed most of the challenged detentions as “routine” border stops, even where the stops included detentions of as long as six hours, handcuffing, or brief displays of physical force.

Thus, Fourth Amendment doctrine presumptively permits FBI and CBP interviews, even those that would not strike the average person as truly “voluntary” or “routine.”

---

52 Id.
54 Id.
57 509 F.3d 89, 92, 95, 98-99 (2d Cir. 2007).
C. Weak Internal Constraints on Interviews

Despite the documented historical use of interviews to suppress political activities, and current concerns over the practice, FBI and CBP internal guidelines impose few constraints on these interviews. In particular, existing guidelines do not limit the circumstances in which investigators can ask questions related to political and religious activities and do not provide effective constraints against selecting people for scrutiny on account of their First Amendment activities.

1. FBI Interviews

Guidelines for FBI investigations have grown progressively less stringent over time and now give FBI officers considerable discretion. The attorney general first issued internal guidelines for FBI domestic security investigations in 1976 in response to public outcry over abuses. The first guidelines required a factual predicate for all investigations and additional procedural requirements for interviews, including, in most cases, a requirement for supervisory approval. But successive versions of the attorney general’s guidelines loosened such constraints, culminating in the newest and weakest version issued by Attorney General Michael Mukasey in late 2008.

The Mukasey Guidelines include the most expansive definition yet of what FBI agents may legitimately investigate, permitting agents to conduct a form of investigation called “assessments” without any information or even allegation of a potential national-security threat. The Mukasey Guidelines

---


Office of the Atty Gen., Domestic Security Investigation Guidelines, reprinted in FBI Statutory Charter: Hearings Before the S. Comm. on the Judiciary, 96th Cong. 20-22 (1978). Under these 1976 “Levi” Guidelines, agents could use interviews in preliminary investigations (the least intrusive tier of investigation) only to gather certain public information or to identify the subject of an investigation; in limited investigations, the next tier, agents could use interviews for other purposes, but only with supervisory approval and after “full consideration of such factors as the seriousness of the allegation, the need for the interview, and the consequences of using the technique.” Id. at 20-21.

See Mukasey Guidelines, supra note 38; see also Jones, supra note 99, at 139-50.

authorize interviews in assessments as well as other methods “of relatively low intrusiveness,” and generally do not require supervisory approval for interviews. The Domestic Investigations and Operations Guide implementing the Mukasey Guidelines permits pretextual interviews, in which an agent fails to reveal an FBI affiliation or the true purpose of the information request. The Mukasey Guidelines do not permit information collection for the purpose of monitoring First Amendment–protected activity, and the Operations Guide states that assessments may not be based “solely” on the exercise of First Amendment rights. That standard, however, appears to permit an assessment conducted mostly based on First Amendment activity but also based on some additional, facially innocent fact—say, an agent’s decision to interview those who recently converted to Islam and serve in the U.S. armed forces.

Beyond the use of interviews for assessments and investigations, the Mukasey Guidelines appear to give the FBI broad authorization to conduct interviews for intelligence planning that goes beyond the investigation of specific cases. They seem to allow interviews that “develop overviews and analyses” of “present, emergent, and potential threats and vulnerabilities” and “their contexts and causes”—a standard that could conceivably justify interviews initiated to inquire into the religious or political “contexts and causes” of extremist threats. Notably, while guidelines for the original post-9/11 interviews of Arab noncitizens forbade inquiries into religious

---

103 MUKASEY GUIDELINES, supra note 38, at 17-18, 20.
104 Id. at 18; OPERATIONS GUIDE, supra note 102, at 63-64.
106 MUKASEY GUIDELINES, supra note 38, at 13. For more detailed FBI interpretation of First Amendment restrictions on its activities, see OPERATIONS GUIDE, supra note 102, at 24-30.
107 OPERATIONS GUIDE, supra note 102, at 44.
108 Moreover, the Guide makes clear that its definition of “First Amendment activities” does not extend to all activities that would be protected by the First Amendment, such as the advocacy of violence. Id.
109 See MUKASEY GUIDELINES, supra note 38, at 29 (permitting FBI to “draw on all lawful sources of information” in intelligence analysis).
110 Id.
beliefs or practices, the Mukasey Guidelines and Operations Guide do not.

The Mukasey Guidelines and Operations Guide include some restrictions to prevent coercing interviewees, although the press accounts described above suggest that the FBI does not consistently follow them. The Operations Guide states that information in interviews “must be voluntarily provided” and that agents should not “state or imply in any way” that “adverse consequences may follow if the interviewee does not provide the information.” In addition, the Mukasey Guidelines state that agents should stop questioning “immediately” if a person indicates a desire to consult a lawyer. Despite these limitations, the Operations Guide does not prohibit FBI agents from disregarding ambiguous or hesitant expressions of desire for legal counsel, even though individuals approached by the FBI may be too intimidated to state that desire definitively. Nor does the Guide disallow pressure tactics short of implying adverse consequences, such as insinuating that a reluctant interviewee “must have something to hide.”

The Mukasey Guidelines also advise that where different investigative methods are each “operationally sound and effective,” agents should use the “least intrusive method feasible,” but the Mukasey Guidelines contain other language to minimize the constraint this principle suggests. The Operations Guide recognizes that interviews with “employers, neighbors, and associates,” or those conducted at the workplace, are more intrusive than interviews in discrete locations. Despite this helpful distinction, the Operations Guide also advises that agents should primarily measure the degree of intrusion based on how much procedural protection established law and the Mukasey Guidelines themselves provide for the investigative method—thus designating interviews as a whole as a relatively nonintrusive choice. In addition, the Mukasey Guidelines give significant discretion to

---

111 GAO, PROJECT TO INTERVIEW ALIENS, supra note 22, at 9.
112 Although the publicly released version of the Operations Guide is redacted, one does not expect that the government would redact a restriction that protects individual rights.
113 OPERATIONS GUIDE, supra note 102, at 63.
114 Id.
116 OPERATIONS GUIDE, supra note 102, at 36.
117 Id. at 35.
agents in interpreting these rules, advising agents not to hesitate to use any lawful method, especially in terrorism investigations.118

2. CBP Interviews

Documents released by the CBP suggest a lack of significant constraints on questioning, but the agency has not publicly released sufficient information on its border inspection policies to fully judge the level of constraint that agents have in choosing whom to interview, for how long, or in what manner. A CBP training manual states that “routine questioning” at the border does not require reasonable suspicion.119 An immigration inspection manual released by the department, possibly outdated, states that “reasonable suspicion” is generally required to detain U.S. citizens for “extensive questioning,” but it appears to vitiate that requirement in the next breath by permitting agents to “continue inspecting for Customs purposes.”120

The CBP appears to have no written policy restricting the questioning of individuals about religious views, political activities, or other expression protected by the First Amendment.121 In fact, one high-level CBP official told community organizations that it was appropriate to question an individual about the mosque the individual attends.122 The agency does issue internal directives that may reflect First Amendment considerations; for instance, one CBP field office advised border agents not to apply special enforcement measures based “solely” on a person’s return “from a

118 MUKASEY GUIDELINES, supra note 38, at 12-13.
120 INSPECTOR’S FIELD MANUAL, supra note 119, at 18. The policy requires probable cause for detentions beyond an unspecified “reasonable period of time.” Id. These provisions, which appear in a 2006 edition released through a Freedom of Information Act request, may be outdated: the section on questioning U.S. citizens refers both to immigration inspectors as well as to “Customs,” while the CBP consolidated the functions of the U.S. Customs Service and Immigration and Naturalization Service in 2003. See KNOW BEFORE YOU GO, supra note 32, at 3. The Inspector’s Field Manual itself notes that the material is “gradually being updated” to reflect CBP policies. See INSPECTOR’S FIELD MANUAL, supra note 119, at 1.
121 ASIAN LAW CAUCUS, supra note 34, at 13-14.
122 Id. at 14-15.
pilgrimage to Mecca,” while also advising that “the large influx of travelers during this time period may be used as a cover by extremists and/or terrorists to enter the United States.”

Thus, existing evidence of internal agency regulations suggests insignificant constraints on the factual basis for initiating interviews in either the FBI or CBP context, and few meaningful constraints on interviews that bear on individuals’ political or religious expression. Neither the Fourth Amendment nor internal regulations offers real protection against the actual intrusion and stigma of FBI and CBP encounters.

II. FIRST AMENDMENT PROFILING

In the last part, this article argued that the coercion, content, and selection criteria behind law enforcement interviews present serious concerns that have largely been unaddressed. This article turns now to one set of interviews that raises particular concern: where individuals’ lawful acts of expression or association trigger a knock on the door or detention at the border, it sends a particularly strong message of exclusion to individuals and their communities and creates a chilling effect on expression. This part defines First Amendment profiling and discusses the profound stigmatic costs and chilling effects of the practice. This article then argues that while First Amendment profiling is sometimes justifiable, it is inappropriate both where the government deliberately seeks to suppress speech and where law enforcement investigations—even those with a legitimate purpose—swipe too broadly and consequently burden lawful speech and association.

When the FBI or CBP agent questions a person because he wrote a letter to the editor criticizing U.S. intervention in Afghanistan, worshipped at a particular mosque, or visited a religious website, they engage in what I call First Amendment profiling: the selection of a person for law enforcement attention because she has engaged in acts of expression or association of a political, religious, or cultural nature that would be protected by the First Amendment. CBP’s investigation of individuals returning to the United States
after attending an Islamic religious conference in Canada, litigated in the Second Circuit in Tabbaa v. Chertoff, provides a paradigmatic example of such profiling.

Murad Hussain described the detentions in Tabbaa as “cultural profiling,” which he defined as government targeting of “expressions of cultural identity” as proxies for “criminality, terrorist connections, or other subversive propensities,” noting that expressions of identity are often “significantly correlated” with membership in a racial, ethnic, or religious group. Others have called investigations based on First Amendment conduct “political profiling” or “First Amendment investigations.” Building on these conceptions, I consider First Amendment profiling to include investigative decisions triggered by expression or association of a cultural, religious, or political nature, whether “pure” speech or expressive conduct, that would be protected under the First Amendment.

Given that the First Amendment protects such a wide range of expression, with no consensus on the amendment’s core meaning, scholars have recognized the difficulty in delimiting the scope of expression that raises First Amendment concerns in the investigative context. Eugene Volokh argues that a broad interpretation of the First Amendment might lead one to the untenable conclusion that because speaking or sending an e-mail are constitutionally protected actions, we should interpret the First Amendment as limiting government “subpoenas demanding that people testify about what someone said or wrote.” My definition focuses on expression or association of a religious, cultural, or political nature, not as a normative statement of the outer limits of First Amendment protection, but because expression or association outside these areas frequently raises concerns of a different kind, and arguably degree. A person who triggers FBI scrutiny by sending an e-mail about purchasing a vast quantity of fertilizer may have engaged in a communicative act protected by the First Amendment (sending an e-mail), but the act holds no particular political, cultural, or religious meaning. A law enforcement investigation into that speech act does not

---

124 509 F.3d 89 (2d Cir. 2007). See supra notes 6-11 and accompanying text.
126 Fisher, supra note 16, at 625 (citing Chip Berlet & Abby Scher, Political Profiling: Police Spy on Peaceful Activities, AMNESTY NOW 20 (Spring 2003)).
127 See generally Rosenthal, supra note 16.
stigmatize particular views or manifestations of identity in the same way as investigations triggered by the political, cultural, or religious aspect of a communication.

While I limit my definition of First Amendment profiling to expression and association of a political, cultural, or religious nature, I expand it in two other respects. First, I consider First Amendment profiling to include not just an initial decision to target a person for investigation, but also any subsequent decision to prolong an investigation on account of First Amendment expression. For example, First Amendment profiling would include a law enforcement agent’s decision to broaden an investigation because of a person’s responses to questions about religious affiliations. It would not include, however, incidental questioning on religious or political beliefs, even if independently objectionable, where it does not trigger the interview or intensify law enforcement scrutiny.129

Second, I include within First Amendment profiling investigative decisions based “predominantly”—not just “solely”—on protected expression. For instance, it would include not just border agents’ decisions, as in Tabbaa, to question people solely because they attended an Islamic conference, but also a decision to question people because they had attended the conference and returned at night, or because they had attended the conference and were young men. The concerns raised by First Amendment profiling, as described in the next section, are not diminished in such cases.

A. The Impact of First Amendment Profiling

1. Stigmatic Harms

The most immediate, and perhaps most pervasive, harm of First Amendment profiling in law enforcement investigations, including interviews, is the imposition of stigma.130 U.S. Muslims have described interviews triggered by

---

129 Inquiries on religion and politics, even without First Amendment profiling, might give rise to an independent challenge based on the Supreme Court’s compelled disclosure cases, especially in compulsory border interviews. For some discussion of these cases, see infra notes 194-207 and accompanying text.

130 A large volume of literature, since the publication of Erving Goffman’s seminal account, has attempted to define and conceptualize stigma. ERVING GOFFMAN, STIGMA: NOTES ON THE MANAGEMENT OF SPOILED IDENTITY (1963). I use the term “stigma” in both the sense defined by Goffman (an “attribute that is deeply discrediting” that reduces the individual “from a whole and usual person to a tainted, discounted one”) and according to Bruce Link and Jo Phelan’s more recent
their religious or political activities as branding them, and their communities, as disloyal or suspicious—as outsiders excluded from “belonging” to the nation.

“We weren’t treated as American citizens. We were treated as suspects,” recounted one Muslim college student whom CBP detained for attending the conference referenced in Tabbaa and then ordered to “stand face-first against the wall,” his legs apart, to be searched for weapons. Such encounters affect not just the individuals questioned, but their broader ethnic or religious community’s sense of belonging in the United States. An Oregon Muslim community leader questioned by CBP agents about why he made a religious pilgrimage to Mecca said that Muslims had grown accustomed to being “pariahs in their own country.”

The harms associated with First Amendment profiling mirror those arising from explicit racial or religious profiling. Where a form of expression is strongly linked to one’s ethnicity, national origin, or religion, government selection of individuals for special scrutiny on account of their expression will “feel” the same as targeting members of that racial or religious group directly. Certainly the Muslim Americans detained in Tabbaa did not perceive their detentions to be less stigmatizing because the trigger was membership at an Islamic conference—not their religion per se—or because CBP might theoretically have stopped any non-Muslims who said they had attended the conference. As the correlation between an expressive practice and membership in a particular racial or religious group approaches 100 percent, the technical distinction between the two collapses altogether: the questioning of seven Muslim men for praying in a convenience store conceptualization of stigma as the co-occurrence of components including: the labeling of a particular human difference, the linkage of that difference with stereotypes, the use of that difference to separate “us” from “them,” and the resulting loss of status and discrimination in a context of unequal power. See id. at 3; Bruce G. Link & J o. C. Phelan, Conceptualizing Stigma, 27 ANN. REV. Soc. 363, 367 (2001).

One might question why an Equal Protection or Free Exercise Clause challenge is not available to challenge these measures. See Hussain, supra note 16, at 944-52 (arguing that where the government selects people for scrutiny based not on their membership in a protected group, but based on behavior that largely correlates with it, the requirement that plaintiffs prove discriminatory intent will impede challenges under both the Equal Protection Clause and under prevailing Free Exercise Clause interpretations).
store parking lot illustrates an instance that is at once religious and First Amendment profiling.\textsuperscript{134}

At the border, questioning and extensive searches of returning U.S. citizens particularly convey a message of exclusion since CBP agents effectively control the terms by which a person can return home. A Sacramento Muslim and naturalized citizen recounted that his experiences at the border—including repeated screenings, questioning on his political views, and searches of websites he visited—made him feel "unwelcome" in his own country.\textsuperscript{135} He said, "I never experienced such a feeling at any international airport in the world, including Third World countries. But I have this feeling when I come home."\textsuperscript{136}

Furthermore, First Amendment profiling sends a message to the nation as a whole, not just affected communities, that Muslims are unequal. As Murad Hussain has argued, law enforcement scrutiny of Muslim Americans' expressive activities labels Muslims "presumptively disloyal and unworthy of empathy" to the polity at large, facilitating hate crimes and private discrimination and stymieing the community's efforts to use civic engagement to achieve social equality.\textsuperscript{137} Indeed, a growing social science and legal literature points to the tangible costs of stigma for individuals and communities.\textsuperscript{138} Not only do stigmatized groups lose self-respect and tend to internalize "at least part of the version of their identities imposed by the stigma," but society correspondingly "acts toward the stigmatized person on the basis of the stigma," leading to a deprivation of material goods from economic opportunities to political representation.\textsuperscript{139} While stigma is

\textsuperscript{134} See Ken Ritter, Muslim Group Says FBI Still on Nevada Prayer Case, KOLOTV.COM (June 21, 2010), http://www.kolotv.com/southernnevadanews/headlines/96852029.html. When one man questioned why their prayers had elicited suspicion, the police officer replied, "I don't know if you're... saying, 'I hope that I kill a police officer today.'... We just want to make sure that you guys are good people." CAIRtv, Video: CAIR Concerned About FBI Questioning of "Henderson 7," YOUTUBE (June 23, 2010), http://www.youtube.com/watch?v=lxGNN3U6Da&feature=player_embedded.

\textsuperscript{135} A SIAN LAW CAUCUS, supra note 34, at 34.

\textsuperscript{136} Id.; see also Kuruvila, supra note 73 (quoting a U.S. citizen describing his experience of CBP questioning as "trying to instill in us a feeling that we don't belong here").

\textsuperscript{137} Hussain, supra note 16, at 938-41.


\textsuperscript{139} KENNETH L. KARST, BELONGING TO AMERICA: EQUAL CITIZENSHIP AND THE CONSTITUTION 26-27 (1989). For recent studies suggesting links between stigma and the physical health of U.S. Arab communities, see Diane S. Lauderdale, Birth Outcomes for Arabic-Named Women in California Before and After September 11, 43 DEMOGRAPHY 185 (2006); Aasim I. Padela & Michele Heisler, The Association of
often seen as a result of equal protection violations, it results no less from First Amendment profiling.

2. Chilling Effects on Expression

Perhaps the most common harm legal scholars posit as resulting from law enforcement investigations into political and religious expression is the chilling impact on such expression. Despite the fact that scholars sometimes favorably contrast interviews to clandestine surveillance, the overt nature of interviews actually makes them more likely to directly and immediately influence behavior than covert investigative methods.

On some occasions, the chilling effect may be deliberate. Law enforcement agencies deliberately used interviews in the past to suppress lawful political activities by sowing mistrust within movements. The extensive congressional investigation of U.S. intelligence operations in the mid-1970s concluded that the FBI’s fifteen-year Cointelpro programs, aimed at civil rights activists and others on the Left, deliberately used interviews to disrupt political activities. An infamous FBI memorandum from 1970 advised that interviews could “enhance the paranoia in these circles” and convey the impression that “there is an FBI agent behind every mailbox.”

Even where law enforcement agencies do not deliberately use interviews to suppress expression, they acknowledge using them to send a message that government agents are watching. The Department of Justice explained its post-9/11 interviews of several thousand Arab immigrants as an attempt to “sow disruption among potential terrorists,” and claimed that the interviews “ensured that potential terrorists sheltering themselves within our communities were aware that

Perceived Abuse and Discrimination After September 11, 2001, with Psychological Distress, Level of Happiness, and Health Status Among Arab Americans, 100 AM. J. PUB. HEALTH 284 (2010).


Fisher, supra note 16, at 673.


law enforcement was on the job in their neighborhoods.” The FBI has also explained other interview programs—including interviews of potential anarchist protestors before the 2004 national political conventions and surveillance of Muslims before that year’s presidential election—as efforts to deter acts of violence through obvious surveillance. While using an investigative technique to dissuade people from committing violence is not problematic in principle, it is unclear that targets of such practices can distinguish that permissible message from others potentially received (e.g., “do not protest” or “do not go to the mosque”).

Thus, when FBI agents interviewed sixty Muslims in Flint, Michigan, regarding their donations to Muslim American charities, many donors interpreted the investigation as intimidation aimed at chilling contributions to Muslim charities. FBI agents visited a number of donors two years in a row, both times on the “eve of Ramadan,” the Muslim holy month in which many individuals choose to give religiously ordained charitable contributions. During these visits, two agents interviewed the donors at their workplaces, while two others simultaneously interviewed their spouses at home. The inquiries, and particularly the return visits, convinced the donors that the government sought to intimidate them into not supporting lawful Muslim charities, with the implied message that “[i]f you keep giving, we’ll keep coming back at you.” Whether or not the government intended that message, the fact that law enforcement agencies continue to acknowledge using

---

144 Final Report on Interview Project, supra note 43, at 1, 7.
147 See Frederick Schauer, Fear, Risk and the First Amendment: Unraveling the “Chilling Effect,” 58 B.U. L. REV. 685, 690 (1978) (distinguishing “benign” deterrence resulting from “intentional regulation of speech or other activity properly subject to governmental control” from “invidious” chilling of activities protected by the First Amendment).
149 Id. at 70.
150 Id.
151 Id. at 71.
interviews to send a message underscores the point that
interviews can serve as an intervention, not just an information-
gathering measure.

Indeed, there is powerful anecdotal evidence along with
additional support from ethnographic and empirical studies of
chilling effects on expression in the Muslim community. These
accounts indicate that government investigative practices,
including questioning, have led some U.S. Muslims to avoid
attending political demonstrations or gatherings, restrain from
donating to political causes or religious charities, avoid speaking out against U.S. foreign policies or express political
opinions, hesitate to join or participate in mosques or community organizations, remove their names from group
membership lists, modify their use of the Internet, stop purchasing political books abroad, refuse to speak publicly
about law enforcement practices, and avoid names or clothing
that express their religious or cultural identities. For instance, a Muslim community leader asserted that the FBI
questioned nearly every donor to one Southern California
mosque, leading to a pronounced decline in donations.

152 Vera Institute Study, supra note 75, at 66; June Han, "We Are

153 An extensive ACLU report documented FBI and CBP interviews of donors
to Muslim charities and presented statements of numerous community members who
had stopped giving money in response to questioning of themselves, friends, or family. Blocking Faith, supra note 148, at 97-100; see also Han, supra note 152, at 14; Vera Institute Study, supra note 75, at 66.

154 Teresa Watanabe & Paloma Esquivel, Muslims Say FBI Spying Is Causing Anxiety: Use of an Informant in Orange County Leads Some to Shun Mosques, L.A. Times, Mar. 1, 2009, at 1; Vera Institute Study, supra note 75, at 58; Matthew Rothschild, FBI Talks to Muslim High School Student About "PLO" Initials on His Binder, Progressive (Dec. 23, 2005), http://www.progressive.org/mag_mc122305 (reporting that high school student became hesitant to express his political views after FBI interview apparently triggered by his writing the initials "PLO" on a binder).

155 Tyler et al., supra note 87, at 396; Watanabe & Esquivel, supra note 154.

156 Vera Institute Study, supra note 75, at 66 (reporting statement of
community organization that membership declined because people called to remove
names from database).

157 Sidhu, supra note 16, at 391.

158 Asian Law Caucus, supra note 34, at 18, 19 (citing examples of
individuals subject to repeated lengthy CBP interviews and searches who no longer
purchase books abroad).

159 Carmen, supra note 44.

160 Tyler et al., supra note 87, at 396; Brian Haynes, Extra Scrutiny Chafes Muslims, Las Vegas Rev.-J., Oct. 4, 2004, at 1A (describing the decision of one Muslim
convert, who was followed in an airport while reading the Quran, not to adopt a
Muslim name for fear of harassment).

161 Watanabe & Esquivel, supra note 154.
These claims are largely anecdotal, and one might question the extent to which individuals and communities have actually ceased to engage in expression or association. Some individuals subjected to First Amendment profiling say they would continue the activities that triggered the scrutiny. In fact, ethnographic studies show that while post-9/11 scrutiny of Muslims led some people to withdraw from activities that would identify them as Muslim, Arab, or South Asian, others became more engaged in civic and political life in an effort to dispel stereotypes and resist unfair treatment. Communities are not monolithic; they can simultaneously exhibit chilling of expression and signs of resistance. Both of these responses, however, stem from the stigma experienced, and thus even the engagement response should not dispel concern.

In two quantitative studies, members of the U.S. Muslim community, but not a majority, reported changing their behavior in response to government scrutiny of their community. In the study of New York Muslims described above, one in five surveyed reported altering behavior in response to general law enforcement scrutiny of Muslims, including changes in attendance at group prayers in a mosque (20 percent of respondents), manner of dress (22 percent), everyday activities (17 percent), and travel behavior (26 percent). A 2007 study of Muslim Americans found more modest changes: while almost three-quarters of Muslims surveyed believed that the government was monitoring the general activities and Internet usage of Muslims, only 11.6 percent of respondents reported changing their general activities due to that concern, and 8.4 percent reported changing their Internet usage.

These numbers are not negligible: assuming a population of two million Muslims in the United States, even 10 percent of U.S. Muslims represents two hundred thousand people who cease to engage in lawful expressive behavior for fear of government scrutiny. In addition, the proportion of individuals who report changing their behavior might be greater in particular subsets of the Muslim community: immigrants, working class community

---

162 Tabbaa v. Chertoff, 509 F.3d 89, 102 (2d Cir. 2007).
164 Tyler et al., supra note 87, at 396.
165 Sidhu, supra note 16, at 390-91.
members, or those who directly experience law enforcement scrutiny might feel particularly vulnerable.

Moreover, studies that measure the extent to which individuals report changing their behavior may underestimate the extent of more subtle, but potentially more pervasive, changes in behavior. A practicing Muslim, for instance, might not stop going to the mosque, but might hesitate to speak freely with other worshippers. Other questions (not asked in these studies) might elicit evidence of subtler changes: “Do you feel that you must reassure co-workers about your loyalty before participating in a casual workplace conversation about events in the Middle East?” “Do you hesitate to speak in Arabic or perform ritual prayers in public for fear of drawing suspicion?”

Focusing on the hard impacts of surveillance (the decision to avoid a political gathering) misses the soft impacts that may affect communities in equal or greater fashion (the diminished sense of trust in others and openness in communication). Several scholars have argued that the surveillance of mosques intrudes on the sense of security that worshippers seek at a place of worship, and surveillance of any community-based or political organization diminishes trust within the group, with longer-term effects on the quality of association. Aziz Huq argues further that law enforcement’s use of religious speech to signal high-risk terrorist threats interferes with religious communities’ “epistemic autonomy”—their “collective interest . . . in determining the content and direction of . . . religious beliefs without interference by the government.” Thus, any assessment of chilling effects must take into account the cumulative impact of subtle, smaller changes and the qualitative aspects of changed behavior.

B. Suppression and Overbreadth in First Amendment Profiling

Accepting that First Amendment profiling, at least in some cases, imposes significant stigmatic costs and chilling

166 Neither study separated out the impact on individuals who had personally faced law enforcement scrutiny from that of the Muslim community at large.
167 Lininger, supra note 16, at 1233-36; Harris, supra note 16, at 166-68; Fisher, supra note 16, at 653. See, e.g., Dennis Wagner, FBI’s Queries of Muslims Spurs Anxiety, ARIZONA REPUBLIC (Oct. 11, 2004, 12:00 AM), http://arizona.indymedia.org/news/2004/10/22062.php (reporting Muslim community leader’s statement that FBI interviews led to “creeping distrust,” leading people to question whether acquaintances might report their words to the government); see also CAINKAR, supra note 20, at 185-86.
effects does not tell us whether the practice in a given instance is justified. Not every law enforcement practice that qualifies as First Amendment profiling is wrong. For instance, few would argue that law enforcement should not scrutinize an influential religious leader advocating terrorist attacks in the United States or a person who has joined al Qaeda, at least in order to determine whether the person is actively recruiting or assisting terrorists. The permissible scope of government investigation exceeds the permissible scope of government criminalization: the government may investigate, in an effort to uncover potential violations of the law or threats to national security, activities that it cannot constitutionally ban. The government could not, consistent with the First Amendment, outlaw the advocacy of terrorism or a person’s membership in a terrorist organization unless those activities cross certain lines—e.g., the advocacy is directed at producing “imminent lawless action and is likely to” have such an effect, the member has a specific intent to further the organization’s illegal aims, or either activity constitutes material support to a designated foreign terrorist organization. Such activities, even without crossing these lines, may be so relevant to uncovering actual violations of the law that the government may legitimately investigate them.

These are relatively clear examples; even civil libertarians might agree that a close enough nexus exists between the protected activity and an actual threat in order to permit First Amendment profiling. On these narrow facts, a three-step justification for First Amendment profiling articulated by scholars and law enforcement officials seems least objectionable. First, because of the potential of grave harm from terrorism, investigators must identify terrorists early and prevent acts of violence in advance of the commission of any criminal act. Second, these scholars and officials argue, because terrorists are elusive and hard to find, investigators need to cast their net broadly to identify those who pose a

173 RICHARD POSNER, NOT A SUICIDE PACT: THE CONSTITUTION IN A TIME OF NATIONAL EMERGENCY 114 (2006); Huq, supra note 16, at 839-42.
threat. Third, they argue that where the threat stems from ideologically motivated violence, evidence of that ideology (whether religious or political) is a relevant and rational basis for suspicion.

Of course, in most cases, the nexus between ideology and threat will not be as clear. At that point, the justifications for First Amendment profiling encounter serious questions. While terrorists may act out of religious or political motivation, the use of individuals’ religious or political views or behavior to “predict” threats is fraught with peril. Claims about the relevance of ideological linkages to violence may appeal to intuition rather than actual evidence. In fact, research suggests that the relationship between religion and terrorism is complicated and contested: for instance, while a New York Police Department report cites a turn to religion as a risk factor for “radicalization” and identifies mosques, student associations, and nongovernmental organizations as examples of hubs where radicalizing Muslims might gather, one academic study suggests that increased religious education and greater immersion in Muslim social institutions diminishes the religious naïveté and social isolation that can feed extremism. Even assuming that the “experts” could arrive at an accurate and sophisticated understanding of the linkages between terrorism and ideology, there is good reason to question how well law enforcement agencies could apply that information.

Beyond these reasons to doubt the effectiveness of First Amendment profiling, the primary objection rests on a broader

---

174 POSNER, supra note 173, at 114.
175 Id. at 116; see also OPERATIONS GUIDE, supra note 102, at 27-28 (providing examples of where it would be “rational and permissible” for the FBI to consider religious adherence, affiliations, or practices in an investigation).
principle: where a large number of innocent people engage in religious or political behavior alleged to be linked to an actual threat, law enforcement should not define threats so broadly as to sweep in substantial amounts of protected expression. If, at one end of the spectrum, an influential religious leader’s advocacy of terrorist attacks in the United States or a person’s membership in al Qaeda justifies law enforcement attention, law enforcement scrutiny directed at those who don Islamic garb or become active in social causes—behaviors identified by the New York Police Department report as “signatures” of the “second stage of radicalization”—is just as clearly unjustified.

So when is First Amendment profiling actually wrong? I argue that there are two situations where the government oversteps its bounds. First, it is wrong for the government to target people for investigation on account of lawful expression or association for the purpose of suppressing protected expression or association (what I label the suppression concern). Second, First Amendment profiling is wrong to the extent that the investigation imposes too great a burden on protected expression in relation to the government interest at hand (the overbreadth concern). Thus, First Amendment profiling directed at the suppression of protected speech is presumptively wrong; First Amendment profiling not involving deliberate suppression is suspect but not categorically wrong.

1. The Suppression Concern

The principle behind the suppression concern is that if the government cannot criminalize expression, it cannot use indirect methods of coercion to achieve the same ends. This principle applies whether the government’s ultimate goal is illegitimate (suppression of dissent or religious views out of hostility or self-interest) or legitimate (the prevention of violence), so long as the immediate objective is suppression of lawful speech or association. Although one might distinguish, at a theoretical level, between a purpose to suppress speech based on hostility to ideas and a purpose to suppress speech to prevent violence, the distinction collapses in the national-

179 Silber & Bhatt, supra note 176, at 30-31; see also Huq, supra note 177, at 57.

180 In using the term “overbreadth,” I am not referring to the separate First Amendment standing doctrine that allows a party to challenge the constitutionality of a statute on the grounds that it impermissibly prohibits the speech of others, even if the party’s own speech could constitutionally be prohibited. See generally Note, Overbreadth and Listeners’ Rights, 123 Harv. L. Rev. 1749 (2010).
security context, where hostility to ideas almost always takes the form of a belief that particular ideas are dangerous and likely to produce harm.\textsuperscript{181} Thus, the suppression concern includes the scenario where a government official who believes that a particular religious movement inspires some individuals to embrace violence harasses adherents to prevent them from worshipping together—even if the official’s ultimate purpose is to forestall acts of violence.

Though investigations today may raise the suppression concern, there is a relatively broad consensus that the purposeful use of investigations to suppress lawful expression is wrong. For instance, the FBI officially proscribes investigations of First Amendment activities solely for the purpose of monitoring or abridging lawful expression.\textsuperscript{182} That principle stems from wide condemnation of the agency’s historical abuses: a congressional investigation concluded that from 1956 to 1971, the FBI had “conducted a sophisticated vigilante operation aimed squarely at preventing the exercise of First Amendment rights of speech and association,”\textsuperscript{183} including the use of wiretaps and bugs to surveil and discredit Dr. Martin Luther King.\textsuperscript{184} Despite near consensus today that investigations should not aim to disrupt lawful speech, that consensus may shift as federal agencies develop a counter-radicalization strategy to prevent and disrupt Islamic extremism within the United States, leading to renewed debate over government-sponsored efforts to suppress speech.\textsuperscript{185}

2. The Overbreadth Concern

Second, though less obvious, First Amendment profiling raises the concern that regardless of motive—even assuming that law enforcement agents are acting for the legitimate purpose of investigating a terrorist threat—the government

\textsuperscript{181} Thus, in this context, the distinction Elena Kagan draws between “ideological” and “harm-based” motives for government actions collapses—as Kagan anticipated in suggesting that in certain cases, “the two kinds of motives become hopelessly entangled . . . in a kind of endless feedback loop.” Elena Kagan, Private Speech, Public Purpose: The Role of Governmental Motive in First Amendment Doctrine, 63 U. Chi. L. Rev. 413, 433-35 (1996).

\textsuperscript{182} Mukasey Guidelines, supra note 38, at 13.

\textsuperscript{183} Church Committee Report, supra note 142, at 3.

\textsuperscript{184} See id. at 81-86. For other examples of FBI investigations focused on First Amendment expression, see generally Coles & Dempsey, supra note 169.

may have defined threats too broadly, overclassifying innocent expression and association to improperly create a legitimate basis for scrutiny. There are two reasons why law enforcement investigations with a legitimate purpose should raise First Amendment concerns. First, even where the purpose of an investigation is not the suppression of expression, hostility to that expression may well influence the scope and shape of the investigation. In Tabbaa, for example, CBP officials ordered intrusive screening measures on Muslims returning from a conference as a response to intelligence concerning the presence of suspected terrorists at the event. But unspoken assumptions that an Islamic conference was likely to be suspicious might well have colored the judgment that it was rational—and fair—to stop all people returning from a diverse gathering of thirteen thousand people based on the possibility that they may have met a terrorist at the event. The unpopular and subordinate status of the communities affected may color threat perceptions, reduce empathy for innocent people affected by an investigation, or diminish fears of political backlash. For instance, the knowledge that political leaders and the U.S. public would not raise howls of protest at the singling out of participants at an Islamic conference might well have lessened the perceived costs of a decision to target them.

The second reason for the overbreadth concern is that even where hostility to speech or speakers does not influence an investigation, First Amendment profiling is likely to result in particularly grave harms. Where an investigative decision on its face turns on First Amendment expression or association, it is particularly likely to result in stigmatic harm and chilled

---

186 See Kagan, supra note 181, at 435 (arguing that hostility to speech may lead the government to overestimate the harm the speech may cause).

187 One commentator writes of Tabbaa that questioning attendees at an Islamic conference based on the presence of a suspected terrorist is no more “profiling” than questioning attendees at a State Fair for similar reasons, because it does not rely on stereotyping but on “suspect description.” R. Richard Banks, Group Harms in Antiterrorism Efforts: A Pervasive Problem with No Simple Solution, 117 YALE L.J. POCKET PART 198, 199 (2008), http://yalelawjournal.org/the-yale-law-journal-pocket-part/civil-rights/group-harms-in-antiterrorism-efforts-a-pervasive-problem-with-no-simple-solution/. This conclusion overlooks the likelihood of group-based assumptions behind investigators’ decisions: it is hard to imagine that CBP would have detained, fingerprinted, and photographed everyone returning from a state fair attended by thirteen thousand people rather than choosing more tailored means to identify those who posed a threat.

188 See Paul Brest, Foreword: In Defense of the Antidiscrimination Principle, 90 HARV. L. REV. 1, 7-8 (1976) (arguing that government decisions not based on racial animus may still be affected by “racially selective sympathy and indifference” to other racial groups); see also Hussain, supra note 16, at 945 (discussing possibility of unconscious and cognitive biases in investigative decisions).
expression. Consider a hypothetical in which the FBI, responding to the attempted Times Square bomber Faisal Shahzad’s claims that U.S. drone attacks in Pakistan motivated him, began questioning hundreds of people who signed a petition protesting civilian deaths from drone strikes. The FBI might justify the interviews as an intelligence assessment to measure potential threats presented by Pakistani Americans opposed to U.S. policies. While such an inquiry might have a “legitimate” purpose—at least one that the current attorney general guidelines deems proper—the means employed would be well out of proportion to the declared objective and seriously likely to discourage others from voicing similar opposition to U.S. policies. Even where hostility to the ideas expressed does not influence an investigation, the government’s use of First Amendment expression as a basis for selecting interviewees can result in grave harm to those questioned and their broader communities.

In recent years, some federal agencies have themselves concluded that certain investigative or intelligence activities they conducted were overbroad. For instance, the Department of Homeland Security distributed and then rescinded intelligence on certain First Amendment activities of U.S. Muslims, deeming it a violation of department policy. A Department of Justice Inspector General review of the FBI’s investigations of several domestic advocacy groups concluded that while the FBI did not deliberately target groups on account of their First Amendment activities, it had weak factual support for opening or continuing certain investigations.

The overbreadth concern also draws support from history: David Cole and James Dempsey have argued, for instance, that the FBI investigation in the 1980s of the Committee in Support for the People of El Salvador, a peaceful political organization, continued and expanded despite a slim factual basis for concern.

---

191 OVERSIGHT & REVIEW DIV., U.S. DEP’T OF JUSTICE OFFICE OF THE INSPECTOR GEN., A REVIEW OF THE FBI’S INVESTIGATION OF CERTAIN DOMESTIC ADVOCACY GROUPS 186-87 (2010). But see A REVIEW OF THE FBI’S INVESTIGATIVE ACTIVITIES CONCERNING POTENTIAL PROTESTORS, supra note 145, at 3 (finding that the FBI did not improperly target potential protestors at national political conventions to chill First Amendment activities).
192 COLE & DEMPSEY, supra note 169, at 22-23.
The FBI later admitted that the investigation, which generated files on 2376 individuals and 1330 groups, was improperly focused on First Amendment activities.\footnote{Id. at 24, 33.}

The most difficult challenge, of course, is making a coherent distinction between permissible First Amendment profiling and impermissible overbreadth, especially in interviews. Indeed, the overbreadth concern, unlike the suppression concern, does not suggest any bright line. Rather, overbreadth suggests a contextual, case-by-case determination of whether there is a tight enough nexus between the targeted expression and an actual threat of violence to justify the particular burden on First Amendment rights.

Two examples of actual interviews involving First Amendment profiling might help illustrate relevant considerations. First, as this article has argued, the First Amendment profiling in Tabbaa was overbroad, and the case was wrongly decided. There, the concern was the possibility of individuals using the mainstream gathering as a cover for meeting terrorist suspects. The size of the event (thirteen thousand participants), however, made any individual participant’s probability of meeting suspects fairly small. In addition, the security measures taken, including interviews, lengthy detentions, fingerprinting, and the like were particularly stigmatizing and intrusive. Moreover, surveillance of specific suspects at the conference itself, which presumably could have been arranged through the cooperation of the Canadian government, suggested a ready alternative.\footnote{The Tabbaa court rejected this alternative on the grounds that “the U.S. government cannot freely conduct surveillance...in Canada,” without any explanation of why U.S. authorities could not coordinate with Canadian law enforcement to arrange any required surveillance. Tabbaa v. Chertoff, 509 F.3d 89, 104 (2d Cir. 2007).} Given these facts, the nexus between the alleged security threat and the First Amendment trigger for the selection of interviewees—attendance at the conference—was too attenuated to justify the particularly burdensome measures taken.

Second, a more difficult case might be the FBI’s numerous interviews of students who attended a sixteen-day Houston Islamic conference in 2008 that Umar Farouk Abdulmutallab, the Nigerian man who attempted to blow up an
airliner in December 2009, also attended. The conference was sponsored by an Islamic religious institute known to teach a particularly conservative brand of Islam, though it did not advocate violence. Following the attempted bombing, the FBI initially sought to interview all 156 participants and indeed interviewed an unknown (but apparently large) number. Perhaps the FBI reasoned that other students who attended could provide information on Abdulmutallab, that a teacher at the conference might have influenced both Abdulmutallab and other attendees, or that the conference might have brought together like-minded individuals. How should we evaluate this case, which at first glance bears some resemblance to Tabbaa?

Several considerations inform this inquiry. One consideration involves the burden imposed: the FBI interviews likely did not create the same direct burden as the compulsory and extensive screening measures taken in Tabbaa. On the other hand, especially if the FBI approached all or nearly all the 156 participants, these interviews may well have chilled those who would otherwise wish to attend the institute's events. A second consideration is the strength of the nexus between the potential terrorist threat and the First Amendment trigger. Here, there had been an actual terrorist attempt, and compared to Tabbaa, the relatively small number of participants and the duration of the conference suggests a greater probability of other students having met Abdulmutallab or some other common source of influence. Also relevant is that a “tiny fraction” of the institute's other students (not those subject to the interviews) had turned to violence. A third consideration is the scope of the questioning: to the extent it focused on Abdulmutallab, it would be much less problematic than broad-based inquiries that supposed that the interviewees at large presented a threat. Ultimately, whether these interviews were overbroad might come down to the strength of the FBI's preexisting evidence that the institute attracted students drawn to violence and the scale and content of the interviews conducted.

These examples suggest the complexity of the fact-specific inquiry required for determining overbreadth. Though

---

196 Id.
197 Id.
198 Id.
deciding when First Amendment profiling is permissible may not be easy, the real stigma and chilling effects of the practice, combined with the historical record of abuses, make that inquiry critical. In the next part, I argue that courts should undertake that inquiry through heightened scrutiny of First Amendment profiling, and that existing First Amendment doctrine provides a foundation for such review.

III. FIRST AMENDMENT PROFILING IN THE COURTS

Given the dual concerns of suppression and overbreadth raised by First Amendment profiling, courts ought to subject investigations involving First Amendment profiling, including interviews, to heightened scrutiny whenever plaintiffs establish that they have faced a substantial burden on their First Amendment rights to freedom of speech or association. This article proposes a two-part test for such cases. A court should inquire first whether the government had a compelling interest in conducting the practice in question. In that step, the government will usually be able to prevail, citing national security, except in the rare case where direct evidence indicates a purpose to suppress speech or some other illegitimate reason for an investigation—in which case, the government would lose without further balancing. Assuming that the government establishes a compelling interest, the court should then query whether the means employed were “narrowly tailored” to serving the declared objective. The availability of less restrictive means for the government to resolve security concerns ought to create a rebuttable presumption that the investigation was overbroad. This heightened scrutiny would be neither traditional strict scrutiny weighted heavily against constitutionality nor a deferential balancing analysis in which the government always wins.

While suppression and overbreadth are conceptually distinct, the test here would seek to address both concerns. The existence of a suppression purpose on the facts of any particular case can often be detected only from a demonstration of

199 I focus in this article on the First Amendment rights to freedom of speech and association, which cover religious speech and association, rather than the separate religion clauses of the First Amendment.

200 A point of comparison might be the approach in freedom of association claims, where courts query whether state interests may be achieved by other means “significantly less restrictive of associational freedoms,” Roberts v. U.S. Jaycees, 468 U.S. 609, 622 (1984), not whether the means employed were the least restrictive alternative.
overbreadth. When government agents conduct an investigation in order to suppress protected speech, the invidious purpose is rarely apparent; law enforcement will almost always provide some facially legitimate reason for an improper investigation. Thus, the most plausible route to “uncover” an actual suppression purpose is through demonstrating such a remote connection between the investigated activity and an actual threat, or a burden on speech so disproportionate to the declared objective as to support the inference that no legitimate investigative purpose existed.201

I do not argue that courts, as opposed to Congress or administrative agencies, are necessarily best positioned to resolve First Amendment concerns regarding law enforcement investigations. Others have extensively debated the effectiveness and desirability of judicial enforcement of civil rights claims.202 But political process theory suggests that where terrorism investigations predominantly affect communities that are subject to widespread public hostility, Congress and executive agencies may lack the political will to fully resolve concerns stemming from such investigations.203 While executive oversight institutions, such as the Department of Justice Inspector General, have triggered impressive reform of certain investigative practices, even strong Inspector General reviews may not substantially constrain agency discretion to prevent future abuse, and these institutions cannot define substantive rights that would bind agency conduct.204 Moreover, even if agency policies and guidelines are more likely to directly influence law enforcement behavior than court decisions, court decisions often set the baseline for agency guidelines: the FBI’s Domestic Investigations and Operations Guide, for instance, references judicial opinions heavily in outlining First Amendment constraints on investigations.205

201 See Kagan, supra note 181, at 440-41.
203 For the classic political process argument for searching judicial review, see JOHN HART ELY, DEMOCRACY AND DISTRUST: A THEORY OF JUDICIAL REVIEW (1980); United States v. Carolene Prods. Co., 304 U.S. 144, 152 n.4 (1938).
205 OPERATIONS GUIDE, supra note 102, at 24-30.
The case for heightened scrutiny is justifiable, though not compelled, by existing law. The law on First Amendment profiling is deeply unsettled, as lower courts since the 1970s have struggled to answer the many questions the Supreme Court has left open. The first important question that has divided lower courts evaluating First Amendment profiling involves the circumstances in which plaintiffs challenging law enforcement investigations have suffered sufficient abridgment of their First Amendment rights. The harm from law enforcement investigations is less direct and certain than the harm from classic restrictions on speech or associational rights because investigations do not typically prohibit someone from expressing a view or associating with an organization. Lower courts since the 1970s have primarily confronted the question of sufficient harm as an issue of standing, and many have denied standing to plaintiffs challenging surveillance or investigations. Despite the restrictive standing doctrine created in this context, I argue that lower court decisions, and the Supreme Court precedent on which they rely, suggest that some plaintiffs challenging law enforcement interviews would be able to establish standing—overcoming a threshold hurdle that has stymied other First Amendment claims against covert surveillance or the observation of religious or political events. While plaintiffs will still need to establish, on the merits, that they suffered substantial harm, standing doctrine in this context already screens out most cases involving minimal harm, and the stigmatic and chilling effects of interviews should further help plaintiffs demonstrate the seriousness of harm suffered.

A second question that has divided lower courts is whether the government’s articulation of a plausible legitimate purpose for a law enforcement investigation ends the First Amendment inquiry, or whether courts ought to apply heightened scrutiny to probe the “fit” between the claimed purpose and the means employed—either because heightened scrutiny is required to uncover an illicit motive (the suppression concern) or because even an investigation with an acknowledged legitimate purpose may still sweep too broadly in burdening First Amendment rights (the overbreadth concern). Despite the fundamental importance of this question, only one recent piece—an article by Lawrence Rosenthal discussed below—has discussed this divide, concluding that First Amendment investigations need not satisfy heightened
scrupulously. I examine the split among lower courts on First Amendment profiling and challenge the view that First Amendment doctrine does not support heightened scrutiny for investigations. In doing so, I address the most significant objection to this view: the idea that narrow tailoring requirements would undermine the ability of law enforcement to identify and preempt potential threats.

A. The Harm from Investigations

1. The Supreme Court’s Chilling Effects Cases Through Laird

In the 1950s and 1960s, the Supreme Court made clear that government actions that indirectly burden, rather than directly restrict, speech or association could violate the First Amendment. Several classic cases establishing this principle arose out of investigations or inquiries into the membership of the National Association for the Advancement of Colored People (NAACP) or “subversive” organizations. Thus, in NAACP v. Alabama, the Supreme Court held that the state of Alabama had failed to show a compelling justification for requiring the NAACP to disclose its membership list, which the state sought as part of an investigation into whether the group had complied with a corporate registration requirement. The Court recognized that although the requirement did not directly abridge NAACP members’ freedom to associate, the practical consequence of disclosure would be to subject members to public threats and reprisals—and chill the organization’s lawful political advocacy.

Although in many cases, the Court surely suspected an illicit government motive behind an investigation—the suppression concern—these decisions did not conclude that the government purposes articulated were pretextual or implausible. For instance, in Shelton v. Tucker, the Court explicitly acknowledged the state’s legitimate interest in investigating a public employee’s associational ties, but

---

208 NAACP, 357 U.S. at 466.
209 Id. at 462-63.
nevertheless concluded that that interest did not justify the breadth of the inquiry undertaken.\footnote{Shelton, 364 U.S. at 490.}

Moreover, the Court sometimes proved willing to engage in a rigorous examination of the facts in finding an insufficient “fit” between the state interest articulated and the information-gathering measure employed. Thus, in \textit{Gibson v. Florida Legislative Investigative Committee}, the Court held that while state legislatures had broad power to investigate possible subversive activities, a Florida legislative committee’s demand that the NAACP disclose whether fourteen “known” Communists were members of the association was not supported by “a substantial relation between the information sought and a subject of overriding and compelling state interest.”\footnote{\textit{Gibson}, 372 U.S. at 546.}

The state presented colorable evidence of Communist attempts to participate in the NAACP: some of the fourteen alleged Communists had attended NAACP meetings, one informant had “been instructed to infiltrate the NAACP,” the organization had passed “antisubversion” resolutions that acknowledged attempts by Communists to take over the group, and one or two of the fourteen had given talks to the organization or distributed leaflets for the group.\footnote{\textit{Id.} at 552-54, 554 n.6.} Yet the Court rejected this evidence, fact by fact, as “merely indirect, less than unequivocal, and mostly hearsay testimony.”\footnote{\textit{Id.} at 555.}

Without deferring to either the government’s explanation for the investigation or its interpretation of the facts, the Court conducted its own, searching inquiry into whether the evidence justified the investigation.\footnote{\textit{Gibson} distinguished several earlier decisions that upheld disclosure requirements aimed at the Communist Party on the grounds that regulating the Communist Party did not present the same constitutional issues as disclosure requirements aimed at other political groups. \textit{Id.} at 547-48.}

In these cases, the Court explicitly characterized the harm resulting from government investigative practices as the chilling effect such practices would have on legitimate expression and association. For instance, Alabama’s attempt to compel disclosure of the NAACP’s membership would hinder the group’s advocacy efforts by dissuading individuals from participating out of fear of exposure.\footnote{\textit{NAACP v. Alabama ex rel. Patterson}, 357 U.S. 449, 462-63 (1958).} Less explicitly, the decisions recognize that the chilling effect on First Amendment
rights results from an exercise of state power that subjects individuals to stigma—either directly or by exposing people to public hostility. In NAACP v. Alabama, the Court compared the compelled disclosure of group affiliation to a “requirement that adherents of particular religious faiths or political parties wear identifying armbands,” drawing an implicit connection to the branding of minority or dissenting communities. In a case involving legislative investigations into First Amendment views, the Court noted that such investigations could impose the “stain of the stamp of disloyalty” as significant as might result from an actual loss of employment.

In spite of these decisions recognizing the stigmatic harm and chilling effects of indirect government burdens on expression, in 1972, the Court curbed the practical reach of these doctrines by restricting standing to challenge surveillance. In Laird v. Tatum, the Court announced that while indirect burdens on First Amendment rights might warrant constitutional review, individuals did not have standing to challenge government surveillance that created only a “subjective ‘chill’” on political activities. Laird remains the Court’s last word on First Amendment claims against government surveillance of political activities, and has profoundly affected First Amendment challenges to law enforcement investigations and surveillance.

In Laird, a group of political activists and organizations sought to enjoin the army’s collection of intelligence regarding their lawful political activities, claiming that the army had surveilled them and continued to maintain files regarding their political activities. The army acknowledged conducting surveillance (mostly through public sources such as the news media and attendance at public meetings) of political activities with the potential to result in civil disorder. The Court distinguished the surveillance in Laird from past “chilling effect” cases, where “the challenged exercise of governmental power was regulatory, prescriptive, or compulsory in nature,” and plaintiffs were or would be “subject to the regulations,

---

216 Id. at 462.
218 408 U.S. 1, 12-14 (1972).
219 Id. at 2.
221 Id. at 6-7.
proscriptions, or compulsions” that they were contesting.\textsuperscript{222} For instance, in past cases, plaintiffs stood to lose out on state bar membership or employment if they did not take loyalty oaths or answer questions about their political associations.\textsuperscript{223} In Laird, by contrast, the Court characterized the plaintiffs’ claims as arising from “the mere existence, without more, of a governmental investigative and data-gathering activity that is alleged to be broader in scope than is reasonably necessary for the accomplishment of a valid governmental purpose.”\textsuperscript{224} A “subjective ‘chill’” alone, such as the plaintiffs’ fears that the army might in the future misuse the information it had collected against them, could not substitute for a “claim of specific present objective harm or a threat of specific future harm.”\textsuperscript{225}

Although Laird did not address the merits of the plaintiffs’ First Amendment claims, the five-to-four opinion suggested a reluctance to credit the stigma of general surveillance in the absence of other state action directly affecting the plaintiffs. But the Court dealt there with information-gathering from largely public sources, noting that the lower court described the investigation as uncovering no more than what a good newspaper reporter might uncover from attending meetings and clipping articles.\textsuperscript{226} The plaintiffs did not allege any law enforcement interviews of individuals or intimidating contacts between individuals and army officers, and learned of the army’s surveillance through a magazine article that described the program.\textsuperscript{227} The Court qualified its concern over federal courts monitoring “the wisdom and soundness of Executive action” by making clear that an actual or immediately threatened injury could properly subject army surveillance to judicial review.\textsuperscript{228} Thus, the Court left open the ability of courts to scrutinize overbroad and stigmatizing investigations in a case where plaintiffs could demonstrate “objective” harm.

2. Establishing Standing for Interviews, Post-Laird

Since Laird, lower courts facing First Amendment claims against law enforcement investigations have grappled

\textsuperscript{222} Laird, 408 U.S at 11.
\textsuperscript{223} Id.
\textsuperscript{224} Id. at 10.
\textsuperscript{225} Id. at 13-14.
\textsuperscript{226} Id. at 9 (quoting Tatum v. Laird, 444 F.2d 947, 953 (D.C. Cir. 1971)).
\textsuperscript{227} Id. at 2 n.1.
\textsuperscript{228} Id. at 15-16.
with the threshold question of whether the plaintiffs have standing to sue. A review of these decisions suggests that at least some plaintiffs challenging law enforcement interviews as unjustified or overbroad should be able to establish standing.

In the case of CBP interviews, where travelers cannot enter the United States without satisfying border inspectors, Laird is unlikely to present a challenge. Beyond a chilling effect, individuals interviewed on account of First Amendment expression can point to objective harm from interviews, minimally the detention they experienced—often in addition to intrusive searches, handcuffing or displays of physical force, or other screening measures. In contexts involving the compelled disclosure of information, the Supreme Court has never held that organizations or individuals personally subject to that compulsion lacked standing. Although CBP presumably cannot deny entry altogether to U.S. citizens who refuse to speak, CBP can delay entry or subject people to more intense searches; even with a legal right to enter, U.S. citizens must still await the permission of border inspectors. Thus, for purposes of standing, courts should view border interviews as a type of involuntary detention, creating an objective injury.

FBI interviews, by contrast, present a more difficult question because they are ostensibly voluntary, and plaintiffs will not always be able to establish objective injury beyond the stigma or chilling effects of the practice. As Scott Michelman and others have described, some courts have interpreted Laird broadly to foreclose standing for claims of chilling effect injuries that did not arise out “of ‘regulatory, proscriptive, or compulsory’ government action.” For instance, the lead opinion in a Sixth Circuit case adopted this position and denied standing to plaintiffs who challenged the National Security Administration’s warrantless wiretapping program; the plaintiffs could not establish that they were “regulated, constrained, or compelled directly by the government’s” program.

---

229 Standing was not an issue in Tabbaa, for instance, where on the merits the Second Circuit found that even without the “clear chilling of future expressive activity,” the screening measures imposed a significant penalty on plaintiffs. Tabbaa v. Chertoff, 509 F.3d 89, 101 (2d Cir. 2007).


231 ACLU v. Nat’l Sec. Agency, 493 F.3d 644, 661 (6th Cir. 2007) (Batchelder, J.); see also United Presbyterian Church in the U.S.A. v. Reagan, 738 F.2d 1375, 1378-79
Courts that embrace the broad view of Laird might find that chilling-effects claims arising out of FBI interviews are nonjusticiable since individuals need not agree to the interviews. On the other hand, plaintiffs approached by the FBI for interviews are subject to direct, personal contact with law enforcement making a request of them—in contrast to the surveillance in Laird, which did not require individuals to do anything at all.

Moreover, other courts have not interpreted Laird so broadly as to foreclose First Amendment standing in the absence of “regulatory, prescriptive, or compulsory” government action. Rather, many courts distinguished cases where litigants showed only a “subjective ‘chill’” from cases where plaintiffs established an “objective” chill, or specific adverse effects, from surveillance or investigations. Most recently, the Second Circuit held that plaintiffs who had taken costly measures to avoid surveillance of their international communications in reasonable fear of being monitored had standing to challenge a new foreign intelligence surveillance law.233

These decisions espouse a narrower view of Laird and suggest at least three bases for asserting justiciable challenges to FBI interviews that burden free expression and association. First, plaintiffs can establish standing where they demonstrate that interviews were aimed at disrupting speech or lawful activities or were undertaken in bad faith.234

(D.C. Cir. 1984) (denying standing where plaintiffs could not show that executive orders authorizing intelligence collection commanded or prohibited them from doing anything).

232 Thus, several courts dismissed challenges to otherwise legal law enforcement data-gathering, photography, and physical surveillance at political events open to the public. See, e.g., Phila. Yearly Meeting of Religious Soc’y of Friends v. Tate, 519 F.2d 1335, 1337-38 (3d Cir. 1975); Fifth Ave. Peace Parade Comm. v. Gray, 480 F.2d 326, 330-33 (2d Cir. 1973); Donohoe v. Duling, 465 F.2d 196, 202 (4th Cir. 1972). One court, finding no tangible harm, denied standing to students and teachers who claimed that the placement of an undercover police officer in two high school classes was ideologically motivated and stifled classroom discussion. Gordon v. Warren Consol. Bd. of Educ., 706 F.2d 778, 780-81 (6th Cir. 1983).

233 Amnesty Int’l USA v. Clapper, 638 F.3d 118, 121-22 (2d Cir. 2011).

234 See, e.g., Anderson v. Davila, 125 F.3d 148, 160 (3d Cir. 1997) (distinguishing Laird where plaintiff alleged that government surveilled him in retaliation for exercise of First Amendment rights); Alliance to End Repression v. City of Chicago, 627 F. Supp. 1044, 1047, 1050-52 (N.D. Ill. 1985) (finding standing where Chicago police placed informants and undercover agents in senior positions in organizations in order to “neutralize” their influence); Founding Church of Scientology of Wash., D.C. v. Dir., Fed. Bureau of Investigation, 459 F. Supp. 748, 760 (D.D.C. 1978) (ruling that Church of Scientology had standing where alleged that federal government had disrupted the organization and interfered with its activities); Berlin Democratic Club v. Rumsfeld, 410 F. Supp. 144, 149-51 (D.D.C. 1976) (holding that plaintiffs had standing to contest surveillance activities that included terminating...
Second, plaintiffs can overcome Laird where the FBI’s actions cause individuals objective harm, such as damage to reputations or employment prospects. This type of harm sometimes occurs when FBI agents approach individuals at work and announce to employers that they want to investigate an employee. Indeed, in another context, the Supreme Court distinguished reputational injuries as a basis for standing from the chilling effects found nonjusticiable in Laird. Lower courts have allowed individuals to challenge employment-related government loyalty investigations, the public disclosure of information on individuals targeted, and even the harm resulting from retention of an investigative file that carried a risk of potential future disclosure.

Third, an organization that has experienced an identifiable decline in membership, support, or reputation from FBI interviews, such as a mosque whose congregants are questioned, could sue on its own behalf. Courts have ruled that organizations have standing to contest law enforcement investigative practices that dissuaded their members from participating and thereby caused the institutions tangible harm, such as a decline in membership or participation. For instance, the Ninth Circuit found that churches sheltering refugees had standing to challenge surveillance of their activities because the chilling impact on their members concretely impaired the churches from carrying out their ministries.


235 Meese v. Keene, 481 U.S. 465, 473-75 (1987) (holding that state legislator who sought to exhibit foreign films had standing to challenge statutory scheme labeling the films “political propaganda” because of potential harm to his reputation and professional interests).


238 Paton v. La Prade, 524 F.2d 862, 868 (3d Cir. 1975).


240 Presbyterian Church, 870 F.2d at 521-22.
While Laird imposes a significant threshold requirement that could screen out many FBI encounters—even large-scale interviews with significant chilling effects, where plaintiffs did not have evidence of bad faith or tangible injuries—it does not foreclose First Amendment challenges to CBP or FBI interviews.

3. Beyond Standing: Establishing Sufficient Harm to First Amendment Rights

Even where courts find standing, they would still need to consider the extent of the harm, either as a threshold inquiry into whether the harms are “substantial,” or at a minimum, in determining whether plaintiffs’ injuries ought to prevail against governmental interests. Interviews that survive Laird’s standing doctrine will usually be those where the government effectively penalized or compelled plaintiffs in some way, as in FBI interviews that cause reputational or economic damage or compulsory CBP inspections.

For CBP interviews, the compulsion involved brings potential challenges within the ambit of the Court’s historic chilling effect cases, where the government threatened to withhold a concrete benefit based on protected expression. When CBP prevents individuals who have engaged in protected expression from entering the country until they comply with extensive and unusual CBP interviews, those individuals experience an actual restraint at least as severe as that invalidated in Lamont v. Postmaster General. There, the Court struck down a rule preventing the delivery of foreign mail that the Post Office determined was “communist political propaganda” until recipients first sent in a reply card affirmatively requesting the delivery.

Though sending a reply card to receive mail did not create the same burden as, say, an employment requirement to disclose one’s political affiliations,

---

241 See, e.g., Tabbaa v. Chertoff, 509 F.3d 89, 101-02 (2d Cir. 2007) (evaluating whether burden on freedom of association was “substantial”).
243 381 U.S. 301, 307 (1965). Whether the First Amendment analysis would differ in border inspections of individuals who are not U.S. citizens or lawful permanent residents is beyond the scope of this paper. While noncitizens residing in the United States enjoy First Amendment rights, Bridges v. Wixon, 326 U.S. 135, 148 (1945), some courts have rejected First Amendment challenges with respect to certain immigration decisions. See Reno v. Am.-Arab Anti-Discrimination Comm., 525 U.S. 471, 488 (1999); Price v. INS, 962 F.2d 836, 841 (9th Cir. 1992).
244 Lamont, 381 U.S. at 307.
the Court found the constraint presented by the regulation sufficient to implicate the First Amendment. The stigma imposed by the designation of certain mail as “communist political propaganda,” not just the degree of restraint, made the burden on First Amendment rights significant; the Court recognized that such a designation would deter individuals from reading “what the Federal Government says contains the seeds of treason” and has “condemned.”

In both FBI and CBP interviews, the stigma resulting from a person’s selection for questioning on account of First Amendment activities should help plaintiffs establish that a law enforcement encounter effectively penalized them for engaging in protected expression. That reasoning appeared in Tabbaa, where the Second Circuit relied on the stigma of the investigative measures taken to conclude that they imposed a substantial disability on the plaintiffs’ First Amendment rights. The court found that the extensive security measures the plaintiffs experienced, “when others, who had not attended the conference, did not have to endure these measures,” qualified as a significant “penalty” and one that might reasonably deter others from attending similar conferences.

In other words, the fact of being singled out for this type of scrutiny, not just the “tangible” harm from the screening measures employed, led the court to find that the investigation substantially interfered with freedom of association.

As discussed above, the Supreme Court’s classic First Amendment cases of the 1950s and 1960s explicitly recognized chilling effects, and implicitly recognized stigma, as a basis for harm. Moreover, the Court frequently noted that the chilling effects on speech or association in those cases were grave precisely because the groups or individuals being investigated espoused dissenting or unpopular views. Thus, the Court considered the actual burden of investigations on the specific communities involved in the specific social and political atmosphere of the time—a backdrop of hostility toward the civil rights movement and pervasive fear of communism. Courts examining the harm of First Amendment profiling

---

245 Id.
246 Id.
247 Tabbaa v. Chertoff, 509 F.3d 89, 102 (2d Cir. 2007).
248 See supra notes 214-16 and accompanying text.
today could—and should—consider broader discrimination and hostility to U.S. Muslims in evaluating the extent to which interviews stigmatize targets and chill expression.

Beyond the First Amendment context, the Court has continued to consider stigmatic harm a judicially cognizable injury. In the equal protection context, the Court recognizes—as it did most famously in Brown v. Board of Education\(^\text{250}\)—that discrimination imposes a noneconomic injury by “stigmatizing members of the disfavored group” as “less worthy participants in the political community.”\(^\text{251}\) The Court considers stigma, when accompanied by more tangible harm, a factor in determining what procedural protections the Due Process Clause requires\(^\text{252}\) and as an important reason that laws prohibiting private, consensual sexual acts between persons of the same sex violate substantive due process.\(^\text{253}\) In criminal procedure, the Court has justified rigorous procedural protections partly because of the stigma imposed by a criminal conviction.\(^\text{254}\)

Thus, the fact that law enforcement interviews do not prohibit someone from expressing a view or associating with an organization should not categorically prevent individuals subjected to FBI or CBP interviews from establishing sufficient harm to First Amendment interests, either as a matter of standing doctrine or substantive law. Plaintiffs will be on stronger ground, however, where they can point to stigma and chilling effects in addition to more “objective” compulsion or harm.

B. Suppression, Overbreadth, and Heightened Scrutiny in the Lower Courts

Post-Laird cases involving challenges to First Amendment profiling suggest broad agreement that the First Amendment bars investigations that are directed at an illegitimate purpose. But courts divide as to whether a showing of a legitimate purpose ends the First Amendment inquiry, or whether investigations with a legitimate purpose may still sweep too broadly in burdening free speech, association, and religion as to violate the

\(^\text{250}\) 347 U.S. 483, 494-95 (1954).
\(^\text{253}\) Lawrence v. Texas, 539 U.S. 558, 575 (2003).
First Amendment. And courts divide further as to whether heightened scrutiny should apply to such investigations.

Lower court decisions in First Amendment challenges to law enforcement investigations have generally concluded that the First Amendment bars investigations that are undertaken for the purpose of stifling dissent or interfering with lawful expressive activities. Beginning in the 1970s, a series of lawsuits challenged federal law enforcement and local police investigative practices that allegedly aimed to stifle dissent. Beyond finding standing where the law enforcement practices in question appeared to go beyond “legitimate surveillance,” a number of courts held that investigations taken for the purpose of harassment, interference with lawful activities, or suppression of speech would violate the First Amendment.

For instance, in Hobson v. Wilson, political activists involved in antiwar, civil rights, and other political causes sued the FBI and Washington, D.C., police for conspiring to deprive them of their rights to free speech and association. Plaintiffs had a rare smoking gun: internal FBI memoranda explicitly described the agency’s “Cointelpro” operation as seeking to “disrupt,” “discredit,” and “otherwise neutralize” the lawful activities of “New Left” and “Black Nationalist” activists in a purported attempt to prevent potential civil strife and

255 See Brief for Respondents, supra note 220, at 6a.

256 United States v. Mayer, 503 F.3d 740, 752 (9th Cir. 2007) (stating that “the government must not investigate for the purpose of violating First Amendment rights, and must also have a legitimate law enforcement purpose”); Anderson v. Davila, 125 F.3d 148, 159-60 (3d Cir. 1997) (finding that tactics that ordinarily do not implicate the First Amendment, such as photographing and surveilling a person in public, violate the law when undertaken for a wrongful purpose, such as retaliation for a person’s exercise of First Amendment rights); United States v. Aguilar, 883 F.2d 662, 705 (9th Cir. 1989), superseded by statute on other grounds, Pub. L. No. 99-603, 100 Stat. 3359 (1986) (concluding that the use of informants to infiltrate an organization must not be for “the purpose of abridging [F]irst [A]mendment freedoms”); Socialist Workers Party v. Atty Gen., 642 F. Supp. 1357, 1364, 1416 (S.D.N.Y. 1986) (concluding that the FBI’s disruption of Socialist Workers Party’s lawful political activities violated First Amendment rights of speech and assembly and awarding damages); Ghandi v. Police Dep’t of City of Detroit, 747 F.2d 338, 349-50 (6th Cir. 1984) (suggesting that informant’s alleged misrepresentation of labor party’s goals, disruption of political campaign, and theft of party documents would violate First Amendment); Jabara v. Kelley, 476 F. Supp. 561, 574 (E.D. Mich. 1979), vacated on other grounds sub nom. Jabara v. Webster, 691 F.2d 272 (6th Cir. 1982) (denying summary judgment on First Amendment claims where factual issues existed as to whether FBI investigation of attorney’s political activities was in good faith); see also Alliance to End Repression v. City of Chicago, 561 F. Supp. 537, 549, 559 (N.D. Ill. 1982) (approving settlement agreement forbidding investigation or disruption because of First Amendment conduct).

violence.\textsuperscript{258} Holding that the alleged interference with plaintiffs’ lawful political activities violated “fundamental and well-established constitutional rights,” the court stated that “it is never permissible to impede or deter lawful civil rights/political organizations, expression or protest with no other direct purpose and no other immediate objective than to counter the influence of the target associations.”\textsuperscript{259} In so ruling, Hobson implies that even where the ultimate goal of an investigation is legitimate (the prevention of violence), the immediate purpose cannot be the disruption of lawful First Amendment activity.

Cases such as Hobson involved a range of law enforcement practices, and one might question whether courts would find interviews undertaken to suppress lawful expression to be per se impermissible. Decisions involving interviews are rare. Yet a case involving the Socialist Workers Party treated interviews no differently from other methods calculated to disrupt the organization’s lawful political activities, finding such methods “patently unconstitutional.”\textsuperscript{260} The court noted that while interviews to gather information are usually a legitimate FBI activity, the FBI in that case used interviews of the organization’s members and their relatives, employers, and landlords to foment paranoia within the group.\textsuperscript{261} In Zieper v. Metzinger, the Second Circuit Court of Appeals held that FBI requests to a filmmaker to remove a film from the Internet that could provoke violence would violate the First Amendment if they crossed the line from persuasion to coercion.\textsuperscript{262} The court held that the First Amendment prohibits such requests where the “totality of the circumstances” amount to an implied threat.\textsuperscript{263} Thus, while courts would likely hold that interviews undertaken to suppress lawful activities through harassment violate the First Amendment, in circumstances like Zieper, courts might consider whether the interviews conveyed an implied threat to refrain from expression.

While lower courts have generally found that a purpose to suppress First Amendment rights cannot support an investigation, courts divide as to whether a showing of a legitimate purpose ends the inquiry, or whether investigations with a legitimate purpose may still sweep so broadly that they violate the First Amendment.

\begin{footnotes}
\item[258] Id. at 10.
\item[259] Id. at 27.
\item[260] Socialist Workers Party, 642 F. Supp. at 1416-17.
\item[261] Id. at 1389.
\item[262] 474 F.3d 60, 65-66 (2d Cir. 2007).
\item[263] Id. at 70-71.
\end{footnotes}
The lower courts further disagree as to whether they should employ heightened scrutiny to evaluate an investigation where the government provides a facially legitimate reason for its conduct—either to “smoke out” an actual illicit purpose or to invalidate the overbreadth of the investigation.

Rosenthal, the only commentator to review this split in authority, concludes that most courts have held that investigations into First Amendment activities require no more than a “good faith” law enforcement interest supporting the investigation. That conclusion, however, understates the extent of conflict even within federal circuit courts of appeals. Thus, the Ninth Circuit, which had ruled that a government investigation threatening First Amendment rights need only be justified by a legitimate law enforcement purpose, modified its decision to state that that purpose also had to “outweigh[] any harm to First Amendment interests.” In an influential early decision, the D.C. Circuit declared that only a legitimate purpose was necessary to subpoena reporters’ phone records. In two later decisions, though, the D.C. Circuit stated that an investigation with a legitimate purpose might still be unconstitutional, and the court called for strict scrutiny of investigations that significantly burdened First Amendment rights.

Amid the confusion, two broad views appear in the case law. One set of cases treats the requirement of a legitimate purpose as the only limitation that the First Amendment places on law enforcement investigations. These decisions

---

265 United States v. Mayer, 503 F.3d 740, 753 (9th Cir. 2007), superseding 490 F.3d 1129.
268 United States v. Aguilar, 883 F.2d 662, 708-09 (9th Cir. 1989) (refusing to suppress, on First Amendment grounds, evidence obtained from the INS’s warrantless use of undercover agents and informants to infiltrate church meetings), superseded by statute on other grounds, Pub. L. No. 99-603, 100 Stat. 3359 (1986); Gordon v. Warren Consol. Bd. of Educ., 706 F.2d 778, 780-81 (6th Cir. 1983) (holding that the placement of undercover police officer in two high school classes to uncover drug trafficking did not violate the First Amendment, so long as investigation was not designed to control content of class discussions); Reporters Comm. for Freedom of the Press, 593 F.2d at 1050 (ruuling that the government’s good faith issuance of subpoenas to phone companies for reporters’ phone records did not abridge First Amendment rights despite incidental burden on news gathering); Jabara v. Kelley, 476 F. Supp. 561, 572-73 (E.D. Mich. 1979) (holding that good faith surveillance of a political activist would not violate the First Amendment), vacated on other grounds sub nom. Jabara v. Webster, 691 F.2d 272 (6th Cir. 1982); Anderson v. Sills, 265 A.2d 678, 688 (N.J. 1970) (holding that
generally do not make explicit how a court would ferret out such an illegitimate purpose, but the language of these cases suggests broad deference to the government’s stated justifications rather than any hint of narrow tailoring.269

Perhaps the strongest statement of this view appears in several Seventh Circuit opinions that narrowed landmark consent decrees entered against the FBI and Chicago Police Department in Alliance to End Repression v. City of Chicago.270 The defendants in that case acknowledged illegal surveillance and disruption of civil liberties organizations, religious activists, and political groups.271 Three years after the district court approved the consent decree against the FBI, the Seventh Circuit interpreted the prohibition on investigating First Amendment activities narrowly to bar investigations only where there was no “genuine concern for law enforcement.”272 Later decisions interpreted the consent decree to exclude “negligent” noncompliance273 and modified the Chicago Police Department’s order to allow the agency to monitor political extremist groups to prevent “ideological terrorism,” regardless of whether the police had reasonable suspicion of any crime.274 The Alliance decisions make the court’s view clear: the First Amendment itself, not just a proper reading of the consent decree, prohibited only those investigations with the purpose of interfering with lawful expression275 and any more exacting requirement would impede law enforcement from protecting public safety.276

A second set of cases, however, rejects the view that under the First Amendment a lawful government purpose is the courts should not interfere with executive information-gathering in the absence of proof of bad faith or arbitrariness).269 One exception is Presbyterian Church v. United States, 752 F. Supp. 1505, 1513-15 (D. Ariz. 1990), which purported to apply strict scrutiny. In any event, that district court’s use of strict scrutiny does not seem compatible with the Ninth Circuit’s deferential inquiry in United States v. Mayer, 503 F.3d 740, 748-49 (9th Cir. 2007).

270 Alliance to End Repression v. City of Chicago, 742 F.2d 1007, 1015 (7th Cir. 1984); Alliance to End Repression v. City of Chicago, 119 F.3d 472, 476 (7th Cir. 1997); Alliance to End Repression v. City of Chicago, 237 F.3d 799, 802 (7th Cir. 2001).

271 Alliance to End Repression, 742 F.2d at 1015.

272 Id. Specifically, the court interpreted the decree to permit investigations of groups based on their advocacy of violence, even where that advocacy was protected by the First Amendment.

273 Alliance to End Repression, 119 F.3d at 476 (holding that FBI investigation into the Committee in Support of the People of El Salvador did not violate the consent decree even where the FBI acknowledged widespread negligence).

274 Alliance to End Repression, 237 F.3d at 802.

275 See, e.g., Alliance to End Repression, 742 F.2d at 1010, 1015-16; Alliance to End Repression, 237 F.3d at 800.

276 See, e.g., Alliance to End Repression, 237 F.3d at 802.
sole test of the validity of a law enforcement investigation and instead also requires some form of heightened scrutiny to evaluate law enforcement investigations that burden free expression. Thus, in Clark v. Library of Congress, the D.C. Circuit required “exacting scrutiny” of an individual’s claim that an FBI investigation into his political activities had chilled his expression and prevented him from obtaining further federal employment. The FBI had conducted a wide-ranging loyalty investigation into the plaintiff, mostly through interviews of his neighbors, friends, and associates because of allegations that he belonged to a lawful socialist organization. The court ruled that regardless of whether the government “intended to punish or coerce the individual . . .[,][w]here the government’s action inflicts a palpable injury on the individual because of his lawful beliefs,” it should demonstrate that it employed the “least restrictive” alternative.

While Clark concerned an FBI investigation for employment purposes, other courts suggest a similar test for law enforcement investigations directed squarely at uncovering security threats. Tabbaa itself, while ultimately ruling against the plaintiffs, applied heightened scrutiny to the plaintiffs’ freedom-of-association claims. The court did not conclude its inquiry in finding a legitimate purpose for the investigation (identifying individuals who might pose a threat based on CBP intelligence on the conference), but further analyzed whether the government’s actions “serve[d] compelling state interests,

\[\text{Clark v. Library of Cong., 750 F.2d 89, 94 (D.C. Cir. 1984); Hobson v. Wilson, 737 F.2d 1, 27, 27 n.85, 28 (D.C. Cir. 1984) (stating that the “clearly established” right to freedom of association included the principle that government action taken for a legitimate purpose would nevertheless be unconstitutional where it “significantly interfered with protected rights of association, unless the [g]overnment . . . demonstrate[d] a substantial . . . or compelling . . . interest” that “could not more narrowly be accommodated” (citations omitted)), overruled in part on other grounds by Leatherman v. Tarrant Cnty. Narcotics Intelligence & Coordination Unit, 507 U.S. 163 (1993); White v. Davis, 533 P.2d 222, 227-28 (Cal. 1975) (requiring strict scrutiny of Los Angeles Police Department’s covert surveillance of university classrooms, which police justified as intelligence-gathering to prevent future criminal activity); Tabbaa v. Chertoff, 509 F.3d 89, 102, 105-06 (2d Cir. 2007). With the exception of United States v. Mayer, 503 F.3d 740, 748-50 (9th Cir. 2007), most courts that express concern for the “overbreadth” of an investigation, not just its purpose, require some form of heightened scrutiny.}\]
unrelated to the suppression of ideas, that [could not] be achieved through means significantly less restrictive.\(^\text{282}\)

Notably, Clark and Tabbaa were both cases that involved claims of more concrete harm—loss of future employment and reputational injuries or detentions with intrusive screening measures—in addition to chilling effects and stigma. These cases suggest that while lower courts are split as to whether First Amendment profiling requires heightened scrutiny, plaintiffs demonstrating more objective harms from interviews are most likely to garner such review.

C. Addressing Objections to Heightened Scrutiny

Critics of heightened scrutiny for First Amendment profiling argue that judicially-imposed narrow tailoring requirements would undermine law enforcement investigations, which depend on discretion and flexibility to identify and preempt potential threats. One argument is that the Supreme Court, out of concern for law enforcement interests, has refused to create a higher standard for law enforcement investigations burdening First Amendment activities in other contexts. A second objection is that heightened scrutiny of First Amendment

\(^{282}\) Id. at 97, 102. Note that while the court applied strict scrutiny, one judge suggested that a less demanding standard might be appropriate because the inspections occurred at the border. Id. at 102 n.5. The extent of First Amendment protections at U.S. borders is unclear. See generally Timothy Zick, Territoriality and the First Amendment: Free Speech at—and Beyond—Our Borders, 85 NOTRE DAME L. REV. 1543 (2010). In United States v. Ramsey, 431 U.S. 606, 623-24 (1977), the Supreme Court ruled that border searches of incoming international mail did not violate the First Amendment where existing regulations forbade such searches without reasonable suspicion and flatly prohibited reading mail without a search warrant. The Court refused comment on the “constitutional reach of the First Amendment” in the absence of such protections. Id. at 624. Lower courts have considered First Amendment claims in recent challenges to warrantless border searches of laptop computers. Two appeals courts rejected First Amendment arguments in refusing to suppress evidence of child pornography uncovered through warrantless searches of defendants’ laptop computers. See United States v. Arnold, 533 F.3d 1003, 1010 (9th Cir. 2008); United States v. Ickes, 393 F.3d 501, 502-07 (4th Cir. 2005). A district court is now considering a Fourth and First Amendment challenge to a Department of Homeland Security policy permitting searches and copying of laptop computers and other electronic devices without reasonable suspicion. See generally Complaint, Abidor v. Napolitano, No. 1:10-cv-04059 (E.D.N.Y. Sept. 7, 2010), available at http://www.aclu.org/free-speech-technology-and-liberty/abidor-v-napolitano-complaint. While the border search doctrine in Fourth Amendment law rests on the idea that individuals have a lesser expectation of privacy at the border, these cases do not suggest that individuals at the border have diminished First Amendment rights not to be singled out on the basis of protected speech or association. Neither Arnold nor Ickes involved a claim of First Amendment profiling, but the separate First Amendment claim that searches of laptop computers require individual suspicion because computers store expressive materials. See Arnold, 533 F.3d at 1006; Ickes, 393 F.3d at 506.
profiling would place courts in the inappropriate role of second-guessing law enforcement judgments in critical terrorism-related investigations. A third argument is that law enforcement agents need clear, consistent rules when making decisions about whom and how to investigate, and narrow tailoring requirements would force law enforcement agents to make complicated “balancing” decisions whenever a First Amendment interest is implicated.

First, the argument from precedent, which appears both in lower court decisions rejecting heightened scrutiny as well as in recent commentary, stems from the fact that the Supreme Court in several cases has refused to create greater protections for law enforcement investigative practices that burden First Amendment rights. For instance, the Court has held that the ordinary Fourth Amendment warrant requirement sufficiently protects First Amendment interests in the seizure of allegedly obscene films or the search of a newspaper office, rejecting arguments that the expressive materials or institutions involved required the provision of a prior adversary hearing or a higher showing of necessity.

Similarly, in Branzburg v. Hayes, the Court ruled that requiring reporters to testify before grand juries on information obtained from confidential sources did not abridge the First Amendment, despite claims that the disclosure of confidential sources would deter people from speaking freely to reporters and diminish the press's ability to gather news. The Court concluded that grand jury subpoenas of reporters required only a good faith law enforcement interest—creating a standard that Rosenthal notes is “remarkably like that employed by the courts that have rejected any form of heightened scrutiny for First Amendment investigations.” The Court deemed the burden on First

---


284 New York v. P.J. Video, Inc., 475 U.S. 868, 875 (1986) (holding that ordinary probable cause standard applied to warrant application for sexually explicit videos presumptively protected by First Amendment); Zurcher v. Stanford Daily, 436 U.S. 547, 567-68 (1978) (holding that Fourth Amendment warrant requirement offered sufficient protection in searches of a newspaper office); Heller v. New York, 413 U.S. 483, 492 (1973) (holding that the seizure of a single allegedly obscene film for evidentiary purposes did not require a prior adversary hearing); see also Univ. of Penn. v. EEOC, 493 U.S. 182, 201 (1990) (rejecting a university's First Amendment claim that subpoenas requesting tenure-related materials should be supported by a higher standard of necessity).


286 Rosenthal, supra note 16, at 54-56.
Amendment rights “incidental” and found any such interest outweighed by the public interest in grand juries exercising broad power to investigate potential criminal conduct.\(^{287}\)

Despite the Court’s unwillingness to create a special standard for searches and seizures of expressive material or testimony that potentially chills expression, these decisions do not require a rejection of heightened scrutiny in cases of First Amendment profiling. These decisions all involved First Amendment challenges to law enforcement procedures where other protections against government overreaching already existed; while affirming the important First Amendment interests involved,\(^{288}\) the Court found those procedures sufficient to accommodate First Amendment interests. For instance, the Court concluded that judges could safeguard First Amendment concerns in searches of newspaper offices by assessing the ordinary preconditions for a warrant: probable cause, specificity regarding the places to be searched or objects to be seized, and overall reasonableness.\(^{289}\) As Daniel Solove has argued, the search warrant cases leave open the question as to how the Court would rule in situations where Fourth Amendment procedural protections were unavailable.\(^{290}\)

At first glance, the explanation that other procedural protections existed seems less persuasive with respect to Branzburg, since grand jury investigations do not provide as much protection: the government does not have to demonstrate probable cause or satisfy Fourth Amendment standards to subpoena witnesses.\(^{291}\) Even so, a person still has recourse to a motion to quash or modify a subpoena where “compliance would be unreasonable or oppressive.”\(^{292}\) While the standard by which a court considers such motions is deferential,\(^{293}\) the very prospect of prior resort to a court provides protection that is lacking in the case of FBI or CBP interviews. In addition, despite the wide latitude that grand juries enjoy in investigating potential crimes, the fact that grand juries are convened by courts, not law enforcement agents, and that they

\(^{287}\) Branzburg, 408 U.S. at 701.

\(^{288}\) See, e.g., P.J. Video, 475 U.S. at 873-74; Zurcher, 436 U.S. at 564-65.

\(^{289}\) Zurcher, 436 U.S. at 565.


\(^{291}\) United States v. R. Enters., Inc., 498 U.S. 292, 297 (1991) (holding that a court can quash a grand jury subpoena only where it finds no reasonable possibility that it would produce relevant evidence, but refusing to consider whether the standard would differ in cases involving First Amendment interests).

\(^{292}\) Fed. R. Crim. P. 17(c)(2).

\(^{293}\) See R. Enters., 498 U.S. at 301.
necessitate the seating of sixteen to twenty-three jurors, provide some practical restraints on overreaching that are simply lacking in FBI and CBP interviews. Moreover, the First Amendment claim in Branzburg differs from claims in potential challenges to First Amendment profiling: the reporters in Branzburg alleged a chilling effect on First Amendment activity, but not that the decision to subpoena them involved any improper First Amendment considerations. The Court has not considered whether a different result might be appropriate for situations involving a greater threat that hostility to the speakers or speech in question affected the decision of who to select for questioning. Thus, the Court’s unwillingness to create additional procedural protections in First Amendment challenges to search warrants and grand jury subpoenas does not dictate the outcome in other First Amendment profiling contexts, including FBI and CBP interviews.

A second objection to heightened scrutiny of First Amendment profiling is that even if Supreme Court precedent is distinguishable, courts should not be in a position to second-guess law enforcement judgments in critical terrorism-related investigations. Especially in the terrorism context, courts and critics maintain that law enforcement must be permitted wide discretion to identify and pursue potential threats, and the...

---

295 While lower courts, following Branzburg, have rejected a First Amendment privilege for journalists challenging grand jury subpoenas, some have suggested a higher First Amendment standard might be appropriate for grand jury subpoenas involving greater risk that the government was targeting individuals on account of expressive activities. See, e.g., In re First Nat’l Bank, 701 F.2d 115 (10th Cir. 1983) (granting evidentiary hearing on First Amendment claim); In re Grand Jury Subpoenas Duces Tecum, 78 F.3d 1307, 1312-13 (8th Cir. 1996) (enforcing subpoenas where government demonstrated a compelling interest in information and a sufficient nexus between information sought and subject matter of investigation); In re Grand Jury Proceeding, 842 F.2d 1229, 1236-37 (11th Cir. 1988) (finding compelling interest and substantial relationship, without deciding standard). But see In re Grand Jury 87-3 Subpoena Duces Tecum, 955 F.2d 229, 232 (4th Cir. 1992) (declining to create “substantial relationship” test for subpoenas involving First Amendment claims).
296 Rosenthal also argues that in cases involving “no prohibition or direct cost on the exercise of First Amendment rights,” the Supreme Court as a general matter does not impose heightened scrutiny but simply balances the degree of inhibition against the government interests involved. Rosenthal, supra note 16, at 49. Deferential balancing on account of the indirect nature of the restraint in question may accurately describe the Court’s approach in certain cases, such as Meese v. Keene, 481 U.S. 465 (1987), a case involving a registration requirement applicable to distributors of foreign films. But Rosenthal’s statement offers too sweeping a characterization of the Court’s practice as a whole. For instance, it does not explain Gibson, where the Court rigorously scrutinized government justifications for the investigation of the NAACP and limited two Cold War decisions cited by Rosenthal as examples of deferential balancing. See Rosenthal, supra note 16, at 47-48.

To the extent, however, that this argument rests on the assumption that courts applying heightened scrutiny to First Amendment profiling will routinely second-guess law enforcement decisions, that assumption is unwarranted. Even courts applying heightened scrutiny have sustained law enforcement practices where they credited law enforcement justifications for the necessity of a practice: heightened scrutiny in this context is not “strict in theory, fatal in fact.”\footnote{Tabbaa, of course, sided with CBP’s determination that the security measures employed were necessary. Other judicial doctrines already protect law enforcement officers from second-guessing of reasonable decisions, including the idea that factual determinations of the executive branch are entitled to some deference and the qualified immunity defense in damages cases. The risk that heightened scrutiny for First Amendment profiling will lead to pervasive overturning of executive determinations is remote. Rather, the alternative rule of categorical deference to law enforcement justifications presents the more immediate risk—that even investigations significantly burdening speech or association, or aimed at the suppression of expression, will altogether escape First Amendment scrutiny.}

A third objection to heightened scrutiny is rooted in concerns over practical administrability. Specifically, this concern centers on the idea that law enforcement needs clear, consistent rules when making decisions about whom and how to interview, and that narrow tailoring requirements would force agencies (or even individual agents) to make complicated balancing calculations whenever a First Amendment interest is implicated. The Supreme Court itself has pointed to the need for “readily administrable rules” in law enforcement contexts, especially where decisions are made on the spot. Some courts and commentators have expressed concern that because the...
First Amendment protects such a wide range of expression and association, heightened scrutiny requirements would force investigators to assess innumerable circumstances potentially implicating First Amendment rights.\textsuperscript{303} Such considerations might favor a clear rule holding law enforcement to account only for investigations without a legitimate purpose—and deferential review of such investigations rather than heightened scrutiny.

Indeed, the distinction between overbreadth and permissible First Amendment profiling is contextual and fact-specific, and therefore often complex.\textsuperscript{304} Law enforcement agencies can reduce the uncertainty by adopting more concrete guidelines to instruct officers on questioning decisions involving First Amendment concerns; the border officer charged with interviewing travelers would not be left alone to ruminate, in the heat of the moment, whether an interview would be “narrowly tailored,” but to implement practical agency rules devised in response to constitutional requirements.\textsuperscript{305} Although such guidelines would not eliminate line-drawing questions, they would reduce uncertainty for law enforcement agents. And the judicial doctrines discussed above, such as the qualified immunity defense, already protect officers from second-guessing of reasonable decisions.

More fundamentally, a court’s decision to adopt a deferential bright-line rule to ensure “administrability” is at bottom a judgment that the interest in clarity trumps the harm to individuals potentially subject to abuses.\textsuperscript{306} It is unsurprising that decisions favoring clarity for law enforcement officers over competing liberty concerns often include language discounting the severity of the civil liberties problem at issue. Thus, the Laird Court declared full confidence that the political branches of government would remedy any actual abuses by the military,\textsuperscript{307} Branzburg opined that the press was not vulnerable

\begin{footnotes}
\footnotetext{304}{See supra notes 192-98 and accompanying text.}
\footnotetext{305}{In the border context, for instance, the CBP might impose stricter rules for interviews longer than a certain period of time or for inquiries into religious or political activities, such as requirements of individualized suspicion, advance supervisory approval, and consultation with agency counsel. Moreover, agency directives, such as the one at issue in Tabbaa, do not require spur-of-the-moment decisions by individual officers.}
\footnotetext{306}{See Atwater, 532 U.S. at 366 (O’Connor, J., dissenting).}
\footnotetext{307}{Laird v. Tatum, 408 U.S. 1, 15-16 (1972).}
\end{footnotes}
to abuse, and the Seventh Circuit court that dismantled the Chicago consent decree proclaimed that “[t]he era in which the Red Squad flourished is history.” Courts less sanguine about the tendency of law enforcement agencies to respect individual rights—and more attuned to the real stigmatic costs and chilling effects on communities subject to First Amendment profiling—should rather conclude that meaningful judicial review is a necessary protection against government overreaching.

CONCLUSION

Law enforcement interviews in the terrorism context involve intrusion and stigma that the existing literature has not recognized or explored. As overt and highly personal encounters between individuals in the U.S. Muslim community and law enforcement, interviews can shape individuals’ and communities’ very sense of belonging. Where interviews result from First Amendment profiling, they are especially likely to stigmatize communities and chill First Amendment–protected expression and association. Moreover, such interviews carry a particular risk that improper considerations influenced the scope of the investigation, even where law enforcement did not intend to suppress speech. In light of the suppression and overbreadth concerns in investigations triggered by protected expression, I have argued that courts should apply heightened scrutiny to provide some accountability in FBI and CBP interviews—and that existing First Amendment doctrine provides a foundation for such review.

First Amendment challenges in this context will not be easy for plaintiffs. Before a case even gets to court, individuals must overcome the chilling effects of interviews and submit themselves to the additional public exposure of a lawsuit. That chilling effect conundrum—those who are most chilled by an investigation are least likely to challenge it—will screen out a great many claims at the very outset. Those who do sue must establish standing according to the exceptions courts have carved out after Laird, convince a court to apply heightened scrutiny, and persuade the court not to defer to law

---


309 Alliance to End Repression v. City of Chicago, 237 F.3d 799, 801 (7th Cir. 2001); see also Atwater, 532 U.S. at 353.
enforcement justifications on the facts of their case. They must also overcome additional barriers facing all civil rights plaintiffs, such as the Court’s new “plausibility” pleading standard\textsuperscript{310} and restrictive standing doctrine in claims for injunctive relief.\textsuperscript{311}

Even under these restrictive conditions, some plaintiffs challenging First Amendment profiling in interviews may prevail. Those with the greatest likelihood of succeeding, of course, are those who can establish significant harm with the least persuasive government justifications. Plaintiffs who can demonstrate a combination of stigma and chilling effects plus some further evidence of tangible harm stand the greatest chance. In the case of CBP interviews, lengthy detentions and additional screening measures at the border might suffice to establish such tangible harm. In the case of FBI interviews, such harm might include reputational or economic injuries incurred where law enforcement sought interviews in highly public contexts; where organizations were the target of numerous interviews conducted of other individuals; or where interviewees suffered additional consequences from First Amendment profiling, such as inclusion on a watch list, along with interviews.

In addition, courts will be most persuaded by challenges to interviews where the relationship to actual security threats is so attenuated as to suggest law enforcement decision-making tainted by possible hostility or group-based generalizations—even without a suppression purpose. While the court in Tabbaa declined to find for plaintiffs, the same court might have ruled for plaintiffs had the threat information been even weaker—say, information that suspects might be gathering at Islamic conferences in general without any evidence pertaining to the particular conference in question. Similarly, courts might reject stigmatizing interviews triggered by political letters to the editor, statements in newspapers, or postings on a Facebook page, where there is no suggestion of violence.

Although legal challenges to First Amendment profiling will not be easy, First Amendment freedom of speech and association doctrine offers potential redress for individuals subject to onerous interviews where there is strong evidence of law enforcement overreaching.


\textsuperscript{311} City of Los Angeles v. Lyons, 461 U.S. 95, 105-12 (1983).
Cooperative NRDA & New Governance

GETTING TO RESTORATION IN THE HUDSON RIVER, THE GULF OF MEXICO, AND BEYOND

Michael B. Runnels†
Andrea Giampetro-Meyer‡

INTRODUCTION

In December 2010, General Electric Company (GE) announced that it would proceed to Phase Two of a two-part plan to remove toxic polychlorinated biphenyls, or PCBs, from the Hudson River. The company had released the toxins into a forty-mile stretch of the river over thirty years earlier. After decades of fighting, GE finally decided to use its innovative drive, technical expertise, and economic resources to clean and restore the pollution’s damage. Less than a year earlier, and fifteen hundred miles away, British Petroleum (BP) was making initial preparations to respond to a massive oil spill. Starting in April 2010, the spill yielded eighty-six days of nearly uncontrolled gushing.

† Assistant Professor, Law & Social Responsibility Department, Sellinger School of Business and Management, Loyola University, Maryland.
‡ Professor and Chair, Law & Social Responsibility Department, Sellinger School of Business and Management, Loyola University, Maryland.

2 Id.
3 For recent law review articles that analyze the Deepwater Horizon disaster, see, for example, Oliver A. Houck, Worst Case and the Deepwater Horizon Blowout: There Ought to Be a Law, 24 TUL. ENVTL. L.J. 1 (2010) (considering why three laws—the National Environmental Policy Act (NEPA), the Outer Continental Shelf Leasing Act (OCSLA), and the Oil Pollution Act (OPA)—failed “to prevent and cope with the explosion of the drilling rig, Deepwater Horizon”). Houck highlights “the power of negative thinking,” or worst-case analysis. Id. at 17-18; see also Sam Kalen et al., Lingering Relevance of the Coastal Zone Management Act to Energy Development in Our Nation’s Coastal Waters?, 24 TUL. ENVTL. L.J. 73 (2010) (considering the role of the states “to influence energy development occurring off their coasts” under the Coastal Zone Management Act); Stanley A. Millan, Escaping the “Black Hole” in the Gulf, 24 TUL. ENVTL. L.J. 41 (2010) (“exploring the reaches of the Oil Pollution Act of 1990,” and how this law and state analogues address oil disasters and consequent damages).
This disaster in the Gulf of Mexico made clear that while BP had invested billions into sophisticated drilling technologies,\(^4\) it had failed to invest in the technologies necessary to deal with what ultimately became the largest oil spill in U.S. history—a spill that continues to threaten marine life and adjacent wetlands in the Gulf.\(^5\)

Federal laws strive to hold polluters accountable. In particular, Congress has enacted laws that make polluters liable for the injury, destruction, or loss of natural resources resulting from hazardous substances’ release into the environment. For instance, the Comprehensive Environmental Response, Compensation, and Liability Act\(^6\) (CERCLA, or Superfund\(^7\)) concerns waste sites, and the 1990 Oil Pollution Act\(^8\) (OPA)

---


\(^5\) See generally Carrie Presnall, Laura López-Hoffman & Marc L. Miller, Can the Deepwater Horizon Trust Take Account of Ecosystem Services and Fund Restoration of the Gulf?, 40 ENVTL. L. REP. NEWS & ANALYSIS 11,129 (2010). Presnall, López-Hoffman, and Miller encourage Mr. Feinberg, administrator of the Deepwater Horizon Oil Spill Trust and the Gulf Coast Claims Facility (GCCF), “to account for ecosystem services when assessing . . . harms,” and they describe strategies and mechanisms for assessing harms. Id. at 11,130-31. “Ecosystem services are the benefits humans receive from functioning ecosystems and the species that comprise them.” Id. at 11,130. These services include “seafood, flood control, carbon sequestration, habitat for resident and migrating wildlife, hunting, sport fishing, wildlife watching and other outdoor recreation, a rich local culture, and more.” Id.

\(^6\) CERCLA provides for the cleanup of sites contaminated by hazardous substances. It: (1) authorizes the federal government to clean up sites using the Hazardous Substance Superfund, (2) imposes liability for cleanup on responsible parties, (3) requires responsible parties to perform the cleanup and reimburse others or the fund for the costs of cleanup, and (4) requires responsible parties to pay damages to state and federal governments for injury to natural resources, including compensation for destruction or loss. Comprehensive Environmental Response, Compensation, and Liability Act, 42 U.S.C. §§ 9601-9675 (2006).

\(^7\) For more information on the Superfund, see Kathleen Chandler Schmid, The Depletion of the Superfund and Natural Resource Damages, 16 N.Y.U. ENVTL. L.J. 483 (2008) (explaining how the depletion of the Superfund has impacted CERCLA remediation and describing how some states have created programs to recover natural resource damages).

\(^8\) OPA imposes liability for removal costs and damages resulting from an incident in which “oil is discharged . . . into navigable waters or adjoining shorelines or the exclusive economic zone.” 33 U.S.C. § 2701 (2006). Like CERCLA, OPA establishes liability for damages for injuries, or loss of, natural resources, and outlines liability limits in specific circumstances. Id. §§ 2701-2761.

For an article that explores some of the history of OPA, see Lawrence I. Kiern, Liability, Compensation, and Financial Responsibility Under the Oil Pollution Act of 1990: A Review of the First Decade, 24 TUL. MAR. L.J. 481 (2000) (critiquing OPA, while recognizing the statute’s contribution to modern oil pollution law in the
concerns oil spills. Both acts authorize a form of environmental cleanup and restoration called natural resource damage assessment (NRDA). NRDA is a legal process that determines the type and degree of restoration in which a polluter must engage to compensate the public for environmental pollution’s harm to natural resources.

The NRDA process is inherently adversarial. Companies—often called Responsible Parties (RPs) or Potentially Responsible Parties (PRPs)—represent the interests of their shareholders; Natural Resource Trustees (Trustees) act on behalf of the public. But each of these parties engages in NRDA with an eye toward litigation under the assumption that the courts will ultimately decide on required remedial action and compensation.

Neither CERCLA nor OPA requires cooperative NRDA, a particular approach to NRDA that emphasizes the RPs’ and Trustees’ need to join together and work toward the common goal of restoring natural resources quickly and efficiently—to get to restoration. Three decades of adversarial Hudson River cleanup and restoration efforts evidence a need for cooperative NRDA approaches, the necessity of which is further underscored by the work awaiting BP in the Gulf of Mexico.

Companies are liable for restoring natural resource injuries beyond normal cleanups under Superfund and the Oil Pollution Act. But they can cause their financial exposure to skyrocket by instinctively using legal and scientific defenses to avoid liability. Instead, industry can reduce costs and government trustees can achieve restoration more quickly by joining together in a cooperative natural resource damage assessment.

---

10. Id.
12. Id. at 19.
The purpose of this article is to examine NRDA's current shortcomings and recommend reforms to incentivize cooperative NRDA and broader New Governance principles in the future. This article proceeds in four parts. Part I delineates the article's theoretical underpinnings by placing cooperative NRDA in the context of New Governance. New Governance is a concept that encompasses various contemporary techniques of law and regulation that foster cooperation among public and private actors, including industries, corporate social responsibility (CSR) advocates, and shareholders. Part II considers the facts and relevant NRDA processes of two distinct cases: GE's release of PCBs into the Hudson River and the BP oil spill in the Gulf of Mexico. Part III applies a particular New Governance theory, Professor Edward M. Epstein's model of New Governance, to the GE and BP cases. Epstein's model highlights six modes of social control—law, affinity group regulation, self-regulation, ethical precepts, the media, and an

---


Rather than oppositional, [this approach] aims for an appreciative positive stance, pulling together disparate ingredients and synthesizing elements from opposing schools of thought. Through new governance approaches, contemporary thinkers can bring together in their research unlikely pairs, such as privatization and democratic theory. The theory itself is thus reflexive, in the sense that it calls for integration in legal practice and correspondingly exemplifies hybridization in the academic field. Indeed, the theoretical basis for [this] vision mirrors its practical application in its inclusive spirit.

engaged civil society—that encourage corporations to engage in socially responsible behavior. Applying and analyzing Epstein’s New Governance model yields three recommendations that advance cooperative NRDA and improve corporate behavior beyond the GE and BP cases. In Part IV, this article therefore recommends the following: (1) reforming CERCLA and OPA, (2) reframing Epstein’s model to include science as a mode of social control, and (3) increasing corporate disclosure requirements. In addition to incentivizing more cooperative NRDA, these reforms will contribute to the transparency and accountability the public has come to expect in the post-Enron era.

I. COOPERATIVE NATURAL RESOURCE RESTORATION, IN CONTEXT

This section integrates New Governance theory with cooperative NRDA. Part A considers business reform, in general, and ways to encourage positive corporate behavior. Highlighting Epstein’s theory of New Governance, this section explains movements that advocate CSR and New Governance principles. Part B explains trends in environmental policymaking that mirror business reform, especially those that embody collaborative rather than adversarial approaches to environmental protection. This section then considers NRDA and ultimately integrates the policy of cooperative NRDA with New Governance principles, highlighting trends toward collaboration among stakeholders.

A. Business Reform and New Governance

New Governance has emerged in response to CSR’s inadequacies in incentivizing ethical corporate behavior. Scholars often focus on the CSR movement and how its maxims may be incorporated into the business-decision-making process. CSR advocates encourage corporations to broaden

---

14 Edwin M. Epstein, The Good Company: Rhetoric or Reality? Corporate Social Responsibility and Business Ethics Redux, 44 AM. BUS. L.J. 207, 212-16 (2007) (arguing that CSR is inherently insufficient in achieving the Good Company). Epstein provides his own framework and describes corporations, by virtue of their economic and political power, as the most efficient proxies through which his framework can encourage the actualization of the Good Company. Id. at 210-12.


16 Id. at 463-66; see also George Cheney, Juliet Roper & Steve May, Overview, in THE DEBATE OVER CORPORATE SOCIAL RESPONSIBILITY 3, 4 (Steve May, George
relationships with multiple stakeholders, engage in meaningful and sustained efforts to improve communities, and conform to society’s laws and ethical customs. Increasingly, however, scholars recognize that the CSR movement’s goals are fundamentally flawed. Though some argue that corporations ought to do good for goodness’s sake, several view such duties, if not tethered to the corporate bottom line, as hopelessly naïve. Moreover, critics argue that corporate marketing

Cheney & Juliet Roper eds., 2007) (explaining that conversations about unchecked corporate power are central to conversations about how to “probe in an informed and systematic way the potentials for positive social change in, through, and around the modern corporation”); John M. Conley & Cynthia A. Williams, Engage, Embed, and Embellish: Theory Versus Practice in the Corporate Social Responsibility Movement, 31 J. CORP. L. 1, 37-38 (2005) (describing CSR as “a complex communication network among public and private actors,” which, “[a]t its best, promises a corporate decision making process in which managers think and talk openly about social and environmental issues and then tell the world what they did and why”).

Milton Friedman, The Social Responsibility of Business Is to Increase Its Profits, N.Y. TIMES, (Magazine), Sept. 13, 1970, at 33, reprinted in BUSINESS ETHICS 17 (Tamara L. Roleff ed., 1996) (describing the responsibility of the corporate executive). Friedman argues that this responsibility “is to conduct business in accordance with [the shareholder’s] desires, which generally will be to make as much money as possible while conforming to the basic rules of the society, both those embodied in law and those embodied in ethical custom.” Id. While Friedman articulates this point as a response to the CSR movement, he fails to consider how ethical custom and the law interact. Indeed, he fails to consider that ethical custom and the law are, in fact, interdependent. See Cyrus Mehri, Andrea Giampetro-Meyer & Michael B. Runnels, One Nation, Indivisible: The Use of Diversity Report Cards to Promote Transparency, Accountability and Workplace Fairness, 9 FORDHAM J. CORP. & FIN. L. 395, 407 (2004).

See, e.g., M. Todd Henderson & Anup Malani, Corporate Philanthropy and the Market for Altruism, 109 COLUM. L. REV. 571, 581 (2009) (characterizing the philosophical underpinnings of the CSR movement as based on the view that corporations have a moral duty to do good for others, even at the expense of the bottom line); see also David P. Baron, The Positive Theory of Moral Management, Social Pressure, and Corporate Social Performance 5 (Rock Ctr. for Corporate Governance, Working Paper No. 36, 2006), available at http://ssrn.com/abstract=913808 (arguing that one of the principles underlying the CSR movement is that corporations have an abstract “moral duty” to do good).

See Elizabeth F. Brown, No Good Deed Goes Unpunished: Is There a Need for a Safe Harbor for Aspirational Corporate Codes of Conduct?, 26 YALE L. & POLY REV. 367, 399 (2008) (explaining the reason why certain corporations do not engage in CSR), Brown argues corporate reluctance is partly due to the fact that following CSR principles is more expensive than not, and corporations cannot always pass the costs to the consumer. Id. Moreover, Brown argues that “[p]art of those added costs are the costs associated with increased risk of litigation that corporations adopting codes that embody CSR principles face.” Id.; see also Janet E. Kerr, The Creative Capitalism Spectrum: Evaluating Corporate Social Responsibility Through a Legal Lens, 81 TEMP. L. REV. 831, 839 (2008) (characterizing CSR as profit-centric). Kerr explains that since the effects of CSR on the bottom line have become quantifiable, the law supports, if not requires, corporate managers to “investigate and consider whether CSR can impact the bottom line.” Id. Kerr further argues that a corporate manager who does not consider such linkages—who does not weigh profits as a consideration—could be considered derelict in her duty. Id.
strategists have effectively co-opted the CSR movement as a tool to preserve branding and public image.\textsuperscript{21}

Principles of New Governance can incentivize the responsible behavior society expects. There is no single model of New Governance; rather, several evolving models have been developed and tried in various industries around the globe.\textsuperscript{22} The underlying premise of New Governance is that corporate-governance mechanisms, if they are to be responsive to public expectations of responsible corporate behavior, must have greater flexibilities. Furthermore, those flexibilities ought to be animated by outcomes—not processes.\textsuperscript{23}

New Governance promotes systems that “use innovative, pragmatic, information-based, iterative, and dialogic mechanisms to gather, distill, and leverage industry learning in the service of a still-robust but better designed—that is, more effective and less burdensome—public regulatory mandate.”\textsuperscript{24}

In every respect, deliberation among stakeholder

\begin{footnotesize}
\begin{itemize}
  \item See e.g., S. PRAKASH SETHI, SETTING GLOBAL STANDARDS: GUIDELINES FOR CREATING CODES OF CONDUCT IN MULTINATIONAL CORPORATIONS 45-63 (2003) (regarding the marketing benefits from CSR and the widespread practice of insufficient or inconsistent implementation); Ruth V. Aguilera et al., Putting the S Back in Corporate Social Responsibility: A Multilevel Theory of Social Change in Organizations, 32 ACAD. MGMT. REV. 836, 838 (2007) (arguing that “some companies introduce CSR practices at a superficial level for window-dressing purposes”); Joe W. (Chip) Pitts, III, Corporate Social Responsibility: Current Status and Future Evolution, 6 RUTGERS J. L. & PUB. POL'Y 334, 373-82 (2009) (finding credible the critiques that consider “CSR as, at best, toothless and marketing-oriented, and at worst a malevolent strategy to co-opt or render powerless the critical forces hoping to tame corporations with the more meaningful constraints of law”); Betsy Atkins, Is Corporate Social Responsibility Responsible?, FORBES (Nov. 28, 2006, 12:00 PM), http://www.forbes.com/corporatecitizenship/2006/11/16/leadership-philanthropy-charity-lead-citizen-oa_ba_1128directorship.html (detailing the disingenuousness of corporate CSR campaigns). Atkins writes that “[t]here are practical reasons why corporations should cloak themselves in the politically correct rhetoric of social responsibility. But marketing should not be confused with significant deployments of corporate assets.” Id.; see also Gill, supra note 15, at 462 (arguing that “CSR has become a business-sensitive, if not business-driven practice”). Gill notes that many critics consider the CSR’s original motive to have been effectively subordinated to corporate marketing strategies. Id.
  \item See New Governance, supra note 13.
  \item See supra note 13 and accompanying text. See generally Epstein, supra note 14.
  \item See Ford, supra note 13, at 5 (describing the B.C. (British Columbia, Canada) model as an example of the New Governance). Ford defines the linchpin of this model as a “substantially [altered] relationship between regulators and industry”—a relationship not defined by inflexible regulators mandating rules that are often incompatible with fast-paced business environments, but a relationship defined by a shared responsibility and a pragmatic responsiveness to “complex real-life social systems.” Id. at 27. Ford goes on to describe this New Governance as the “most effective mechanism for making decisions in complex organizational structures.” Id. at 27-28. This, Ford argues, is an “opportunity for dialogic and transparent securities regulation,” viewed from the perspective of industry, regulators, shareholders, stakeholders, and CSR advocates. Id. at 27-28, 60.
\end{itemize}
\end{footnotesize}
groups is central to New Governance approaches. Rather than emphasize checklist-style compliance with prescriptive legal rules—which often incentivizes corporate actors to discover and abuse loopholes—New Governance principles encourage groups to orient themselves in the underlying policy priorities of those rules. In turn, New Governance approaches continually revise both means and ends to solve problems as they arise. 

New Governance approaches also strive to foster transparency and accountability; envision corporate decision making as a collaborative, rather than an adversarial, process; and “provide[] a rational, systemic alternative to draconian rule-making and [its] often adverse effects on business.”

Statute and policy reform (in the environmental arena and elsewhere) have emerged as a result of New Governance principles. For example, British Columbia’s Bill 38 offers a principle-based and outcome-oriented approach to securities regulation. One part of Bill 38 replaces detailed, compliance-based rules governing dealers and advisors with rules arranged under broader standards. These broader standards motivate dealers and advisors to exercise sound judgment, whereas the compliance approach motivated mere adherence to rules that might have been misconstrued.

Political scientists Christopher McGrory Klyza and David Sousa, in their review of American environmental policy, recounted two further examples of collaborative approaches in environmental protection. Quincy Library Group in Northern California brought a variety of participants together—local politicians, environmentalists, and members of the timber industry—to create a logging plan that better supports the

---

25 See supra note 13 and accompanying text.
26 See Ford, supra note 13, at 29 (arguing that while corporations strictly adhering to the letter of the law may appear ideal, such strict adherence, paradoxically, may generate the very moral hazards that undermine corporate governance objectives—through incentivizing the corporate discovery and abuse of regulatory loopholes).
27 Id. at 29-30.
30 Ford, supra note 13, at 17-18.
31 Id. at 19.
Another group of ranchers and environmentalists, the Quivara Coalition, collaborated on an environmentally conscious cattle-ranching plan. These examples illustrate mechanisms that de-emphasize draconian rule-making, which often typifies inflexible regulation. Instead, these examples emphasize cooperation among business, government, and additional stakeholder groups as a means to improve decision making. These examples are typical of the New Governance movement.

In one application of New Governance theory, Epstein considered factors that induce corporations to become “Good Companies”—companies that act as ethical corporate citizens. In particular, Epstein found, six “modes of social control” encourage corporations to engage in socially responsible behavior—law, affinity group regulation, self-regulation, ethical precepts, the media, and an engaged civil society. Epstein’s model provides a practical framework with which scholars can systematically develop—and redevelop—methods to galvanize positive corporate behavior.

In brief, Epstein defined the modes of social control as follows: Law is the articulation of public policy enforced by government. Affinity group regulation refers to standards of behavior established by members of a particular profession.

---

33 Id. at 233-36.
34 Id. at 236-40.
35 While New Governance taxonomies are often contested, a core element in virtually all the theory’s formulations is that private-public associations and networks animated by a series of new regulatory frameworks may achieve social and public good. See generally Colin Scott, Regulation in the Age of Governance: The Rise of the Post-Regulatory State, in THE POLITICS OF REGULATION: INSTITUTIONS AND REGULATORY REFORMS FOR THE AGE OF GOVERNANCE 145 (Jacint Jordana & David Levi-Faur eds., 2004) (exploring theoretical approaches to regulation and providing a foundation for New Governance scholarship); see also Ford, supra note 13, at 28 (conceptualizing a New Governance framework for securities regulation, Ford explains that it would entail a regulatory structure that “spans the so-called public/private divide, pulls industry experience into regulatory decision making, and establishes robust ongoing communication mechanisms (rather than an information-hoarding, adversarial relationship) between industry and regulator”).
36 See Epstein, supra note 14, at 210-13 (arguing that the modern corporate social responsibility movement is inherently insufficient in encouraging ethical corporate behavior; Epstein provides his own framework as the means by which one can incentivize the actualization of the Good Company).
37 Id. at 210-12.
38 Id. at 212-16 (arguing that CSR is inherently insufficient in achieving the Good Company). Epstein provides his own framework and describes corporations, by virtue of their economic and political power, as the most efficient proxies through which his framework can encourage the actualization of the Good Company. Id.
39 Id. at 210.
such as medicine.\textsuperscript{40} Self-regulation is the voluntary adherence to issue-specific standards (such as standards concerning climate change) set by nongovernmental organizations (NGOs).\textsuperscript{41} Companies that self-regulate are expected to comply with standards voluntarily and in good faith.\textsuperscript{42} Ethical precepts are beliefs “derive[d] from religion, humanistic philosophy, social customs, mores, and traditions.”\textsuperscript{43} Ethics often inform or inspire laws.\textsuperscript{44} Vigilant and responsible media respond to corporate malfeasance by reporting material that renders corporate behavior transparent.\textsuperscript{45} Finally, an engaged civil society refers to direct-citizen action through the leveling of pressure on government officials.\textsuperscript{46}

While law has traditionally served as the centerpiece to control corporate behavior, legal and CSR scholars increasingly recognize that corporate malfeasance is highly context-specific. For example, in a given situation some modes may work better than others to encourage socially beneficial corporate behavior.\textsuperscript{47} Epstein’s model will be applied below in Part III.

B. Environmental Policymaking, NRDA, and Cooperative NRDA

The rise of New Governance in business brought with it an attempt to create new cooperative approaches to environmental regulation. Legal scholars have noted that the shift from command-and-control approaches was concomitant with business reform’s shift to New Governance.\textsuperscript{48} Regulators

\begin{flushleft}
\textsuperscript{40} Id. at 210-11.
\textsuperscript{41} Id. at 211.
\textsuperscript{42} Id.
\textsuperscript{43} Id.
\textsuperscript{44} Id. at 211-12.
\textsuperscript{45} Id. at 212.
\textsuperscript{46} Id.
\textsuperscript{47} See generally Runnels, Kennedy & Brown, supra note 28. Epstein lists his modes of social control in descending order of importance: law, affinity group regulation, self-regulation, ethical precepts, the media, and an engaged civil society. Epstein, supra note 14, at 210. In addition to arguing that the modes be used in a prescribed order of importance, he also contends that each mode must be used to galvanize ethical corporate behavior. Id. at 212. Runnels, Kennedy, and Brown discount Epstein’s argument that his modes must be both considered in a precise order and in concert. They suggest instead that context matters in terms of which, and how, the modes are applied. Runnels, Kennedy & Brown, supra note 28, at 513.
\textsuperscript{48} Lisa Blomgren Bingham, The Next Generation of Administrative Law: Building the Legal Infrastructure for Collaborative Governance, 2010 Wis. L. Rev. 297, 300. Blomgren Bingham describes developments in statutory administrative law. She indicates that “[c]ollaborative governance represents an emerging alternative to traditional command-and-control approaches to making, implementing, and enforcing
that utilize command-and-control approaches command particular environmental goals and issue instructions, or controls, detailing how to reach those goals. Collaborative approaches, however, bring stakeholders together to share information and seek solutions that support a both-and agenda.

Similar to a “win-win” situation, a both-and agenda seeks to both achieve environmental goals and factor in economic concerns. Today, a range of environmental policy areas endorse problem solving that is consistent with the both-and policy—collaborative, voluntary, and transparent (e.g., collaborative watershed management and collaborative energy management).

CERCLA and OPA, however, rest on the outdated approaches of yesteryear. Congress passed CERCLA in 1980 at the end of the Environmental Era—the golden era of environmental lawmaking, from 1964 to 1980. Congress passed OPA in 1990 as the shift from command-and-control to collaborative approaches was beginning. Not surprisingly, the
NRDA processes outlined in both statutes and their accompanying regulations exemplify the command-and-control approach.\textsuperscript{57}

Recall that CERCLA outlines legal requirements for cleanup and restoration after releases of hazardous substances at waste sites,\textsuperscript{58} while OPA outlines legal requirements for cleanup and restoration after oil spills.\textsuperscript{59} Both CERCLA and OPA define natural resources broadly—fish, wildlife, drinking water supplies, land, and other such resources that the United States somehow controls (e.g., property it manages or holds in trust).\textsuperscript{60} Though CERCLA and OPA tailor responses to different environmental needs, the NRDA processes under both statutes are similar.

In NRDA, state and federal natural-resource agencies, such as the Department of the Interior (DOI) and the National Oceanic and Atmospheric Administration (NOAA), serve as Trustees. During restoration, the RP replaces the resources or acquires equivalent resources. Additionally, RPs may compensate the public for the interim loss of natural resources contaminated by the incident. Trustees achieve natural-resource restoration by negotiating with RPs or forcing RPs to act through litigation.\textsuperscript{61} Scientists, engineers, economists, and

\textsuperscript{57} Although OPA includes the word cooperative, RPs and Trustees still assume an adversarial relationship in practice under OPA’s NRDA process. See infra notes 218-22 and accompanying text.

\textsuperscript{58} CERCLA was enacted in response to the Love Canal disaster. Basic Information: What Is Superfund?, U.S. ENVTL. PROT. AGENCY, http://www.epa.gov/superfund/about.htm (last visited Sept. 30, 2011). For an article exploring aspects of the complicated CERCLA process, which are beyond the scope of this article, see Jason J. Czarnecki & Adrienne K. Zahner, The Utility of Non-Use Values in Natural Resource Damage Assessments, 32 B.C. ENVTL. AFF. L. REV. 509 (2005). Czarnecki and Zahner urge CERCLA trustees to advance public interest by engaging in proper valuation of natural resource damages. In particular, they urge trustees to include “non-use” values in NRDA. Id. at 512. Non-use value includes “the simple knowledge that something exists . . . , the potential for its use . . . , or the expectation that it will be of use to future generations.” Id. at 511. These attributes are existence, option, and bequest value. Id.

\textsuperscript{59} OPA was enacted in response to the Exxon Valdez oil spill. Oil Pollution Act Overview, U.S. ENVTL. PROT. AGENCY, http://www.epa.gov/osweroe1/content/lawsregs/opaoverview.htm (last visited Sept. 30, 2011). For an interesting article that explores the complex, uncertain NRDA process that followed the Valdez spill, see Sanne Knudsen, A Precautionary Tale: Assessing Ecological Damages After the Exxon Valdez Oil Spill, 7 U. ST. THOMAS L.J. 95 (2009).


\textsuperscript{61} Conner & Gouguet, supra note 11, at 20.
regulatory specialists join Trustees and RPs as they follow the DOI and NOAA methodologies of NRDA. \textsuperscript{62}

The CERCLA NRDA Process includes four phases: (1) preassessment screening, (2) assessment planning, (3) assessment implementation, and (4) postassessment. \textsuperscript{63} During preassessment, Trustees determine whether natural resources have been injured. \textsuperscript{64} In the second stage, the assessment plan, Trustees confirm that pollutants have affected trust resources. They then develop an assessment plan that outlines the procedures Trustees will use to assess damages. \textsuperscript{65} During assessment implementation, Trustees gather data to quantify injuries and determine damages. \textsuperscript{66} Scientists conduct field and laboratory studies to evaluate injuries to natural resources. \textsuperscript{67} Trustees then compare information from injury investigations to baseline conditions in order to develop resource or service loss estimates. \textsuperscript{68} The final phase, postassessment, identifies restoration options and establishes monitoring protocols to ensure the success of the selected restoration projects. \textsuperscript{69}

The OPA NRDA Process includes three phases: (1) preliminary assessment, (2) injury assessment, and (3) restoration implementation. \textsuperscript{70} During preliminary assessment, Trustees determine the impacts of pollutants on natural resources. \textsuperscript{71} Scientists collect time-sensitive data and review scientific literature to determine how a particular hazardous substance has affected trust resources. \textsuperscript{72} If Trustees determine that resources are injured, they proceed to the second phase. \textsuperscript{73}

\textsuperscript{62} NRD Primer, supra note 59 (outlining the details of NRD assessments under both CERCLA and OPA). The Primer outlines the DOI’s methodologies with regard to NRD under CERCLA. It also outlines the NRDA methodologies under OPA.


\textsuperscript{64} Id.

\textsuperscript{65} Id.

\textsuperscript{66} Id.

\textsuperscript{67} Id.

\textsuperscript{68} Id.

\textsuperscript{69} Id.


\textsuperscript{71} 15 C.F.R. § 990.42.

\textsuperscript{72} Mathematical models designed to help predict the fate and effects of the spill on trust resources may also be used. Pre-Assessment: Phase One: Studying the Impacts, NOAA: GULF SPILL RESTORATION: DAMAGE ASSESSMENT, REMEDIATION & RESTORATION PROGRAM, http://www.gulfspillrestoration.noaa.gov/assessment/pre-assessment/ (last visited Feb. 19, 2011).

\textsuperscript{73} Id.
In this phase, injury assessment, Trustees quantify the scope of the injuries and “identify possible restoration projects.”

Trustees rely on both economic and scientific studies to assess the injuries to natural resources and the impact on the public use of those resources. Additionally, Trustees use these studies to develop a restoration plan that details ways to work toward quick recovery of the injured resources and payment to parties who suffered losses due to natural resource injuries. Trustees then evaluate the restoration options and ask the public to comment on a draft of the restoration plan.

The Trustees move on to the final phase of OPA’s NRDA process once the injury assessment is complete. In this phase, restoration, the restoration plan is implemented while NOAA and other Trustees monitor its effectiveness. The Trustees first identify the full spectrum of injuries and determine the optimal restoration procedures. The Trustees, working in tandem with the public, then select and implement the restoration projects. The costs of assessment and restoration are borne by the RPs, who work cooperatively with the Trustees. If the RPs refuse to bear these costs, the Trustees may file a lawsuit or submit a claim of damages to the Oil Spill Liability Trust Fund. The United States may then seek to recover the costs paid out by the fund from the RPs. The restoration that Trustees implement may fall into any or all of the following categories: emergency restoration, primary restoration, compensatory restoration, or early restoration.

---

75 Id.
76 Id.
77 Id.
78 Id.
80 Id.
81 Id.
82 Id.
84 See infra note 88.
85 Restoration: Final Phase, supra note 79. NOAA defines “emergency restoration” to include “actions that are taken by the trustees prior to the completion of the damage assessment and restoration planning process to prevent or reduce..."
NRDA can be both time-consuming and expensive. And despite the seemingly collaborative nature of the process, in the end, RPs and Trustees are poised to litigate, not negotiate. Cooperative NRDA processes, however—where Trustees and RPs voluntarily work together to resolve natural resource liability—reduce costs and increase restoration's efficiency. In cooperative NRDA, Trustees and RPs engage in strategic collaboration to generate accurate information and share that information with an eye toward restoration.

Michael Ammann, a Staff Environmental Scientist at ChevronTexaco Energy Technology Company, explained that cooperative NRDA is more likely when Trustees and RPs develop a commonly held vision of what a successful restoration project would look like. Ammann described an example of a specific pollution incident, an oil spill that occurred in Bay Point, California, when a Chevron pipeline ruptured. Ammann indicated that, when NRDA for the oil spill began, a representative of California's Department of Fish and Game's Office of Spill Prevention and Response suggested a specific restoration project the parties could use. This representative's suggestion laid the foundation for what
successful restoration would look like; the RP and Trustees would work towards a common goal.\textsuperscript{95} Moving forward with the shared vision, the RP (Chevron) and Trustees voluntarily reached agreements regarding injuries and how to respond to future risks.\textsuperscript{96} The parties settled the NRDA in three meetings\textsuperscript{97}: “By focusing on restoration, managing uncertainty, and avoiding unnecessary studies, restoration was achieved faster and all parties realized savings in transactions costs, especially consultant and legal fees.”\textsuperscript{98}

From the opposing perspective of a community organization, Mark Davis, Executive Director of the Coalition to Restore Coastal Louisiana in Baton Rouge, recounted a story of public and private actors who worked together to protect and restore Bayou Trepagnier.\textsuperscript{99} The process involved a coalition of stakeholders, including Shell Oil Company (the RP), state Trustees (including the Louisiana Department of Environmental Quality, or DEQ) and federal Trustees (the National Oceanic and Atmospheric Administration, or NOAA). The parties worked with the Coalition to Restore Coastal Louisiana (which represented local interests) to resolve complex, contentious issues.\textsuperscript{100} In evaluating the NRDA process, Davis wrote, “[W]hat NRDA did was provide a forum and enough compulsion to begin to work issues through.”\textsuperscript{101} The parties reached a settlement, and Davis highlighted the importance of community groups in bringing about the cooperative process.\textsuperscript{102}

\begin{itemize}
\item \textsuperscript{95} Id.
\item \textsuperscript{96} Id.
\item \textsuperscript{97} Id.
\item \textsuperscript{98} Id.
\item \textsuperscript{99} Id. Cooperative NRDA can be highly technical. It includes these two features: (1) restoration-based assessments, and (2) integration of NRDA processes and cleanup. Conner & Gouguet, supra note 11, at 22. Restoration-based assessments allow Trustees and RPs to focus on “restoration of the injured resources rather than on valuation of the lost resources and services.” Id. The focus is on “getting to restoration,” and seeking “the cost of carrying out the restoration as the measure of damages rather than the value of the injured resources and lost services.” Id. Similarly, “prospective restoration” refers to processes that consider restoration at the beginning of an NRDA claim. Prospective restoration allows for some natural resources to be restored, providing ecological services and saving costs, while the formal NRDA process unfolds over the course of several years. Stephen K. Davis, Lawrence D. Malizzi & Nel Yoskin, How Prospective Restoration and Planning Can Be Use in the Settlement of Dredging Natural Resource Damage Cases (unpublished manuscript) (on file with authors).
\item \textsuperscript{100} Mark Davis, Community Organizations Can Make the Process Work, ENVTL. F., May-June 2004, at 25.
\item \textsuperscript{101} Id.
\item \textsuperscript{102} Id.
\end{itemize}
As these examples illustrate, cooperative NRDA provides the collaborative, consensus-based approach that promises to work more effectively than command-and-control approaches. Unfortunately, the cooperative NRDA examples this section has described remain the ideal. The next two cases illustrate adversarial NRDA—the reality.

II. GE, BP, AND NRDA PROCESSES UNDER CERCLA AND OPA

A. GE, PCBs, and the Hudson River

In 1942, GE began using polychlorinated biphenyls, or PCBs, in its capacitators.\footnote{See Frances F. Dunwell, The Hudson: America’s River 301 (2008); see also Austl. & N.Z. Env’t & Conservation Council, Identification of PCB-Containing Capacitors (1997), available at http://www.environment.gov.au/settlements/publications/chemicals/scheduled-wastepubs/pcb.pdf.} As part of the manufacturing process for electrical transformers and capacitors,\footnote{Bayard Webster, E.P.A. Bans Discharge of PCB’s Directly into the Nation’s Waters, N.Y. TIMES, Jan. 20, 1977, at 22. Monsanto Industrial Chemicals Company was, as of 1977, the only American maker of PCBs. PCBs were also used in metal casting plants and in the recycling of wastepaper. Id.} GE discharged PCBs into the Hudson River,\footnote{Id.} among other locations.\footnote{Id.} In 1977, the EPA banned the direct discharge of PCBs into U.S. waters.\footnote{Id.} At that time, the synthetic compounds were described as a “highly toxic...chemical”\footnote{Id.} that resists biological degradation and is, consequently, “one of the most serious of the...environmental contamination problems prevalent today.”\footnote{Id.} The Environmental Defense Fund, a Washington-based environmental law firm, led the actions that challenged PCB discharges and asked the EPA to rule on the proposed ban.\footnote{See Dunwell, supra note 103, at 301. The same firm had also led the actions that yielded the DDT ban in 1972. Webster, supra note 104, at 22.}
By the time the EPA banned the discharge of PCBs into U.S. waters, GE had already released approximately 1.3 million pounds of PCBs into the Hudson. In 1976, commercial fishing in the Hudson River was banned due to the PCB releases. Scientists determined that Hudson River fish ingested much more than the permissible level of the chemical to be safe for human consumption. In particular, scientists believed that consuming fish from the upper Hudson could increase the risk of cancer. GE negotiated a settlement with the State Department of Environmental Conservation wherein GE agreed to cease PCB dumping, pay $3 million toward cleansing the river, and provide $1 million for PCB research. This settlement was significant because, after it, New York citizens no longer had recourse against GE under state law.

In 1980, Congress passed CERCLA, which required GE to clean or neutralize PCBs in the river. In the early 1980s, talk of dredging the river bottom began. GE relied on scientists and lawyers to create and implement strategies to avoid dredging. GE accused the EPA of “shoddy science” and argued that tightening the spigot at or near the polluting plants was sufficient for restoration. GE maintained that sediments from the past were stagnant and no cause for concern. By this point, an environmental group, Scenic Hudson, emerged as the primary advocate for the public regarding the natural resources in and around the Hudson River.

The EPA issued a report detailing the status of the Hudson in 1997. Although GE had managed to stop new PCB leaks, heavy concentrations of PCBs remained at the bottom of Thompson Island Pool, a six-mile stretch of river downstream from GE’s plants in Ford Edward and Hudson Falls. But more

---

111 Revkin, supra note 106.
112 Dunwell, supra note 103, at 312.
115 Dunwell, supra note 103, at 312.
116 Id.
117 Kamienecki, supra note 113, at 144.
118 Id. at 145.
119 PCB War, supra note 114.
120 Id.
121 Id.
122 Id.
importantly, PCBs had not biodegraded or detoxified.\textsuperscript{123} Undegraded, PCBs that were buried for decades entered the food chain where they could biomagnify to toxic concentrations.\textsuperscript{124}

In 2000, after studying PCBs and health hazards, the EPA ordered GE to spend a half-billion dollars over five years to dredge toxic PCBs embedded in the river bottom north of Albany.\textsuperscript{125} In particular, the EPA determined that PCB levels in the upper Hudson’s fish were still unacceptably high.\textsuperscript{126} Carol Browner, the EPA administrator at the time, asserted, “This river needs to be cleaned up. It will not clean itself.”\textsuperscript{127} New York Governor George Pataki endorsed the strategy of forcing GE to dredge and pay for the dredging.\textsuperscript{128}

In protest, GE argued that “someone would have to eat a half-pound of fish a week for forty years to run even a remote cancer risk.”\textsuperscript{129} Moreover, GE issued a reminder of the state’s catch-and-release policy, which required fisherman to return caught fish to the water.\textsuperscript{130} The EPA and environmental groups responded simply that the river should be cleaned.\textsuperscript{131} GE then launched a public relations campaign to argue that dredging the river would “devastate the lives of upriver communities.”\textsuperscript{132} From the mid-1990s to the early 2000s, GE conducted a multimillion-dollar campaign aimed at both politicians and citizens, highlighting the dangers of removing PCBs from the river.\textsuperscript{133} GE also asserted that the dredging would “disrupt river life for a generation.”\textsuperscript{134}

The EPA issued a record of decision in 2002 that laid out a legally binding cleanup plan.\textsuperscript{135} The decision—devised

\begin{thebibliography}{9}

\bibitem{123} Id.
\bibitem{125} \textsc{Dunwell}, supra note 103, at 312.
\bibitem{127} Id.
\bibitem{128} Plausible Plan, supra note 124.
\bibitem{129} Id.
\bibitem{130} Id.
\bibitem{131} See id.
\bibitem{132} Id.
\bibitem{133} \textsc{Kamieniecki}, supra note 113, at 145.
\end{thebibliography}
under the Clinton administration and ratified by President Bush’s EPA administrator, Christine Todd Whitman—was considered “a mix of politics and science.” In April 2002, after exhausting nearly all avenues to appeal the decision, GE changed its position and offered to clean the Hudson River.

In 2005, after years of denying responsibility, GE signed a consent decree committing itself to the removal of PCBs from the Hudson River. The plan called for dredging to begin in 2007, but environmental groups wondered whether GE would continue “a timeworn pattern of grinding delay.” The decree called for two phases of cleanup and bound GE to just the first phase—a year-long, $100 million project to remove the thickest PCB deposits, 2.65 million yards of tainted mud. Phase Two—a five-year, $500 million project—calls for the dredging of the remaining 90 percent of sediment. The second phase covers sediment spread over a much larger, though less heavily contaminated, area.

In 2006, GE submitted a detailed cleanup plan to the EPA and announced that the company could start dredging in 2008 at the earliest. On May 16, 2009, GE began the Phase One dredging. Mile-long freight trains carried the dredged sediment to a hazardous-waste landfill in Texas. The plan requires the company to replace the sediment with uncontaminated soil and native plants. In the meantime, GE continued to fight, this time challenging CERCLA’s constitutionality.

136 Revkin, supra note 106.
137 DUNWELL, supra note 103, at 150. For an article arguing against dredging of the Hudson, see generally Erik Claudio, Comment, How the EPA May Be Selling General Electric Down the River: A Law and Economics Analysis of the $460 Million Hudson River Cleanup Plan, 13 FORDHAM ENVTL. L.J. 409 (2002) (arguing that, under a cost-benefit analysis, dredging is an improper solution to the Hudson River cleanup).
140 Id.
141 Id.
142 Id.
143 Id.; see supra note 1 and accompanying text.
144 Waiting, supra note 135.
145 Revkin, supra note 106.
146 Id.
147 Id.
Like the cleanup process, the NRDA process is also underway. The Hudson River National Resource Trustees are the DOI, NOAA, and New York State Department of Environmental Conservation.\(^{149}\) The EPA coordinates with the Trustees per CERCLA requirements.\(^{150}\)

The Trustees released the Hudson River Preassessment Screen in October 1997\(^ {151}\) and subsequently undertook injury assessment.\(^ {152}\) They have planned or completed numerous studies, including wildlife injury, water- and air-quality injury, and pathway determination (e.g., floodplain evaluation).\(^ {153}\) On September 16, 2002, the Trustees issued the assessment plan for the Hudson River. The plan identifies procedures that the Trustees must use to evaluate injuries to natural resources caused by PCBs.\(^ {154}\) The Trustees are currently in the process of implementing the Assessment Plan.\(^ {155}\) As part of this process, the Trustees continue to engage in studies on biological resources (such as fish, birds, amphibians, reptiles, and mammals) and other natural resources (such as surface water).\(^ {156}\) During the implementation phase, the Trustees engage in ongoing monitoring of actions related to the Hudson cleanup. On May 7,
2010, for example, the Trustees offered support for the EPA’s decision to enforce dredging of the Hudson River.\textsuperscript{157}

B. BP, the Deepwater Horizon Oil Spill, and the Gulf of Mexico

While BP brands itself “Beyond Petroleum,” an environmentally conscientious energy company, its reputation is now stained\textsuperscript{158} by the April 2010 explosion of its Deepwater Horizon drilling rig.\textsuperscript{159} The explosion, forty miles from the ecologically fragile Louisiana coast,\textsuperscript{160} killed eleven people and released millions of barrels of oil into the Gulf of Mexico.\textsuperscript{161} After the explosion, oil gushed from a cracked BP pipeline a mile beneath the ocean’s surface.\textsuperscript{162} At first, BP claimed 1000 barrels of oil were released per day\textsuperscript{163} but later changed its estimate to 5000.\textsuperscript{164} BP maintained that the 5000-barrel figure was accurate\textsuperscript{165} even as outside scientists protested that BP deliberately low-balled the actual figure.\textsuperscript{166} Despite BP’s purported estimates, Congressperson Edward J. Markey, Chair of the House Energy and Environment Subcommittee, sought and secured the release of an internal BP document that revealed BP’s damage analysis and actual estimation—a worst-
case scenario of 100,000 barrels per day. Ultimately, the
government declared the flow rate to be 62,000 barrels a day. To date, the Deepwater Horizon oil spill, beyond question, is the worst oil spill in U.S. history.

Soon after the rig explosion, BP executives faced harsh questioning from Congress regarding the foreseeability of the disaster. “The leaders of the House Committee on Energy and Commerce cited five areas in which the company had made decisions that increased the danger of a catastrophic well”: the well’s design, improper maintenance of the blowout preventer, inadequate preparation and testing of the well casing’s cement job, misleading assurances that the well was properly sealed, and the lackadaisical preparation of a government-mandated oil-spill-response plan.

President Obama appointed a commission to explore the causes of the BP oil spill. The commission confirmed the findings of the House Committee on Energy and Commerce—

---


172 Id.

173 Id.

174 Id.

175 Id.

the disaster was ultimately avoidable. Specifically, the commission’s report found that (1) mistakes, both governmental and private, onshore and on the Deepwater Horizon rig itself, increased the risk of a well blowout; (2) the cumulative risk resulting from these decisions and actions was as unreasonably large as it was foreseeable; and (3) the risk of a catastrophic blowout manifested on April 20 and several of the cited mistakes were contributing factors to the blowout. Ultimately, the disaster resulted from clear mistakes made in the first instance by BP, Halliburton, and Transocean, and, subsequently, by U.S. government officials. These officials, relying heavily on the oil industry’s assertions regarding the safety of their operations, failed to create and apply a program of sufficient regulatory oversight to minimize the risks associated with deepwater drilling.

By mid-July 2010, BP’s incurred costs reached nearly $4 billion. Under pressure from the Obama administration, BP, to compensate victims of the spill, pledged an additional $20 billion to an escrow account—an account administered by a BP-appointed and Obama-administration-approved arbiter. As late as February 2011, however, BP objected to the settlement terms between the arbiter and victims, claiming that the terms were too generous. Regardless of these costs, the ultimate cost to restore the natural resources in the Gulf will be far higher.

In addition to the cleanup process, the NRDA process has begun. NOAA coordinates with the Trustees per OPA

178 Id. at 115.
179 Id. at 127.
185 See generally FRAMEWORK FOR EARLY RESTORATION ADDRESSING INJURIES RESULTING FROM THE DEEPWATER HORIZON OIL SPILL (2011), available at
requirements. The Deepwater Horizon Trustees include the U.S. Fish and Wildlife Service, the National Park Service, the DOI, the Department of Commerce, the Bureau of Land Management, and designated state trustee agencies for the states of Alabama, Florida, Louisiana, Mississippi, and Texas. As of February 2011, NOAA is engaged in the preliminary assessment phase. While both Trustees and RPs provide oil-exposure data to the public, which hastens the restoration process, “they reserve the right to withhold information from [those] studies” in which either party contests the results. Withholding information is clearly inconsistent with standard scientific investigation. The practice of withholding


Matthew P. Coglianese, The Importance of Determining Potential Chronic Natural Resource Damages from the Deepwater Horizon Accident, 40 ENVTL. L. REP. NEWS & ANALYSIS 11,100, 11,101 (2010). Coglianese points out that “efforts to understand chronic, sublethal, and cumulative effects have only begun.” Id. at 11,100. He introduces the legal and regulatory framework that will allow the government to hold parties responsible for the natural resource damage in the Gulf of Mexico. Id. With regard to the current federal assessment, and the Damage Assessment, Remediation, and Restoration Program (DAARP), “[i]ndependent researchers have claimed that the ‘big money’ for NRDA research, access to information, and information dissemination are controlled too tightly by the federal government and by BP.” Id. at 11,104. One researcher stated:

The problem is that researchers for BP and the government are kept quiet, and their data is unavailable to the rest of the community. When damages to the Gulf are assessed in court or Congress, there might not be enough objective data to make a fair judgment. Transparency is vital to successful science: researchers must subject their proposals to the scrutiny of colleagues, and publications require peer review.


Petersen, supra note 190 (Noting the importance of the withheld scientific data, Tom Brosnan, an environmental scientist with NOAA, argues that “[t]his is a
information stems from the myriad lawsuits the disaster prompted—much of the data will provide material support for the Trustees’ and RPs’ legal strategies. These legal strategies hinge on the retention of scientific expert witnesses (many of whom also engage in the NRDA process), and result in the systematic withholding of essential information. Only after the NRDA restoration plan is approved, or litigation is exhausted, will the parties disclose oil-exposure data. Paradoxically, “[m]itigating the long term impact of the oil spill . . . require[s] an open exchange of scientific data and analysis.” Keeping such data confidential will likely result in delaying Gulf restoration. As the BP case illustrates, in the Environmental Era that is supposed to be collaborative, any move toward cooperation is strained, at best.

III. Epstein’s New Governance Model, GE, and BP

This section applies Epstein’s modes of social control to the GE and BP cases. The analysis in this section lays the groundwork for the recommendations in Part IV, which support the narrow goal of incentivizing cooperative NRDA and affirming broader New Governance principles, especially increased transparency and accountability.

In an open and democratic society, an engaged civil society, working in tandem with a vigilant and responsible media, is critical in highlighting unethical corporate behavior and spurring change. These two modes are essential forms of social control in both the GE and BP cases. In the GE case, environmentalists’ engagement has pressured government

very pointed investigation into what has been injured, what has been lost and what is required to compensate the public.”


193 See Gagosian & D’Elia, supra note 192. Stan Stenner, the director of conservation science at the Ocean Conservancy, noted the typicality of hiring experts following an ecological disaster. He explained that this is “par for the course . . . anytime you have an event like this, everyone goes out and recruits experts.” Petersen, supra note 190. Noting the 1989 Exxon Valdez oil spill in Alaska and Exxon’s aggressive campaign to hire experts, Senner offered that the expert’s “mission was not to find out what the harm was from the spill; their mission was to cast doubt on any conclusions drawn about harm from the spill.” Id.

194 Gagosian & D’Elia, supra note 192.

195 Petersen, supra note 190.

196 Gagosian & D’Elia, supra note 192.
actors to force GE’s cleanup and restoration. Scenic Hudson, Friends of a Clean Hudson, New York Public Interest Research Group, The Riverkeepers, and the local chapter of the Sierra Club have been especially tenacious.

The media, including the New York Times, have exhibited both vigilance and responsibility in providing information. Reporters exposed deficiencies in GE’s cleanup and restoration efforts. Journalists have also been willing to highlight GE’s illegal and unethical behavior, including GE’s continually obstinate and adversarial behavior. For example, GE launched a public relations campaign arguing “that the risk of leaving PCBs in the silt was very low, while the cost of dredging the contaminated bottom would be very high, possibly costing as much as several hundred million dollars.” This media campaign garnered some support, which meant that some members of an engaged, albeit misinformed, civil society stood with GE and wanted to prevent dredging.

Regarding BP, an engaged conglomeration of regional elected officials, concerned citizens, environmentalists, and myriad others spoke through a vigilant media to pressure BP and government actors into mounting a transparent and sustained cleanup effort. For example, concerned by a “disinformation campaign” BP waged to underestimate the impacts of the spill and the United States’ lethargic response to it, James Carville, a former strategist for President Clinton

---

197 See supra text accompanying note 121.
198 KAMIECKI, supra note 113, at 147.
199 Epstein, supra note 14.
200 KAMIECKI, supra note 113, at 147.

[w]hat [Nungesser] brings to the table is something visceral and raw and brave and at times unhinged.... He has an umpire’s skill of calling balls and strikes—he calls it the way he sees it. He's able to go after BP and the Obama administration with equal fury. People are counting on him to be the last uncompromised man in Louisiana.

Id.
203 Id.
and current Gulf resident,\footnote{James Carville was chief strategist for Bill Clinton’s 1992 presidential campaign. Carville is a resident of New Orleans, Louisiana, where he teaches political science at Tulane University. Press Release, Tulane University, James Carville Joins Faculty (Nov. 8, 2008), available at http://admission.tulane.edu/livecontent/news/28-james-carvillejoins-faculty.html; see also Peter Baker, These Days, Carville Praises Oil Response, \textit{N.Y. Times Caucus Blog} (Aug. 6, 2010, 1:27 PM), http://thecaucus.blogs.nytimes.com/2010/08/06/these-days-carville-praises-oil-spill-response/ (“Mr. Carville’s exasperated criticism proved devastating to the White House, letting loose a storm of criticism from other liberals and emboldening conservatives. That particularly pained the White House, where Mr. Carville has many friends from the Clinton campaigns and presidency, most notably, of course, the White House chief of staff, Rahm Emanuel.”).} famously challenged the Obama administration on \textit{Good Morning America}. He stated: “Man, you got to get down here and take control of this! Put somebody in charge of this thing and get this moving. We’re about to die down here!”\footnote{Jake Tapper & Huma Khan, ‘Political Stupidity’: Democrat James Carville Slams Obama’s Response to BP Oil Spill, \textit{ABC News} (May 26, 2010), http://abcnews.go.com/GMA/Politics/bp-oil-spill-political-headache-obama-democrats-slam/story?id=10746519.} When making one of several appearances on \textit{Anderson Cooper 360}, Billy Nungesser, the combative Parish President from Louisiana,\footnote{See Robertson, supra note 202 (characterizing the combative style of Nungesser and describing him as “pugnacious”).} illustrated the mutually reinforcing nature of media and society. Recounting a conversation with President Obama, Mr. Nungesser indicated that the White House took note of Mr. Cooper’s vigilant presence and broadcasts in the Gulf. Mr. Nungesser stated, “[The President] made me commit and I agreed that, if we have the same mess-up in chain of command, or things not getting done, that I will give him a call at the White House before I call you, Anderson.”\footnote{Brian Stelter, \textit{Cooper Becomes Loud Voice for Gulf Residents}, \textit{N.Y. Times}, June 18, 2010, at A19, available at http://www.nytimes.com/2010/06/18/us/18cooper.html.} Weeks later, Mr. Carville wrote an op-ed piece noting the much-improved and vigorous government response:

\begin{quote}

We need our government to remain vigilant in addressing this. We need a lot of research into the science of the effects of the spill. And in the words of Interior Secretary Ken Salazar, we need to continue to have the heel of our boot on the neck of BP. . . . [W]e need to stay vigilant and aggressive in being sure that the inevitable “It’s time to move on” mentality does not set in. Trust me. The last thing we need to do is move on until our precious coastline is both restored and renewed.

\end{quote}
Mr. Carville also commended the President’s successful negotiation with BP that resulted in a $20 billion fund to compensate the spill’s victims.  

Accepting Mr. Carville’s overture, the White House sent a copy of Mr. Carville’s essay to reporters. To this day, an engaged civil society and a vigilant and responsible media, working in tandem, continue to keep Gulf cleanup and restoration issues in the limelight.

Law, the articulation of public policy enforced by government,212 provided the foundation for environmentalists’ and journalists’ work in both the GE and BP cases. CERCLA, as a mode of social control, was designed to address just the kind of scenario the GE case presents: to hold RPs accountable for cleaning hazardous-waste areas and restoring natural resources.214 But law is complex. GE’s adversarial stance on cleanup and restoration is consistent with the incentives built into CERCLA’s design.215 CERCLA, at its foundation, assumes an adversarial relationship between Trustees and RPs. Moreover, GE used law to delay its response to the PCB discharges and to challenge the constitutionality of CERCLA.216 Though that litigation strategy ultimately ended in failure in 2010,217 legal mechanisms provided GE with the opportunity to both deny its responsibility for polluting the Hudson River and then defend its pollution for decades.

Similarly, law constitutes a necessary foundation to restore the Gulf. Like CERCLA, Congress tailored OPA to respond to the precise scenario the Deepwater Horizon incident presents.218 Although OPA, in contrast to CERCLA, incorporates “cooperation” in its text, the specter of litigation is real. Consequently, restoration processes under OPA are best described as adversarial. As mentioned, the RPs and Trustees are now locked in a “battle of the experts”219 whereby both sides

---

210 Id.
211 See id.
212 See Epstein, supra note 14, at 210.
213 See supra Part I.B.
214 See supra Part I.B.
215 See supra notes 129-37 and accompanying text.
216 See supra note 148 and accompanying text.
217 See supra note 148 and accompanying text.
will rely on experts to prove their case. Trustees will attempt to maximize financial damages and RPs will attempt to minimize them. This predictable dynamic, which the systematic withholding of scientific data compounds, all but eviscerates the cooperative elements of OPA. Counterintuitively, these “cooperative” efforts delay Gulf restoration. Legislative changes to OPA that incentivize genuine cooperation may change the dynamic between RPs and Trustees.

Epstein’s three remaining modes—affinity group regulation, self-regulation, and ethical precepts—did little to harness responsible corporate behavior from GE or BP. Recall that affinity group regulation refers to standards of behavior established by members of a particular profession, such as medicine. The case studies noted above evoke no sign of professional influence. In the GE case, managers developed a stonewalling strategy. In the BP case, managers decided how the company would engage with Trustees and the public. Management is a practice, not a profession. Management has not yet formally recognized a duty to serve the greater good, nor has management adopted an ethics code. Not surprisingly, then, managers in the two industries considered here, electrical manufacturing and oil, have not developed professional codes to articulate how their work will consider the stakeholders’ needs.

Self-regulation is the voluntary adherence to standards set by NGOs concerned with specific issues, such as climate change. In the GE case, consider a scenario where Scenic Hudson had created standards for hazardous-waste disposal and GE had voluntarily complied with those standards in good faith. This would have been an act of self-regulation. In the real scenario, though, environmentalists seemed to know they

---

220 Kamieniecki, supra note 113, at 147 (describing the incentives at play when litigation looms).
221 Id.
222 Id.
223 See Epstein, supra note 14, at 210-11.
224 See supra text accompanying notes 118-20.
225 Henry Mintzberg, The MBA Menace, FASTCOMPANY.COM (June 1, 2004), http://www.fastcompany.com/magazine/83/mbamenace.html. Arguing that “no one can become a manager in the classroom,” Mintzberg notes that management is a “craft” tempered by experience. Id.
226 Thomas Kostigen, The Business Oath: Commentary—Let’s Strive for a More Ethical 2011, MARKETWATCH (Dec. 31, 2010), http://www.marketwatch.com/story/heres-to-more-ethical-business-practices-2011-12-31. Professions such as law and medicine have recognized their duty to contribute to society. Moreover, they have adopted and enforced codes of conduct. Id.
227 See Epstein, supra note 14, at 211.
were in for a long fight from the start.\textsuperscript{228} They did not seek voluntary compliance with guidelines.

“Ethics” highlights beliefs derived from religion, humanistic philosophy, social customs, mores, and traditions.\textsuperscript{229} Ethical precepts can harness excellence in corporate behavior. Today, for instance, GE’s Ecomagination campaign inspires “the brightest minds to collaborate, invest and innovate” in clean energy.\textsuperscript{230} Although business interests necessarily provide the primary foundation for the campaign, GE’s statements about clean energy suggest that the company values the environment, and that GE has an obligation to consider the needs of future generations.\textsuperscript{231} These statements, at a minimum, reflect GE’s ethical awareness. Unfortunately, GE’s Ecomagination is entirely forward-looking: the plan disregards the Hudson River disaster.\textsuperscript{232} In fact, GE’s restoration strategy is decidedly unimaginative. From CERCLA’s passage, in 1980, to 2002, GE stonewalled and continually avoided action that would restore the natural resources the corporation damaged or destroyed.

Ironically, during the stonewalling years, Jack Welch’s rationalizations for stall tactics were peppered with words that hint at an awareness of CSR principles. In essence, he asserted that GE was “doing the right thing” by doing nothing. Welch stood his ground against natural resource restoration by stating that he wanted the “truth” to win out.\textsuperscript{233} Regrettably, Welch’s truth was that GE had a right to refrain from restoration because PCBs were not harmful.\textsuperscript{234} After Welch retired in 2001, GE’s new CEO, Jeffrey R. Immelt, changed strategies and

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{228} See supra notes 116-34 and accompanying text.
\item \textsuperscript{229} See Epstein, supra note 14, at 211-12.
\item \textsuperscript{231} Id.
\item \textsuperscript{232} Although the company did agree to go ahead with Phase 2.
\item \textsuperscript{233} \textit{Jack Welch & John A. Byrne, Jack: Straight from the Gut} 283 (2003) (“But I take great pride...our people get up every morning all over the world and compete like hell with absolute integrity....[Our people] see no conflict between taking on the world’s best, every day, all over the globe, giving 110 percent and more—to compete and win and grow—and at the same time maintain an instinctive, unbendable, commitment to absolute integrity in everything we do.”).
\item \textsuperscript{234} Id. at 283-94. Welch noted:

\begin{quote}
Nothing is more important than a company’s integrity....It not only means that people must abide by the letter and spirit of the law, it also means doing the right thing and fighting for what you believe is right....On PCBs, we've assured ourselves that they are not harmful to our employees or our neighbors.
\end{quote}

Id. at 284.
\end{itemize}
\end{footnotesize}
moved towards ending the dispute. Perhaps as a more neutral outsider, Immelt could see that the stonewalling would have to end because it became clear that PCBs were, in fact, harmful.

Self-regulation and ethical precepts were similarly unhelpful modes of social control in BP’s case. Certain company actions do suggest that BP has at times had self-regulation or ethical precepts in mind. For example, under former BP CEO John Browne, the company unveiled a new motto, “Beyond Petroleum.” The motto accompanies an “insignia of a blooming flower,” an image meant “to portray the company as one . . . responsive to growing public concerns [regarding] climate change.”

Unfortunately, though, BP’s new motto and logo seem more about marketing than responsible behavior. In reality, BP “has a worse health, environment[,] and safety record than many other major oil companies”; the Deepwater Horizon oil spill is only the latest costly blunder in a larger series. The company does not demonstrate the proactive stance inherent in corporations that self-regulate and act ethically. Indeed, despite a catalog of crises and near catastrophes in recent years, BP demonstrates a chronic inability or unwillingness to learn from its mistakes. Three incidents are especially significant.

In 2005, an explosion at a BP refinery in Texas City, Texas, killed fourteen workers and injured many more. In its investigation, the government discovered more than 300 safety

---

235 Revkin, supra note 106.
236 GE demonstrated an escalating commitment to a losing course of action. It is possible GE decided to stick with its stance on PCBs because acknowledging that PCBs are harmful would have made Welch’s initial decision seem incorrect or bad. See LINDA K. TREVINO & KATHERINE A. NELSON, MANAGING BUSINESS ETHICS: STRAIGHT TALK ABOUT HOW TO DO IT RIGHT 91 (5th ed. 2011) (discussing decision makers’ tendency to consider sunk costs in determining whether to escalate a commitment).
237 See Krauss, supra note 158.
239 Robbie Brown, Panel Presses BP on Its Safety Record, N.Y. TIMES, Aug. 27, 2010, at A13, available at http://www.nytimes.com/2010/08/27/us/27hearings.html (Noting the serial safety violations in BP’s history, Capt. Hung Nguyen of the Coast Guard, who is also part of a team of federal investigators questioning BP’s record, commented to BP officials that “one dot is a point, two dots in a line, and three dots is a trend. . . . There’s a trend there about the safety culture of BP. These things keep happening.”)
and BP agreed to pay $21 million in fines, which, at the time, was an industry record. Telas Group, a consulting firm contracted to examine conditions at the facility, reported that they “had never seen a site where the notion ‘I could die today’ was so real.” After inspectors from the Occupational Safety and Health Administration (OSHA) revisited the Texas City facility in 2009, they discovered more than seven hundred safety violations and proposed a record fine of $87 million. OSHA stated that many of the penalties stemmed from BP’s failure to meet its responsibilities under the previous Texas City settlement. Ultimately, in August 2010, BP agreed to pay $50 million to settle penalties for its failure to correct safety issues between 2005 and 2009. This was another record fine for the industry.

In another incident, Thunder Horse, a platform in the deepwater Gulf of Mexico, was vulnerable when Hurricane Dennis passed over the platform in 2005. Thunder Horse, a $1 billion crowning glory in deepwater-drilling technology, listed (or tilted) precariously to one side and appeared to be sinking. Investigations later revealed that a backwards-installed valve caused the vessel to flood, imperiling the project before any oil was pumped.

Finally, in 2006, a cracked BP oil pipeline in Alaska forced one of the nation’s largest oil fields, Prudhoe Bay, to shut down. BP was subsequently fined $20 million after prosecutors demonstrated BP’s negligent maintenance of the pipeline. Prudhoe Bay Oil Field remains vulnerable to an

242 Id.
243 See Lyall, supra note 238.
245 Id.
247 Id.
248 See Lyall, supra note 238.
249 Id.
250 Id.
251 See Krauss, supra note 158.
accident that industry insiders believe could rival the Deepwater Horizon spill.\footnote{252}

Given BP’s continuing struggle to comply with current laws, there remains little to suggest that the company is capable of adopting any heightened standard, whether derived from NGOs or affinity groups (which have yet to exist). Moreover, BP’s accident history makes unlikely the possibility that the company will avail itself of any argument that it ought to behave ethically. To date, Epstein’s final three modes have proven ineffective at incentivizing BP to engage in responsible corporate behavior.\footnote{253}

As we think beyond GE and BP, it is useful to consider what typically incentivizes positive corporate behavior.\footnote{254} Generally, corporations with an international presence tend to prefer reputation-saving self-regulation to government intervention.\footnote{255} GE and BP are exceptions to this generalization. When thinking about Epstein’s modes of social control and New Governance in general, these principles are best considered in the context of a particular company’s history—its record of success and failure, and its past responses to failure, in particular.\footnote{256} Although both GE and BP have resisted government intervention and made attempts to create a positive environmental reputation, any self-imposed strategies have lacked substance. Further, for both companies, the stakes of owning up to real problems are high. GE has several Superfund sites waiting to be restored.\footnote{257} Any self-imposed regulation obligates the company to continue cleaning and restoring with no limit to the corporate spending necessary to become a solid environmental citizen.\footnote{258} BP has already


\footnote{253} It is possible the ethics and self-regulation modes will work in the years ahead.

\footnote{254} See supra note 47.


\footnote{256} Some companies can face an environmental disaster and it can serve as a wake-up call to change the organization’s culture.


\footnote{258} According to the EPA, GE is an RP in 52 Superfund sites in the country. Revkin, supra note 106.
invested so much in one particular strategy—aggressive cost cutting combined with aggressive moves to increase market share by capitalizing on its expertise in deepwater drilling—that even the Deepwater Horizon disaster is unlikely to inspire a major shift in corporate culture.

IV. GETTING TO RESTORATION: LOOKING BEYOND THE HUDSON AND GULF

Ideally, GE would have voluntarily assumed responsibility to remedy natural-resource injuries in the Hudson River. If GE voluntary engaged in cooperative NRDA, this behavior would have demonstrated a commitment to New Governance principles and, moreover, demonstrated corporate integrity. GE needed the courage of its convictions in sustainable value creation. The company waited too long to demonstrate a commitment to using its “Ecomagination” and resources to efficiently and sustainably restore the Hudson River. Similarly, BP is poised for adversarial, rather than cooperative, NRDA. The significance of BP’s decision to embrace adversarial strategies is that doing so prevents Trustees and RPs from getting to restoration. BP still has the power to demonstrate corporate integrity and a commitment to New Governance principles. Modeling a new approach would set an example for future work in responding to environmental disasters.

---

259 Integrity has these three characteristics: (1) One must take pains to try to discern what is right or wrong; (2) One must be willing to shape one’s actions in accord with that discernment, even when it is difficult or painful to do so; and (3) One must be willing to acknowledge publicly what one is doing. STEPHEN L. CARTER, INTEGRITY 7 (1996).

260 For an excellent analysis of culture and accidents in the context of high risk technology, see DIANE VAUGHAN, THE CHALLENGER LAUNCH DECISION: RISKY TECHNOLOGY, CULTURE, AND DEVIANCE AT NASA (1996). Vaughan’s work is an appropriate lens through which to consider reviews of the Deepwater Horizon disaster. See NATIONAL COMMISSION, supra note 177. Also, CCRM Deepwater Horizon Study Group concluded that “those who worked on the Deepwater Horizon Macondo well project [failed to make] conscious ‘well informed’ decisions to trade safety for money . . . . [T]here were perceived to be no downsides associated with the uncertain thing.” Letter from Robert Bea, Professor, Ctr. for Catastrophic Risk Mgmt. Deepwater Horizon Study Grp., to Nat’l Comm’n on the BP Deepwater Horizon Oil Spill & Offshore Drilling 1 (Nov. 24, 2010), available at http://www.eoearth.org/article/Deepwater_Horizon_study_group?topic=50364. The Study Group explains that [T]he Macondo well permitting documentation clearly shows that both BP and the MMS believed the likelihood of a catastrophic blowout were not significant . . . . [A]n organization’s safety culture takes time (several decades) to develop and has to be grown from within . . . . [A]t the time of the Macondo blowout, BP’s corporate culture remained one that was embedded in risk taking and cost-cutting . . . . Cultural influences that permeate an organization and an industry and manifest in actions that can either promote and nurture a high
The following section provides three concrete recommendations to make cooperative NRDA more likely. The first addresses cooperative NRDA directly. The second and third suggestions affirm broader New Governance principles. All three recommendations, however, rely on New Governance theory.

A. Recommendation One: Amend CERCLA and OPA to Incentivize Cooperation

Scientists and policymakers who advocate cooperative NRDA generally see OPA’s statutory language as superior to CERCLA’s because OPA incorporates the concept of “cooperation” in more than one provision of the statute.\(^{261}\) Fundamentally, however, both CERCLA and OPA assume that decisions about NRDA and restoration will likely be made through, or as a consequence of, litigation. As mentioned, while both Trustees and RPs provide scientific data to the public, they reserve the right to withhold information from studies in which either party contests the results.\(^{262}\) This practice results in the systematic withholding of data until litigation is exhausted. Consequently, restoration is delayed.\(^{263}\) Since NRDA takes place in the context of litigation as the default dispute-settling mechanism,\(^{264}\) the end game is still defined by victory for the opposing party, no matter the gloss of civility that opposing sides create during the NRDA process. That is, the cleanup and restoration of injured natural resources is not the primary motivation: attorneys dominate and control the NRDA

---

142 BROOKLYN LAW REVIEW [Vol. 77:1

Id. app. A at 7 & 9.

Insularity is an additional culture problem. “[I]ll-advised [corporate] strategies are often the result of a company dialogue restricted to a narrow group of individuals who confer only with each other.” Peter Firestein, Insularity: The Hidden Killer of Corporate Reputation: BP Has Paid Dearly for Failure to See Its Actions in the Context of Broad Social Interests, BLOOMBERG BUS. WK. (Sept. 7, 2010, 3:51 PM), http://www.businessweek.com/managing/content/sep2010/ca2010092_593603.htm.

For example, OPA mandates citizens’ councils for Prince William Sound and Cook Inlet. These councils are designed to promote partnership and cooperation among local citizens, industry, and government. See Introduction, PRINCE WILLIAM SOUND REG’L CITIZENS’ ADVISORY COUNCIL, http://www.pwsrcac.org/about/index.html (last updated July 14, 2011).

See supra note 193 and accompanying text.

See supra note 193 and accompanying text.

Partisan advocates are supposed to be diligent, productive contributors to finding the truth. However, scholars who study the costs and benefits of an adversarial system have established defects in the truth-finding process; most notably, lawyers hell-bent on victory are not incentivized to aid the court in discovering all the facts.

Attorneys influence both Trustees and RPs, which means that both sides are incentivized to consider factors other than reaching restoration. Certainly, RPs are concerned with costs, and are therefore encouraged to use adversarial flaws to gain the lowest judgment possible. But Trustees are also equally vulnerable to flaws in the adversarial system. For example, Trustees are not clearly bound by fiduciary duties to protect natural resources even if that means revealing the complete truth (e.g., that some natural resources have not, in fact, been harmed by a hazardous release). Given the disincentives of seeking justice through the adversarial system, there is no wonder that critics of the current NRDA process blame its defects on dueling lawyers and their expert scientists, who demonstrate considerable skill at generating evidence with a particular objective in mind—a favorable outcome for the client.

One significant change to both CERCLA and OPA would set the stage to get to restoration faster and more efficiently. Congress, or the EPA in its regulatory capacity, should move the locus of control in evidence-gathering to a judge or other neutral party, rather than an attorney. Under this system, a

---

265 See Nicholas J. Lund & Niki L. Pace, Deepwater Horizon Natural Resource Damages Assessment: Where Does the Money Go?, 16 OCEAN & COASTAL L. J. 327, 351-53 (2011) (describing how lawyers are involved in the NRDA process (e.g., Trustees often retain outside legal counsel)).

266 See Strier, supra note 264, at 482.

267 In fact, Trustees and corporate executives are likely to be attorneys.


269 Strier, supra note 264, at 466.

270 Another neutral party would be an arbitrator. Alternative dispute resolution (ADR) is another avenue that could lead to more effective NRDA. See Sarah L. Inderbitzin, Nicholas Targ, Jamies L. Byrnes & Bruce A. Johnson, The Use of Alternative Dispute Resolution in Natural Resource Damage Assessments, 70 WM. & MARY ENVTL. L. & POL'Y REV. 1, 28-29, 31 (1995). Inderbitzin, Targ, Byrnes, and Johnson urge agencies to use alternative dispute resolution, or ADR, when conducting assessments, especially to “reduce the amount of money spent on data collection, increase the data’s acceptability, and reduce litigation.” Id. at 28. These scholars emphasize the value of non-adversarial approaches to NRDA.

Other authors have suggested statutory change. For example, Klyza and Sousa state generally, that “[w]ithout statutory changes to protect . . . collaborative experiments, they will often be vulnerable in a political system that offers many points of access, many points of attack.” KLYZA & SOUSA, supra note 32, at 8.
judge or other neutral party would oversee (1) the selection of an independent panel of scientists to engage in NRDA and (2) the approval of a defined budget for the NRDA process. In other words, from the beginning, an objective party would set the parameters of the NRDA process. Shifting control from dueling attorneys to a process with judicial oversight supports the classic both-and agenda: it would achieve substantive improvements in environmental protection and accommodate all stakeholders’ legitimate concerns in the economic and social costs of implementing both CERCLA and OPA.  

B. Recommendation Two: Consider Science as a Mode of Social Control in New Governance Frameworks

Science, similar to vigilant and responsible media, could serve as a mode of social control to encourage positive corporate behavior. As Jane Lubchenco has stated, “[O]ne of the most important roles of science is to inform, to provide information, so that decision makers can take that information into consideration and understand the full ramifications of a course of action.” Science as a mode of social control, used alone or in tandem with other modes, could incent ethical behavior by offering factual truth—truth that informs the public about the consequences of corporate behavior. Considering science as a mode of social control is consistent with the historical view of science as a producer of reliable knowledge.

The role of science in society historically focused on the production of “reliable knowledge” toward understanding the world and solving practical problems. According to Lubchenco, the relationship between science and society is predicated upon an unwritten social contract—a commitment that scientists will not only create new knowledge but also communicate knowledge broadly so citizens and policymakers

---

271 See Patrick E. Tolan, Jr., Natural Resource Damages Under CERCLA: Failures, Lessons Learned, and Alternatives, 38 N.M. L. REV. 409 (2008) (exploring NRD litigation, alternatives to litigation, and corrective action that would allow NRD to realize its potential). Tolan sees tremendous promise in the idea of cooperative NRDA, but suggests that regulatory and legislative changes are necessary to set the stage for increased cooperation. Id. at 452.


273 Id. Dr. Jane Lubchenco currently serves as the Under Secretary for Oceans and Atmosphere and Administrator of NOAA.

can use the knowledge to make informed decisions.\textsuperscript{275} Importantly, the form of research Lubchenco had in mind assumes that research will stay true to the ideals of inquiry, rather than advocacy.

Some scholars distinguish the “plain-and-simple inquirer” from the “advocacy researcher” and suggest that individuals view these two types of truth seekers as two extremes of a continuum.\textsuperscript{276} The inquirer seeks all evidence, looks for answers, and finds answers, no matter what the answers turn out to be. The inquirer is likely to generate reliable evidence.\textsuperscript{277} The advocacy researcher, by contrast, may “minimize the importance of unfavorable evidence he/she can ignore or explain away.”\textsuperscript{278} In other words, “[S]cientific work can be distorted and impeded when it gets entangled with litigation.”\textsuperscript{279} Advocacy research is inconsistent with disinterested research. Indeed, advocacy research is much more likely to be biased.\textsuperscript{280} Consequently, advocacy research is often termed “lawsuit science,” “junk science,” or “litigation-driven science.”\textsuperscript{281} Any research “skewed by the desire to advance one side in litigation” raises legitimate issues of scientific integrity.\textsuperscript{282}

Federal Rules of Evidence\textsuperscript{283} and state equivalents\textsuperscript{284} give courts the power to appoint experts at their discretion. In other

\begin{itemize}
  \item \textsuperscript{275} Lubchenco, supra note 272; see also Jane Lubchenco, Entering the Century of the Environment: A New Social Contract for Science, 279 Sci. 491, 494-95 (1998).
  \item \textsuperscript{276} Susan Haack, What’s Wrong with Litigation-Driven Science? An Essay in Legal Epistemology, 38 SETON HALL L. REV. 1053, 1072, 1074 (2008).
  \item \textsuperscript{277} For more background on what “reliable” evidence means, see Ryan Hackney, Flipping Daubert: Putting Climate Change Defendants in the Hot Seat, 40 ENVTL. L. 255 (2010).
  \item \textsuperscript{278} Haack, supra note 276, at 1070, 1072.
  \item \textsuperscript{279} Id. at 1056.
  \item \textsuperscript{280} Id. at 1075. Haack writes: “To describe research as ‘litigation-driven’ may mean either (a) that the need for this work arises out of litigation, or (b) that the work is undertaken for the purpose of finding evidence favoring one side in litigation, and explaining away or otherwise playing down evidence favoring the other side.” Id.
  \item \textsuperscript{281} See id.; see also Danielle Marie Stager, Comment, From Kepone to Exxon Valdez Oil and Beyond: An Overview of Natural Resource Damage Assessment, 29 U. RICH. L. REV. 751, 753 (1995) (focusing on environmental damage assessment, including the “emergence of litigation-driven science following the Exxon Valdez oil spill”). Stager noted that the Exxon Valdez litigation featured lawyers “in all post-spill activities,” which influenced the NRDA process. Id. at 785. Scientists raised concerns about the legitimacy of research generated by scientists Exxon hired—research the state of Alaska sealed. Additionally, evidence gathered by government actors was sealed. Stager pointed out that sealed evidence not only erodes public trust, but also adversely affects emerging oil spill law. Id. at 786.
  \item \textsuperscript{282} Haack, supra note 276, at 1081.
  \item \textsuperscript{284} Id.
\end{itemize}
words, judges are not required to rely on experts presented by adversarial parties. For example, the Court-Appointed Scientific Experts Project of the American Academy for the Advancement of Science aids in the process of determining the truth. As the previous recommendation indicates, courts should convene independent panels of scientists to conduct the research needed to get to restoration effectively and efficiently. Together, these first two recommendations repair fundamental flaws in both CERCLA and OPA implementation, as both statutes currently incentivize Trustees and RPs to prepare for court battle.

C. Recommendation Three: Increase Corporate Disclosure Requirements

Investors and consumers expect voluntary corporate disclosures about both environmental performance and green initiatives. These disclosures should be clear, accurate, and complete. Moreover, these disclosures should be internally consistent. For example, when BP ranks number twenty-five in the top 100 “toxic companies” while touting its commitment to sustainability, BP ought to explain this apparent inconsistency. New Governance principles provide guidance to remove these inconsistencies. In particular, ethical precepts require clear, accurate, and complete disclosure. When corporate actors fail to “walk the talk,” a vigilant and responsible media is likely to offer the transparency that corporations fail to provide. When patterns of secrecy and inconsistency become clear, investors and consumers act. In response, they expect state and federal legislators and regulators to intervene on their behalf.

Legal scholars should urge the Federal Trade Commission (FTC) and Securities Exchange Commission (SEC) to articulate and enforce public policy via laws and regulations that (1) prevent deception and unfairness in the marketplace.

287 Toxic 100 Air Polluters, POL. ECON. RES. INST. (Mar. 2010), http://www.peri.umass.edu/toxic_index/.
and (2) encourage informed investment decisions. The FTC already guides marketers to refrain from using general terms such as “environmentally-friendly.” Currently, the FTC is proposing guidelines to marketers that want to use product certifications and seals of approval. The FTC’s proposed regulations make clear that marketers cannot deceive consumers by implying that an independent third party certifies their products. The FTC gives an example. If a company that places a label (a “GreenLogo for Environmental Excellence,” for instance) on a product, that company may mislead consumers to believe that an independent third party awarded the seal. In short, the FTC may provide effective tools to regulate transparency in corporations’ environmental policies. Unfortunately, the FTC’s power in this space does not extend beyond product marketing.

The SEC also requires accurate disclosure. The SEC asks companies to disclose material information to potential investors. When companies file reports with the SEC, they must include accurate information on environmental risks and liabilities. Generally, though, public companies have not been forthcoming about environmental risks.

---


290 See Coppolecchia, supra note 288, at 1378.


292 Id.

293 Currently, consumers are challenging S.C. Johnson’s labeling. In particular, litigation contends that S.C. Johnson’s “Greenlist” system—which gives its own products, including Windex and Shout, its “Greenlist” seal of approval—is deceptive because it implies its products have been certified by an independent third party. Instead, “Greenlist” is actually a patent held by S.C. Johnson. Lawsuit Says S.C. Johnson’s Green Labeling Deceptive, WEAU NEWS (Eau Claire, WI) (Feb. 9, 2011, 7:11 AM), http://www.weau.com/home/headlines/Lawsuit_says_SC_Johnsons_green_labeling_deceptive_115631579.html.


295 Id. at 343.

296 Id. at 325-26.
Currently, investor organizations and researchers are asking for increased transparency from corporations. In essence, they expect the SEC to improve monitoring compliance with environmental disclosure requirements. The Sarbanes-Oxley Act empowers regulators to increase expectations of corporations with regard to transparency, including transparency about environmental information. In particular, companies should inform investors and the public about all environmental risks: for GE, transparency in the Hudson’s PCB dredging; for BP, transparency in the cleanup and restoration of natural resources in the Gulf. For companies beyond GE and BP, this might include transparency in how a particular company plans to prevent environmental disasters and how the company would act should it face an environmental crisis.

CONCLUSION

Considerations regarding the role and place of nature in human life are more important today than ever before. RPs have a duty to safeguard the corporation’s welfare—which obligates them to balance the multiple, and sometimes conflicting, claims of stakeholders, including shareholders, employees, and residents of local communities. Trustees have a duty to act on behalf of the public, restoring, rehabilitating, replacing, or acquiring the equivalent of injured resources and services after releases of pollutants. As a society, we have made considerable progress in environmental protection by asking government regulators to command and control. More recently, a wide range of stakeholders have asked for more collaborative, cooperative approaches that are flexible enough to respond to the particular facts of an environmental situation.

297 Id.
298 Id. at 343-45.
299 Id. at 325-26.
300 Lubchenco, supra note 272, writes:

We are beginning to understand that human health is an environmental issue, that social justice is an environmental issue, that the economy is in reality an environmental issue, and even national security is an environmental issue. As we begin to appreciate the fundamental ways that humans are dependent upon the functioning of intact ecological systems of the planet, we realize that those systems provide not only goods but the services that collectively are our life support systems. These life support systems determine our health, our economies and our future in ways we are only beginning to appreciate.
In NRDA specifically, the promise of cooperative NRDA—expedited processes with reduced overall costs and improved solutions—has yet to be realized. Flaws in the law and the limited usefulness of other avenues of social control have created hurdles. Hurdles include litigious railroading, lack of neutral scientific information, and weak disclosure requirements. These hurdles have impeded effective collaboration and, in turn, delayed New Governance principles from taking complete effect.

We expect little debate about whether cooperative NRDA and New Governance are worthwhile ideas. We do, however, expect arguments over the solutions we have suggested: (1) reforming CERCLA and OPA to incentivize cooperation, (2) reframing science as a mode of social control, and (3) increasing corporate disclosure requirements. These recommendations change the status quo, and change is difficult even when necessary. By putting a stop to feuding between RPs and Trustees, insisting on neutral science that inspires better behavior, and asking companies to demonstrate more transparency, RPs and Trustees will be poised to comply with their duties and meet their responsibilities to get to restoration in U.S. waterways. Moreover, the recommendations we suggest support corporate behavior that is environmentally and socially responsible.
THE ABRAHAM L. POMERANTZ LECTURE

Don’t Blink
SNAP DECISIONS AND SECURITIES REGULATION

Frank Partnoy†

Modern securities markets move at record speed. Trading decisions are faster than ever. Average investors can immediately acquire information. Rapid technologies have benefits, particularly reduced costs. But fast-moving markets can also be dangerous. Few people had time to think carefully during the financial crisis of 2008 or the “flash crash” of May 6, 2010, when stocks plunged 5-6 percent in minutes and then rebounded almost as quickly.

This article explores the consequences of this speed for securities markets. It addresses the extent to which securities regulation should take into account the pace of decision making. It discusses recent scholarly research on snap decisions and suggests legal reforms, some designed to harness the power of quick decisions and others directed at their dangers. It proposes that regulators slow down the markets with proposals ranging from the improbably difficult (steps to respond more deliberately to crises) to the improbably simple (adding a lunch break to the trading day).

† George E. Barrett Professor of Law and Finance, University of San Diego School of Law. I am grateful to Michael Cone and Andrew Mundt for research assistance, and to Laura Adams, James Fanto, Kent Greenfield, Kristin Johnson, and Shaun Martin for helpful comments. I also want to thank Elizabeth Alper and the staff of the Brooklyn Law Review for help throughout the process. Finally, I am grateful to the Pomerantz family for establishing this lecture series and for including me as one of its honored presenters.
INTRODUCTION

This essay addresses snap decisions and securities law, so it seems appropriate to begin with a story of one lawyer's snap reaction to Abraham Pomerantz, the renowned and respected plaintiffs' securities litigator who pioneered the use of derivative suits by shareholders against corporate officials and whose life and career this lecture series honors.¹ I have this story as double hearsay from Ed Labaton, another well-known plaintiffs' lawyer, who heard it during the 1960s when his firm was four floors below the Pomerantz firm and shared its library.²

It was either a Tuesday or a Thursday, the days on which motions were heard back then, outside room 506 of the federal courthouse in Manhattan. Abe Pomerantz and a defense lawyer from a white-shoe New York law firm had just finished arguing a motion. Outside the hearing room, in the lingering heat of the argument, the defense lawyer snapped. He made a nasty personal attack on Abe, calling him, among other things, a “strike-suit lawyer.” Abe didn’t take insults quietly, so he got in the other lawyer’s face, pointed his finger, and exclaimed, “I’ll see to it that I never sue one of your clients again.”

If that lawyer had paused for a moment to think about why he was able to make a living as a securities defense lawyer, he might have held his tongue. In this piece, I argue that the same kind of pause that might have helped this lawyer also might be good policy in the securities markets. Indeed, I intend to show that much of the wisdom of securities regulation is directed at limiting or lengthening snap decisions by market participants. I also suggest that the study of time and timing might be a fruitful area of exploration for securities-law scholars and that notions of delay should play a more prominent role in the study of markets and corporate and securities law, and in policy.

In some areas of corporate and securities regulation, the law introduces delay and probably gets it about right. There is a waiting period before registration of securities for an initial public offering, and it is illegal to sell securities during that

¹ The Pomerantz Lecture honors the life and work of Abraham L. Pomerantz, a 1924 graduate of Brooklyn Law School. The lecture series focuses on topics of corporate securities law and related issues of professional responsibility. The law firm of Pomerantz Haudek Grossman & Gross LLP, of which Abraham Pomerantz was the founding partner, provides continuing support for this series.
² Correspondence with Edward Labaton (Feb. 26, 2011) (on file with author).
period. There is a review and comment process with the Securities and Exchange Commission (SEC) for registration statements. Private placements are exempt and consequently can be created and sold much more quickly. (It is worth noting that the bulk of troubled assets sold during the financial crisis, including collateralized debt obligations (CDOs), fell within the private-placement exemption.) Resales by security holders are restricted by Rule 144—no resale for a year. Proxy regulation delays voting for up to a year. Various provisions of the Williams Act delay tender offers. There are delays related to gun jumping, Hart-Scott-Rodino, and other regulatory review processes. Section 16(b) has a six-month disgorgement rule. Various securities filings are delayed to limited extents, including insider ownership forms and Schedules 13D and 13F. 

---

3 John C. Coffee, Jr. & Hillary A. Sale, Securities Regulation: Cases and Materials 165 (Robert C. Clark et al. eds., 11th ed. 2009) (There is an average of two to three months of work that must be accomplished before the registration statement can be filed.).
4 Id. (The period before the Commission finally issues its letter of comments can vary greatly. The current SEC policy calls for thirty days but it can take up to one hundred days. It often takes longer at the end of the calendar quarter or in March for financial statement filings.)
5 Id. at 74 (Private placements do not require a registration statement and the purchasers are often sophisticated and can be reached quickly and personally.).
7 Coffee & Sale, supra note 3, at 531; see also 17 C.F.R. § 230.144 (2010) (Qualified institutional buyers (QIB) may purchase a restricted security, but that QIB still may not resell it to a non-qualified purchaser until after one year has passed.)
8 Tom Burnett, The Key Points to Look for in a Corporate Proxy Statement, AM. ASS’N OF INDIVIDUAL INVESTORS J., Feb. 2001, at 8, available at http://www.aaii.com/journal/article/the-key-points-to-look-for-in-a-corporate-proxy-statement (“All publicly traded companies—with the exception of the tiny ones that are listed on the Nasdaq Bulletin Board—must file a proxy statement once a year in advance of their annual meeting”); see also 15 U.S.C. § 78n(c) (2006) (prior to the annual meeting, even if no proxy solicitation is made, the issuer must still file with the Commission and transmit to all holders information equivalent to that found in the proxy).
9 Coffee & Sale, supra note 3, at 726-29. Shareholders can withdraw their tendered shares from seven days until sixty days after commencement. Other provisions also reduce the pressure to tender.
10 E.g., 15 U.S.C. § 77e (It is illegal to sell or offer to buy securities before a registration statement has been filed); id. § 18a (This latter code section requires a filing and waiting period before any person that doesn’t meet an exception may acquire voting shares.)
11 Id. § 78p.
12 Id. § 78m (Schedule 13D is ten days and Schedule 13F is forty-five days); see also Schedule 13D, U.S. SEC. & EXCHANGE COMMISSION, http://www.sec.gov/answers/sched13.htm (last visited July 30, 2011); Form 13F—Reports Filed by
capital gains are taxed at higher rates than long-term gains.¹³ Even the most ardent supporters of market efficiency use one-day, or even multiday, event studies, rather than instantaneous analyses of price changes, to assess loss causation and damages in securities litigation.¹⁴ Deal litigation is fast-paced and frenetic, but deal-protection devices create time for directors, lawyers, and judges to consider mergers more carefully.¹⁵ All these provisions illustrate an unspoken, yet overarching, objective of corporate and securities law—to slow us down.

Conversely, consider the dangers when regulators or legal rules favor a quick response. Critics of the government's response to the financial crisis, including Sheila Bair, former head of the Federal Deposit Insurance Corporation (FDIC), have noted that panic and quick reactions led to poor decisions, particularly in the rescue of Bear Stearns, AIG, and other banks (and in the opposite failure to anticipate the complex fallout from the Lehman Brothers bankruptcy).¹⁶ The public, and relatedly legislators, tend to react quickly and negatively to short sellers, even though short selling played a valuable and important role in uncovering and publicizing financial misstatements at various financial institutions, as well as Enron.¹⁷ Many critics claim corporate officers and directors are increasingly focused on short-term share price maximization instead of long-term sustainable profits.¹⁸ Financial reporting is


¹³ Ivo Welch, CORPORATE FINANCE: AN INTRODUCTION 322 (Donna Battista ed., 2009).

¹⁴ Frank Torchio, Proper Event Study Analysis in Securities Litigation, 35 J. CORP. L. 159 (2009) (Frank Torchio is president of Forensic Economics, Inc., and teaches finance and economics at the William E. Simon Graduate School of Business Administration at the University of Rochester. His article is referenced here to showcase the prevalence of event studies in today's market.). For a general synopsis of event studies, see Alan Palmeter & Frank Partnoy, CORPORATIONS: A CONTEMPORARY APPROACH 91 (2010).

¹⁵ See Palmeter & Partnoy, supra note 14, at 959-60.


done on a quarterly basis even though most investors have much longer time horizons. Executive compensation is also relatively short term. Even annual bonuses create an incentive mismatch when the risks associated with employee action are borne over longer periods. These issues have become especially challenging in recent years, given the crush of technology, the press of constant e-mail, the temptations of the Internet, and the resulting focus on the short term.

In this article, I argue that regulation often takes, or should take, the approach of encouraging or ordering delay, of saying “Don’t Blink.” There are two senses in which I mean “Don’t Blink.” One is about keeping our eyes wide open and looking closely so we don’t miss something important. The other is about the dangers of making snap decisions, as fast as the blink of an eye. My goal in exploring these two ideas is to encourage scholars to examine the role of delay in financial-market decision making, and to explore the intersection of decision making and time management, so we can better understand the benefits associated with waiting and the art and science of delay.

This article will focus on two prominent examples in which timing posed particularly important policy challenges: the financial crisis of 2008 and the flash crash of May 6, 2010. Part I describes the financial crisis, comparing the ways in which Goldman Sachs and Citigroup handled the situation and analyzing the government’s decision not to save Lehman Brothers. Part II discusses the flash crash, documenting the concerns with high-frequency trading and proposing implementation of circuit breakers and lunch breaks.

costs, encourage dangerous risk-taking, and put stakeholder’s long-term investments in jeopardy.)
I. THE FINANCIAL CRISIS

I begin my discussion of the financial crisis with a story about Lehman Brothers and the book Blink, and an introduction to the notions of snap decisions and delay in the financial markets. Next, I compare the way Goldman Sachs dealt with the financial crisis with the approach taken by Citigroup, and I analyze the companies’ procedures through the lens of the first sense of “Don’t Blink”—staying focused on what is important. Finally, I describe the weekend government meeting that determined the fate of Lehman Brothers and show how this approach comports with the second sense of “Don’t Blink”—making a quick decision.

A. Lehman Brothers and Blink

Several months after September 2008—the time most people think of as the peak of the recent financial crisis—there were reports that Lehman Brothers, the investment bank that filed for bankruptcy on September 15, 2008, and thereby triggered a credit freeze, had assigned Blink as part of its training program for top executives. Lehman had invited Malcolm Gladwell, the author of Blink, to speak to the firm’s elite group of future leaders. Blink, a bestselling book from 2005, proposed that we should focus on the first two seconds of our decisions. Gladwell called on readers to “acknowledge there can be as much value in the blink of an eye as in months of rational analysis.” Although the later parts of Blink also explored some of the dangers associated with biases in decision making, the media and the public received the book primarily as a justification for intuition and snap decisions.

Before September 15, 2008, my personal views were decidedly pro-Blink and anti-Lehman. I greatly admired Gladwell, and I had publicly excoriated Lehman. In fact, on

---

20 ANDREW ROSS SORKIN, TOO BIG TO FAIL: THE INSIDE STORY OF HOW WALL STREET AND WASHINGTON Fought to Save the Financial System—And Themselves 120 (2009).
22 Id. at 17.
23 See, e.g., id. at 252-53.
September 14, 2008, just hours before Lehman’s bankruptcy, I published an opinion piece in the Financial Times entitled “Hubris—is thy name Richard Fuld?” My basic take on these topics was: Blink good; Lehman bad.

So my initial reaction when I heard about Lehman’s Blink reading assignment—my snap response, in fact—was that Lehman officials must have overreacted to the book’s praise for quick reactions. The last thing Richard Fuld, the head of Lehman, and his fellow managing directors needed was The Power of Thinking Without Thinking, which is the subtitle to Blink. Indeed, the media portrayed this incident as an example of the folly of Lehman’s obsessively short-term focus, and several journalists pointed to Joseph Gregory, Lehman’s former president, as the architect of the firm’s speedy mindset.

As academics, we have two main weapons: a skepticism about received wisdom and a lot of time on our hands. So I explored the details of this anecdote about Lehman and Blink with the hope that I might learn something about Lehman’s culture. My interviews and research revealed two interesting things.

First, Lehman’s leadership program did not embrace snap decision making; instead, it did the opposite, stressing the dangers of snap decisions. Joseph Gregory and other senior managers at Lehman created a cutting-edge, intellectually rigorous training program, taught by leading social science scholars. The substance of the program was highly skeptical of intuition and snap decisions. Lehman even paid to develop a customized Implicit Association Test, or IAT, to demonstrate to its own officials how they were biased regarding race, age, gender, and politics. The program’s participants and content were diverse along just about every axis.

Malcolm Gladwell’s talk was at the end of the program, a capstone designed to get managing directors from around the world together to discuss the firm’s global approach to decision making. But the folks at Lehman didn’t study a caricature of Blink. They read and studied the whole book, including chapter

---

25 Gregory was terminated in June 2008, months before the firm’s bankruptcy. See Ben White, Lehman Chief Accepts Blame for $2.8bn Loss, FT.com (June 16, 2008, 9:46 PM), http://www.ft.com/intl/cms/s/0/50a84d4c-3b99-11dd-9cb2-0000779fd2ac.html#axzz1MpxRSu65.
26 During 2011, I interviewed and corresponded with several former Lehman employees, who wished to remain anonymous.
6 covering the Bronx police shooting Amadou Diallo and the concluding chapter on gender bias in music auditions.

And yet Lehman’s employees made colossally bad decisions. They took on too much subprime mortgage risk. They hid liabilities from shareholders. They made these bad decisions over months and years, and their elite senior leadership did not spot or correct them.

Some scholars have argued that the popular interpretation of Blink’s thesis was oversimplified and incorrect. Seventh Circuit Judge Richard Posner criticized Blink’s assumptions and opined that the book was filled with attenuated anecdotes, poor analyses, and overreaching assumptions. More recently, social science researchers published empirical work that demonstrates the dangers associated with snap decisions. This literature shows that two seconds is rarely the optimal amount of time in which to make a decision.

Even for what we call snap decisions, people frequently benefit from waiting at least several seconds, up to a minute or so. Some critics argue that analysis trumps intuition for most decisions and suggest that even snap decisions are the result of longer-term analysis. Within particular time frames, ranging from a split second to years, people are often better off making decisions at the end of the relevant time period—at the very last possible instant. Although this kind of delayed action is


28 Fraud allegations were raised against Lehman and its accounting firm Ernst & Young. See Going for the Auditors: The Ultimate Target of the Lawsuit May Be Lehman’s Former Bosses, ECONOMIST (Dec. 29, 2010), http://www.economist.com/node/17800083.

29 See, e.g., Big Think Interview with Christopher Chabris, BIG THINK, http://bigthink.com/ideas/20582 (last visited July 30, 2011) (“We are, in a way, taking on the impression that a lot of people have from books like, ‘Blink,’ by Malcolm Gladwell, and others in that category, which is sort of an uncritical belief in the power of intuition and snap judgments and so on, and the idea that you should rely on them whenever possible.”).


32 See, e.g., Dana R. Carney, C. Randall Colvin & Judith A. Hall, A Thin Slice Perspective on the Accuracy of First Impressions, 41 J. RES. PERSONALITY 1054 (2007) (studies showing accuracy frequently increasing with response times of greater than two seconds).

often labeled procrastination, it is really more a form of delay management, a process of understanding when to go fast and when to go slow.\textsuperscript{34}

Is it possible to generalize from this new perspective on Lehman and Blink to gain any insight into the causes of and response to the crisis? One way to assess the importance of relying on longer-term analysis versus short-term intuition is to compare two institutions that arguably were at opposite poles of the financial crisis: Goldman Sachs and Citigroup.

B. Goldman Sachs, Citigroup, and the Gorilla: The First Sense of “Don’t Blink”

There are many versions of why the financial crisis occurred, but they often boil down to the following condition: “If you had only taken a step back and thought this through . . . .” For example, many financial market participants—bankers, investors, and regulators—relied on credit rating agencies and mathematical models for analytical shortcuts that were woefully inaccurate and inadequate. They used ratings and math as mnemonic devices to streamline a massive flow of information into something they could understand.\textsuperscript{35} Then they decided that, if this complex structured instrument is rated triple-A, or even higher than triple-A, it must be low risk.\textsuperscript{36}

Consider this sentence: if a Monte Carlo simulation based on historical correlation assumptions predicts that the probability of subprime mortgage defaults rising to a level that would impair a super senior tranche of a synthetic collateralized debt obligation is sufficiently small that the tranche is virtually risk free, then we can hold tens of billions

\textsuperscript{34} Professor Manuel Utset has suggested that the notion of delay management can be captured by the mental process of cost-benefit analysis, instead of intuition: if the cost of acting immediately outweighs the long-term benefits, the person will act. Manuel A. Utset, Procrastination and the Law, in The Thief of Time: Philosophical Essays on Procrastination 253, 253-55 (Chrisoula Andreou & Mark D. White eds., 2010).

\textsuperscript{35} See, e.g., Why Economists Failed to Predict the Financial Crisis, KNOWLEDGE@WHARTON: FIN. & INVESTMENT (May 13, 2009), http://knowledge.wharton.upenn.edu/article.cfm?articleid=2234.

\textsuperscript{36} When the realization hit that the CDOs were much riskier than the rating they carried, the crediting agencies downgraded them and the banks had to change not only their formulas, but also their balance sheets. Between the third quarter of 2007 and the second quarter of 2008, mortgage securities had been downgraded by $1.9 trillion. See J on Birger, The Woman Who Called Wall Street’s Meltdown: Star Bank Analyst Meredith Whitney Says the Economy Is About to Sink Into a Deep Recession, CNNMONEY (Aug. 6, 2008, 11:57 AM), http://money.cnn.com/2008/08/04/magazines/fortune/whitney_feature.fortune/index.htm.
of dollars of that exposure without worry, or indeed without even disclosing it. One positive result of the financial crisis is that many more people understand the preceding sentence than did a few years ago. But not many people took the time to understand the principles underlying the text of this sentence or to ask for a detailed analysis of why so many triple-A-rated synthetic instruments could be created without high-quality underlying assets.

With that background, it is strange that Goldman Sachs would emerge as the supposed villain of the financial crisis. By every account, Goldman engaged in vigorous, deliberative risk management. Groups of senior managers discussed every major position. They not only marked positions to market on a daily basis, but they analyzed worst-case scenarios. They publicly disclosed value-at-risk, or VAR, numbers, but internally they did not take them on faith. They listened to and learned from their counterparties, particularly hedge funds, who were betting against mortgages. When the ABX indices of subprime mortgages began declining in 2006, Goldman’s analysts undertook a detailed internal study and produced a thorough internal report about the risks in its mortgage business. They studied actual prices, and actual profits and losses, instead of trusting ratings and models. As a result, in December 2006, Goldman determined that it was too exposed to subprime mortgages, and it reduced that exposure.

Goldman has been vilified for how it reduced its exposure to the mortgage market, particularly for selling approximately $25 billion of CDOs during the eight months

---

40 Hearings, supra note 38, at 95-96 (testimony of David Viniar); see also Nocera, supra note 37.
41 See Hearings, supra note 38, at 95-96 (testimony of David Viniar); see also Nocera, supra note 37.
after it decided to reduce its positions. Goldman paid $550 million to settle SEC allegations that the firm failed to disclose information about CDO deals it sold to clients. But whatever you might think of Goldman’s behavior, these actions didn’t cause the financial crisis. Indeed, Goldman was one of a handful of financial institutions that survived the crisis intact because it reduced its mortgage exposure in late 2006.

Goldman spotted the eight-hundred-pound gorilla in the financial markets in late 2006—the huge risk that a housing price decline would lead to highly correlated system-wide defaults, which would erode the value of subprime mortgages and particularly super-senior tranches of synthetic CDOs. Goldman Sachs avoided the kind of inattentional blindness that plagued other banks with exposure to these instruments. Goldman stepped back and questioned its own judgment. Now consider Citigroup as a counterexample to Goldman.

Citigroup made some of the most egregious mistakes of any financial institution during the crisis. Its board and executives made snap judgments based on intuition and mnemonics and did not undertake more thoughtful analysis. It lost billions of dollars on super-senior positions. Citigroup’s snap judgments and a failure to step back and think should have destroyed the firm.

Although the Financial Crisis Inquiry Commission report was a trifurcated political mess, some of its hearings provided useful color, particularly about Citigroup. Citigroup’s directors and officers accepted naïve risk management perspectives. For example, Robert Rubin, the former Treasury Secretary, who was paid more than a hundred million dollars in cash and stock during his eight years at Citigroup, testified that “I don’t think anybody focused on the CDOs. This was one business in a vast enterprise, and until the trouble developed, it wasn’t one that had any particular profile.” Rubin said he relied on Thomas Maheras. He said, “You know, Tom Maheras was in charge of trading. Tom was an extremely well regarded

---

44 Id. at 193.
trading figure on the street. . . . And this is what traders do, they handle these kinds of problems."  

What about Tom Maheras? Maheras, the co-CEO of Citigroup’s investment bank, made more than $34 million in 2006 but admitted he spent “less than 1% of his time thinking about CDOs.”  

Yet another of Citigroup’s managing directors, Susan Mills, had warned about the deteriorations in Citigroup’s subprime loan quality in early 2006. This was the gorilla: default rates that were doubling or even tripling, that threatened Citigroup’s tens of billions of dollars of super senior CDO positions. Yet the senior managers didn’t see this weakness. They didn’t have the perspective Goldman did because they didn’t step back. Instead, they sold more CDOs and retained even more subprime risk. The Federal Reserve found that Citigroup’s senior management “did not appropriately consider the potential balance sheet implications of this strategy.”  

Why didn’t Citigroup see the gorilla? When we are distracted, we don’t see gorillas. I am using the term “gorilla” deliberately, to reference the visual awareness experiments at Harvard conducted by Christopher Chabris and Daniel Simons. Chabris and Simons showed their subjects a short film depicting two teams of people, one dressed in white and one in black, moving around and passing basketballs. They asked their subjects to silently count the number of passes made by players wearing white shirts. Halfway through the video, a student wearing a full-body gorilla suit walks in, stops in the middle of the players, thumps her chest, and walks off. She spends nine seconds on screen, about one-sixth of the entire video. Yet when Chabris and Simons queried their subjects after they watched the video, one-half of people did not notice the gorilla. They had “inattentional blindness.” They devoted their attention to one part of the world, and did not recognize

48 Id.
49 Id. at 198.
50 Id. at 260.
51 Id. at 199.
52 CHABRIS & SIMONS, supra note 31, at 5.
53 Id.
54 Id. at 6.
55 Id. at 5-6.
56 Id. at 6.
57 Id.
the striking events in the other part. Numerous studies have replicated this result.\footnote{See id. at 39-40 (citing, among other examples, people who miss safety infractions right in front of them, high school teachers and administrators who fail to notice bullying, and fair-minded employers who do not notice discriminatory practices).}

Goldman saw the gorilla; Citigroup did not. Numerous hedge funds saw the gorilla; regulators did not. The financial crisis boils down to a simple problem: not enough people saw the gorilla. This is the first sense of “Don’t Blink”—don’t let your attention lapse, stay focused on the important matters.

What, if anything, can securities regulation do about this problem? There are two things, and they are not new. Indeed, they are the twin pillars of the 1930s securities laws: mandatory disclosure and ex post anti-fraud enforcement.\footnote{See Frank Partnoy & Lynn E. Turner, Bring Transparency to Off-Balance Sheet Accounting, in MAKE MARKETS BE MARKETS: THE REPORT (2010), available at http://makemarketsbemarkets.org/report/MakeMarketsBeMarkets.pdf} As Enron illustrated, adequate disclosure does not mean burying opaque references in footnotes. Enron’s infamous footnote 16, which purported to disclose some of the firm’s off-balance sheet risks, set off a few reporters and short sellers, but it didn’t adequately inform investors of the firm’s risks in a salient way.\footnote{See STAFF OF S. COMM. ON GOVERNMENTAL AFFAIRS, 107TH CONG., FINANCIAL OVERSIGHT OF ENRON: THE SEC AND PRIVATE-SECTOR WATCHDOGS 35-36 (Comm. Print 2002).} Likewise, Citigroup’s disclosure of subprime risk was impenetrable, buried in a web of complex off-balance sheet entities.\footnote{See RooseveltNYC, Frank Partnoy on Off-Balance Sheet Transactions (MMBM—Roosevelt Institute), YOUTUBE (Mar. 6, 2010), http://www.youtube.com/watch?v=xpqYl_xEoVo (last visited July 30, 2011).}

If you want to ensure that people spot the gorilla, you have to tell them there might be a gorilla. Citigroup could have made its risks salient to investors and senior managers. If Citigroup had disclosed worst-case scenarios in its financial statements—that it would lose tens of billions of dollars if housing prices declined significantly—surely Bob Rubin and Tom Maheras would have paid more attention.\footnote{Some have questioned whether the gorilla wasn’t missed, but rather ignored, largely due to the profitability of the risky investments before the crash. It remains unclear how many bank executives were consciously aware of their institutions’ risk exposure but were hubristic about the risk, or whether senior personnel really failed to understand the key mathematical algorithms. See Donald C. Langevoort, Chasing the Greased Pig Down Wall Street: A Gatekeeper’s Guide to the Psychology, Culture and Ethics of Financial Risk-Taking, 96 CORNELL L. REV. 1209, 1221-22 (2011).}

According to Chuck Prince, Citigroup’s CEO, his and the firm’s decisions should not be criticized in hindsight. He said,
If someone had elevated to my level that we were putting on a $2 trillion balance sheet, $40 billion of triple-A, zero-risk paper, that would not in any way have excited my attention. It wouldn't have been useful for someone to come to me and say, “Now, we have got $2 trillion on the balance sheet of assets. I want to point out to you there is a one in a billion chance that this $40 billion could go south.” That would not have been useful information.63

This statement is hard to reconcile with Prince's 2006 comment: “When the music stops, in terms of liquidity, things will be complicated. But as long as the music is playing, you've got to get up and dance. We're still dancing.”64

What the FCIC investigators should have asked Prince was this: “What if someone had elevated to your level the risk that the bank would become insolvent if housing prices declined 30 percent? Would that have excited your attention? Would that have been useful information?” Worst-case scenarios are the gorillas of the financial markets, and they should be disclosed in far greater detail. Prince's reference to “complicated” in his “music” quote suggests that he was aware, at least in part, of these risks. In any event, even if the awareness was buried at a lower level within the bank, there should have been a mechanism that led to its disclosure.

Unfortunately, the Dodd-Frank Act and other reforms do not require disclosure of these kinds of facts, either to the public or to regulators. Some companies try to do this privately through risk management,65 emulating Goldman Sachs. But securities regulation reform could help encourage managers and shareholders to engage in more long-term analysis than short-term intuition by requiring more robust and salient disclosure of worst-case scenarios.

The second pillar—antifraud—is also relevant. If the directors and officers of Citigroup are not held responsible for failures to disclose gorillas, why would they disclose gorillas? Yet private rights of action by shareholders have been substantially restricted in recent years,66 and government prosecutions have not filled the gap. The early shareholder

63 FCIC REPORT, supra note 43, at 260.
derivative litigation against Citigroup was dismissed; it remains to be seen whether federal class actions against Citigroup, and other financial institutions, will fare better. As of late 2011, there had been no major criminal cases against individual Wall Street employees.

Last year, the SEC civilly charged Citigroup with repeatedly making misleading statements about its exposure to subprime-mortgage-related assets. According to the SEC, “Between July and mid-October 2007, Citigroup represented that subprime exposure in its investment banking unit was $13 billion or less, when in fact it was more than $50 billion.” For that epic fraud, Citigroup paid a $75 million penalty. Its former chief financial officer Gary Crittenden agreed to pay $100,000, and former head of investor relations Arthur Tildesley, Jr., agreed to pay $80,000. These numbers are obviously inadequate to deter financial fraud. If Citigroup and its executives had known these would be the penalties in advance, they would not have had any economic incentive to behave differently. A short-term-focused illegal decision is more profitable than a long-term-focused legal one.

Another useful policy tool would encourage market participants to rely, as Goldman did, on market measures of risk instead of ratings and financial models. The Dodd-Frank Act requires the elimination of regulatory references to ratings, and that important project is in progress. The SEC has proposed rules for Forms S-3 and F-3 filings and related documents, based on the presence of a deep market for such seasoned issues, but there is some reluctance to substitute

---

71 Id.
market-based measures of credit risk for ratings in regulation.\textsuperscript{74} International regulators continue to rely on ratings.\textsuperscript{75} The future of ratings in regulation remains unclear.

A market-based approach would improve risk-related incentives at financial institutions. Consider how much more difficult it would have been for Citigroup's CDO desk to take on large amounts of subprime risk in 2006 if—instead of reporting internally that the risk was triple-A, or better than triple-A, or that the probability of loss was one in a trillion according to the model—it had reported the market price of the risk and noted that the price had increased from ten to thirty basis points. Based on its valuation model in late August 2006, Citigroup reported that losses on its super senior tranches might range from $15 million to $2 billion.\textsuperscript{76} Yet market prices, as reflected in credit default swaps, had already plummeted by that time. Markets are not always correct, of course, but market measures of risk can provide an early warning detector, a gut check, to help people avoid using their gut reactions in the wrong way. That is one reason why Goldman decided to reduce its exposure in December 2006, nearly two years before Lehman’s bankruptcy.

Investment decisions based on ratings might not seem like high-speed snap decisions. But they are decisions that are made quickly in a relative sense. If people are accustomed to equating triple-A ratings with safety, then when they see triple-A, they will anchor around the idea that the triple-A-rated instrument is safe. Most of the decision about safety and risk is made immediately, at a pre-conscious level. Hopefully, the snap reaction that people have today to the triple-A symbol is dramatically more skeptical than the snap reaction people had a few years ago.

C. Lehman’s Bankruptcy and a Weekend at the Federal Reserve: The Second Sense of “Don’t Blink”

A different financial crisis decision was “Don’t Blink”-like in the second sense I’m discussing, in that the decision was made too quickly. This is the decision, not by Lehman’s traders about the bank, but about Lehman by its regulators. On

\textsuperscript{74} See Yali N’Diaye, US SEC to Address Reliance on Credit Ratings Yet Again, \textit{Market News Int’l} (Apr. 25, 2011, 2:49 PM) (on file with author).


\textsuperscript{76} FCIC Report, supra note 43, at 264.
September 10, 2008, the “[Federal Reserve Bank of New York] staff put together a draft gameplan for a ‘liquidity consortium’ of major Wall Street banks to provide a forum where these firms can explore possibilities of joint funding mechanisms to avert Lehman’s insolvency.” But the gameplan provided that “[c]onsortium members would be given ‘[v]ery little advance’ notice, ‘2 hours max,’ in order to ‘minimize the risk of outside leaks.’” Then, the consortium banks would have only the weekend—no longer—to perform due diligence on Lehman. If no plan emerged, the Federal Reserve officials would “reach out to regulators in DC and abroad to inform them of potential market disruptions at the opening of business on Monday.”

On Friday night, September 12, 2008, twelve investment bank CEOs were summoned to the Federal Reserve’s headquarters at 33 Liberty Street in New York. Over the weekend, they agreed to provide $20 billion to support a purchase of Lehman by Barclays Capital. But it was a deal that was doomed to fail—or at least doomed to fail by Monday. Barclays executives were not invited to the consortium meeting; they were separately conducting due diligence that night and over the weekend to decide whether to acquire Lehman. Barclays made it clear that in order to guarantee Lehman’s financial obligations, a requirement of any deal, it would need shareholder approval, something that could not happen before Monday. When British regulators confirmed that this requirement would hold, the deal fell through, and Lehman filed for bankruptcy on Monday.

We can never know what might have happened if the government had waited a bit longer. Yes, the credit markets were tight during mid-September 2008. But interest rates were still relatively low. LIBOR was below five percent. Would some parties have been willing to provide short-term loans at higher

---

78 Id.
79 Id. at 1520 (quoting FRBNY, Timeline—Liquidation Consortium (Sept. 11, 2008), at 2 [FRBNY to Exam. 003514]).
80 Id. at 1524.
81 Id. at 1528.
82 Id. at 1524-25.
83 Id. at 1527-28.
84 Id. at 1527, 1535.
rates, or would the markets really have frozen? It is very possible. But regulators could step back and think about these possibilities. That weekend, they were trapped by a snap decision.

Although commentators disagree about a wide range of issues related to the collapse of Lehman, it is undeniable that the decisions about whether to rescue Lehman were made quickly, perhaps more quickly than any financial regulatory decisions in history. The overwhelming pressures from technology—e-mail, the web, computing power, smart phones—sped up the pace of responses. Regulators struggled to avoid the crush of this time pressure.

It is ironic that government officials, who so often are thought to act too slowly, should have acted so quickly under time pressure during the financial crisis. The problem is that regulators have not considered a challenge for all types of leaders in the modern technological age: how to manage delay. Some time-pressured scenarios are unavoidable, but the clear message of recent research is that they should be avoided whenever possible and that, to the extent they cannot be avoided, there should be emergency plans in place so that the senior regulators are experts, not novices, in crisis situations. Regulators, like market participants, should consciously address the art and science of delay.

II. The Flash Crash

The second “Don’t Blink” topic is the so-called “flash crash.” First, I describe the events of May 6, 2010, and discuss the role played by high-frequency trading. Then, I explain that regulation should frame these issues as part of the second sense of “Don’t Blink.” Finally, I propose two strategies to encourage delay: implementation of circuit breakers and the introduction of a lunch break.

A. A Thirty-Six Minute Roller Coaster

At 2:32 p.m. on May 6, 2010, an employee of Waddell & Reed, a mutual fund company headquartered about a mile from my childhood home in Overland Park, Kansas, clicked start on a computerized trading software program.* The firm’s goal was

---

* The details about the “flash crash” were reported in U.S. Commodity Futures Trading Comm’n & U.S. Sec. & Exch. Comm’n, Findings Regarding the Market Events of May 6, 2010: Report of the Staffs of the CFTC and SEC to the
to reduce its exposure to $4.1 billion of stocks it owned by selling something called “E-Mini” futures contracts. The “E-Mini” is based on the Standard & Poor’s 500 Index of top stocks, except that it is traded in small amounts (hence, “Mini”), and it goes through an electronic trading platform instead of the frenzied “open outcry” method still used for other futures contracts (hence, “E”). To hedge $4.1 billion of stocks, Waddell & Reed would need to sell 75,000 E-Mini contracts.

Instead of having its own employees manually enter these orders or calling a broker, Waddell & Reed used this automated computer program. Each minute, the program calculated the number of E-Mini contracts traded during the previous minute. It then automatically sold nine percent of that number. The program was designed to take several hours, or perhaps even days, to sell 75,000 E-Mini contracts.

Instead, the program triggered the fastest roller coaster ride in the history of financial markets. At first, when Waddell & Reed’s computers started to sell, high-frequency traders, with their own computer programs, stepped in to buy. The market was calm and balanced—for about nine minutes.

But after nine minutes, at 2:41 p.m., high-frequency traders began selling the contracts they had accumulated in order to zero-out their positions. High-frequency traders do not typically maintain significant long or short positions for more than a few minutes. During the first minute of their switching sides, trading volume increased, and Waddell & Reed’s automated program responded by selling a larger number of E-Mini contracts. Then, in the second minute, more traders sold, and so did the automated program. During the third and fourth minutes, everyone sold even more, in a kind of high-speed computerized trading death spiral.

---


87 Id. at 2.
88 Id.
89 Id.
90 Id.
91 Id.
92 Id.
93 Id. at 2-3.
94 Id. at 3-4.
95 Id.
96 Id.
97 Id.
98 Id.
By 2:45 p.m., trading volume was exploding, and the E-Mini futures contract was collapsing. Its price had fallen 5 percent in just thirteen minutes. The high-frequency computer programs were a large share of the market at this time. During one fourteen-second period, high-frequency traders accounted for 27,000 E-Mini contracts, about half of the total trading volume.

The decline in the E-Mini contracts instantly spread to the rest of the market. Some of the contagion was bizarre, even inexplicable. Some was due to “stub orders” set at absurdly low prices. Some was due to computer algorithms that hadn’t anticipated this kind of shock. Many high-frequency traders exited their positions entirely, running for the virtual hills. Stock indices other than the E-Mini also collapsed, as did individual stocks. At 2:47 p.m., shares of Accenture plc, the consulting firm, fell from nearly $40 to $30, and then suddenly—in just seven seconds—plummeted to one cent. A few minutes later, shares of Procter & Gamble, the consumer products company, fell from more than $60 to $40. Shares of blue-chip companies such as IBM, Apple, 3M, and General Electric also declined abruptly.

But then, just as fast, the market snapped to life. Accenture traded near $40 again; Procter & Gamble was back above $60. Within minutes, the E-Mini contract and all these stocks recovered. By 3:08 p.m., the market settled, and prices were about the same as they were before Waddell & Reed started the computer program. The program had finished selling 75,000 E-Mini contracts, and then, by instruction, it shut down. The entire ride, the bust and boom now known as the “flash crash,” had taken just thirty-six minutes.

B. High-Frequency Trading

Many critics have blamed high-frequency traders for the flash crash. They say high-frequency trading is socially wasteful and dangerous. Yet there is overwhelming empirical
evidence showing that under “normal” conditions high-frequency traders constitute a powerful positive force in the markets. So-called low-latency trading improves traditional yardsticks for market quality, such as liquidity and short-term volatility. 108 Recent empirical work shows that high-frequency traders did not trigger the flash crash.109 The trigger was the computer program at Waddell & Reed.

However, the evidence also shows that during periods of high market uncertainty—such as May 6, 2010—high-frequency trading is associated with increased volatility.110 High-frequency trading appears to be most dangerous when new information is entering the market, when it can cause prices to swing more dramatically.111

High-frequency trading is a dominant force in modern markets. Estimates suggest that it accounts for almost three-quarters of dollar trading volume in the United States.112 As the SEC has recognized, proprietary high-frequency traders have largely replaced specialists and market makers in stock trading.113 High-frequency traders’ strategies vary widely, and some are more defensible than others.114

High-frequency trading isn’t going away. So what should be done about it? First, it is worth noting that just as computers have beaten human beings at chess and on

111 Id. at 3.
Jeopardy, it is unlikely that human regulators would have much of a chance against high-speed computer trading software. These algorithms move and change too quickly for regulators to act or react in any meaningful way. By the time the SEC/CFTC report on the flash crash was published on September 30, 2010, market participants already had switched to new strategies. No one would use Waddell & Reed’s trading program today. The algorithms that worked last month probably no longer work.

Nevertheless, regulators can try to play the same role they have played in markets generally, even when they are outmatched. No one believes that prosecutors can stamp out all insider trading, but most people still favor some regulatory efforts to deter insider trading.

Likewise, the government can bring cases against high-frequency traders who violate the law. It is unclear how much high-frequency trading is illegal. It probably isn’t a large percentage, but it isn’t zero, either. But for these kinds of illegal, fraudulent, and manipulative activities, there is not a need for new statutes or regulations. Front running and insider trading are already illegal under current law. If regulators are unable to bring cases against illegal high-frequency trading, a new regulatory regime might defer to private rights of action. If investors are disadvantaged by high-frequency traders, one way to police the practice would be through a private attorney-general role. To the extent high-frequency traders are engaged in manipulative market practices, regulators should either prosecute that activity or encourage private actions that deter it.


C. Lunch Breaks and Circuit Breakers: The Second Sense of “Don’t Blink”

The other—and perhaps even more important—policy for regulators to implement is a “Don’t Blink” strategy. They should encourage delay. Indeed, regulators should heed one of the lessons that market participants in the high-frequency trading area are learning: a crucial element of successful trading is delay management. For some strategies, it is best to be first. But for other strategies, it is better to wait a little bit. There are various catch phrases that describe this, like “the first-mover disadvantage” or, conversely, “the second-mover advantage.” Or “the second mouse gets the cheese.” A complete analysis of this phenomenon is beyond the scope of this article, so I will simply note that UNIX, a high-frequency trading firm, was ranked as the top execution-only broker in numerous trading categories from 2005 through mid-2007, and one major reason for its success was that its trading executive was slightly delayed, by a few dozen milliseconds.117

Regulators follow the lead of this private market philosophy and slow down the markets by introducing explicit pauses. One of this article’s themes is that decision makers should take time to step back and think. Yet given the speed of modern markets, there is little time for market participants to do that. Regulators could create more time with circuit breakers.

Regulators have introduced circuit breakers already to force markets to shut down when they have declined by certain specified amounts. After the flash crash, the SEC adopted a pilot program to introduce a five-minute pause if the price of any stock in the S&P 500 Index fell by 10 percent or more during a five-minute period. After the five-minute pause, the primary listing market would use an auction process to determine the new opening price.118 In September 2010, the pilot program was expanded to the Russell 1000 Index and some exchange traded funds. Trading would halt for five minutes.119 These pauses are a


119 Other post-crash solutions include a “limit up-limit down” mechanism that prevents trading outside of a set price band, the elimination of stub quotes, and the use of an
sensible supplement to the circuit breakers that apply generally during major market downturns.

But it is also worth asking more generally if markets would benefit from the introduction of longer pauses, of breaks during the day designed to encourage thinking and deliberation before action. When I worked in Morgan Stanley’s Tokyo office during the 1990s, I was struck by the impact of the ninety-minute lunch break on trading. Not that Morgan Stanley’s traders were models of propriety during lunch: some of the most egregious trades described in my book F.I.A.S.C.O. were created in Tokyo, and were conceived during those breaks.\textsuperscript{120}

Still, as a general matter, Tokyo’s pause in market trading led to more rational thinking about the trading day and often helped cooler heads prevail. During the morning’s two-hour trading session, traders and salespeople focused on prices and deals. There was relentless pressure to execute. But then there was a break from 11:00 a.m. until 12:30 p.m. During that time, traders, salespeople, and clients had conversations that actually lasted for more than a few seconds. They pondered new investment strategies or ideas. They read. Sometimes during the lunch break they even had lunch. Following this break, there was another two-and-a-half hour trading session during the afternoon. The stock exchanges in Hong Kong, Shenzhen, and Singapore followed a similar approach, with ninety-minute, mid-day breaks. In contrast, all the world’s other stock exchanges, including those in the United States, have been and are still open continuously from the morning until the closing bell.

Unfortunately, the Asian markets are now moving toward the Western model. In February 2011, the Tokyo Stock Exchange announced that it would shorten its lunch break by thirty minutes beginning in May, shifting the start of lunch back to 11:30 a.m. from 11:00 a.m.\textsuperscript{121} Some commentators criticized the move. One market participant suggested that extending the time for trading probably would not boost volume because “trading tends to focus around the beginning and end

\begin{quote}
\end{quote}
of trading sessions."

Interestingly, 70 percent of corporate and individual investors opposed the plan to extend trading hours. Yet it passed.

A typical law review article about financial market regulation might propose an intricate and complex reform of computer algorithm-driven trading. Instead, let me offer a more basic reform idea: force traders to break for lunch. I favor the introduction of a lunch break at the New York Stock Exchange and NASDAQ.

A lunch break would create a much-needed pause for reflection and thought during the trading day. Breaks have the additional benefit of creating another opening time, after lunch, when prices would be set based on a pool of bids and offers. At the extreme, the modern trading day might consist simply of two or three brief auctions, with breaks between them. Critics will argue that there would be less liquidity under such a regime, and there very well might. Only an experiment could answer questions about this concern. But the benefits of giving market participants more time to engage in thoughtful discourse and analysis could substantially outweigh any potential loss of liquidity.

Moreover, it is worth asking how much liquidity is necessary in today's securities markets. How often do even the most active traders need to move in and out of positions at particular moments during the day? Most high-frequency traders maintain a flat trading profile, and they zero-out positions right away, or at the latest by the end of the day. Few fundamental traders need to move the bulk of their positions at particular times. And introducing pauses could also deter retail investors from day trading, which is an addictive (and on balance a destructive and losing) strategy. Intraday pauses could create time for people to engage in more productive uses of their time.

Imagine this thought experiment. What if you could trade in U.S. equities or any financial instrument whose value is derived from U.S. equities for only one hour in the morning and one hour in the afternoon? Trades at any other time would be unenforceable. Any purchases and sales during paused periods would be void. Or, in a less extreme version, trades during paused periods could be subject to a transaction tax. Of course, there would be pressure for trades to occur outside of

---

122 Id.
123 Most Investors Oppose Ending TSE Lunch Break, Japan Times Online (Sept. 25, 2010), http://search.japantimes.co.jp/cgi-bin/nb20100925a4.html.
this legal framework for so-called regulatory arbitrage. The policy would require international coordination, but that is getting easier as the major exchanges merge. What if each of the major global markets agreed to trade for an hour or two only, and refused to enforce trades outside those time periods?

Many market participants would complain about a reduction in liquidity. And perhaps they would be right. But liquidity at what cost? And why would one conclude that there would be greater liquidity during a twenty-four hour, constant international trading day, which is where current trading trends are headed? There is a strong argument and evidence that constant trading merely gives the illusion of liquidity.

For example, one lesson from the flash crash is that the supposed liquidity provided by high-frequency traders and others can evaporate very quickly. When there is an error—the next Waddell & Reed program, or the next extra-zero input by a “fat finger”—it will occur at a random time, when liquidity will be limited. Wouldn’t it be better for such errors to occur during a compressed timeframe, when everyone is trading? On a per-trading-moment basis, there arguably should be greater market depth and liquidity after the timing of trading is restricted. Indeed, limiting trading hours might be an especially good idea for less liquid stocks, which would have deeper markets, albeit for a shorter time. The major disadvantage would be that some people who want to trade a few hours earlier will not be able to do so. But who fits within that category?

Interestingly, the SEC/CFTC investigation of the flash crash indirectly supports this pause idea: “Another key lesson from May 6 is that many market participants employ their own versions of a trading pause—either generally or in particular products—based on different combinations of market signals.” The study confirmed that “a liquidity crisis can develop if many market participants withdraw at the same time.”


125 CFTC/SEC FINDINGS, supra note 86, at 6.
policy comes from market-based insights. If markets are pausing, shouldn't regulators?

One final point: although circuit breakers are designed to kick in only when markets are collapsing, why should they have effect only in times of downward stress? The dot.com and housing bubbles are just the latest examples of the dangers associated with rapid moves up as well as down. Perhaps if market participants paused when markets surged they might question more why markets were surging.

CONCLUSION

Delay plays an important, though often hidden, role in financial markets and financial market regulation. Delay management can be an important policy tool for regulators. But rather than conclude by repeating my thoughts about delay and securities regulation, I want to make the point in a more oblique way, by explaining briefly how the idea of “Don’t Blink” applies to the writing of this article.

Professor James Fanto first contacted me about the Pomerantz lecture during mid-September 2010, and we agreed on a lecture date of March 15, 2011. That gave me six months to prepare for the talk. We discussed possible topics, and I promptly did nothing about the lecture for nearly two months.

The topics went into my deep subconscious, where they brewed until early November 2010, when I received an e-mail from Elizabeth Alper of Brooklyn Law School requesting that, by November 15, I confirm the title and write a brief description of the topic for publicity materials. The topic bubbled in my brain for a while. I did some research, and generally spent more time reading about the financial crisis and the flash crash. I wrote a paragraph about the talk, and finally signed off on the description on December 3, 2010, more than two weeks late. (I don’t want to mention how long I took to complete speaker permission and expense forms; that is just too embarrassing.)

I was still nowhere near starting to draft this article, though I was reading and researching the two topics more, and I was getting a better idea of what I would cover. I hired two research assistants to help me gather background materials for each topic.

Then, after the winter break, on February 9, 2011, Jim Fanto sent me the following gentle e-mail reminder: “What do you think? The talk is on Mar. 15. Kristin and Kent are
expected to give about 5–10 minutes each of comments. Can you give them something two weeks before? One week before?” I responded that I would send them something by March 1, stating that “I’m working on it and will get all of you something as soon as I can.” I read and researched more, and I thought more. I began to outline some of my thoughts.

March 1 passed and I still hadn’t begun drafting the article, though I now had a decent idea of what I planned to say and I understood the details about the financial crisis and the flash crash reasonably well. I was scheduled to give a talk to my law school’s board of visitors on March 4th, so I decided to discuss the timing of the Pomerantz lecture and how I hadn’t yet started writing the article in the context of a larger book project on the role of delay in decision making (which I had also barely started, and which was due in a few months). Later that day, Elizabeth Alper sent me a reminder e-mail that the lecture would be on March 15. Then, I shifted gears and worked furiously for a week. On March 10, I circulated a draft of the talk, which I continued to edit during the remaining days. Jim Fanto, Kent Greenfield, and Kristin Johnson were all gracious enough not to mention that I had given them only a few days to prepare a response to my draft (though, in my defense, they had a general idea of what I would cover well before that, and I believe they would not have begun preparing their comments until after March 10 in any event).

Next, immediately after the talk, I met Shawna MacLeod, the Editor-in-Chief of this law review. A week later, as promised, she sent me a detailed note about timing, giving me the choice of sending a first round draft on May 27 or a final manuscript on July 15. I chose the latter. Then I promptly did nothing for a month. She sent a follow-up e-mail on June 20, with eight apparently strict deadlines that would follow soon after I delivered the manuscript.

After the Fourth of July weekend, I finally went through all the materials my research assistants had gathered, and I spent the next eleven days finishing the research and writing of the article. The editorial process then went smoothly, and we approved the final manuscript on November 11.

I include all of this detail for two reasons. First, I know from discussions with many other academics that my various delays, though they might seem like irresponsible procrastination to people with real-world jobs, actually are consistent with a common and reasonable approach to scholarly writing. If I had written this article right away, I
would have missed many important details, thoughts, and research. I would not have had the opportunity to let the ideas brew for several months before putting them down on paper. Waiting until the last minute isn't always bad; it is often precisely what we should be doing: taking as long as we possibly can to consider a research project and then finishing at the last possible moment at the highest possible speed.

Second, I want to set forth the details of this account to remind academics that, although the Internet and the temptations of publishing in speedier venues are attractive alternatives to the slower pace of traditional academic articles and books, there are benefits to longer-form, longer-term writing that these other media do not have. The academics' comparative advantage is to take more time, to think through complex issues more deeply than others. Given the increasing speed of other approaches to writing, more leisurely-paced scholarship in academic articles and books is increasingly important today. That isn't to say that academics should avoid writing for other media, but rather that there is a special place for delayed thinking.

If the snooty lawyer confronting Abe Pomerantz in the story at the beginning of this article had taken more time to think, he would not have insulted Abe—he would have thanked him for suing his clients. Many lawyers, plaintiff- and defense-side alike, make judgments about the other side based on limited information. We all are prone to biases and cognitive error. But a fairer, more experienced counsel would have understood that securities litigation can and often does serve an important social purpose—deterring the fraudulent conduct that threatens investors and markets. Or at least he would have understood that without plaintiffs' lawyers he would not have a job.

If we learn no other lesson from the recent financial crisis and the flash crash, hopefully lawyers and regulators—and academics—will understand that there are dangers associated with snap decisions, and that we should include timing and delay among the factors that matter in regulating financial markets.
TRIBUTE

DAVID TRAGER: JURIST

Jeffrey Brandon Morris†

Earlier this year, the federal judiciary and the City of New York lost an able judge and one of the city's most public-spirited citizens. As a former United States Attorney for the Eastern District of New York, Judge David G. Trager took over what historically had been a patronage-driven office and transformed it into one that is highly professional and motivated to strike political corruption. As dean of Brooklyn Law School for over a decade, Trager was the central figure in transforming that institution from a local to national law school.

While I have written elsewhere about the United States District Court for the Eastern District of New York, as well as its judges, I cannot pretend to be completely objective in this essay, although I have attempted to be. I was a Visiting Professor at Brooklyn Law School during David Trager's deanship, and he left an indelible imprint on my career as a teacher and scholar. As dean, David Trager was a memorable personality. Large in size, he dominated any room he walked into by the strength of his personality. He propelled Brooklyn Law School forward with his vision—and with his ability to take the law school in new directions by anticipating and overcoming opposition through negotiation. In the end, much of his success can be attributed to the fact that he was able to persuade others that his motivation was truly not personal aggrandizement, but rather the improvement of the institution.

† Professor of Law, Touro Law School.
1 See, e.g., United States v. Margiotta, 688 F.2d 108 (2d Cir. 1982).
3 Id.
When he was dean, David Trager, often without being asked, was of enormous assistance to the careers of others. When the recipient of such gratuitous assistance attempted to express his or her appreciation, it tended to be received with what can only be described as an “embarrassed growl.”

The demands upon a federal district judge are very different from those of a dean. While far too early to attempt a definitive assessment of Trager’s seventeen-year career from 1993 to 2011 as a district judge, one could at this point make a few observations about his work.

The District Court for the Eastern District of New York that Trager joined in 1994 was already a distinguished court. Although it had lost men of John F. Dooling, Jr., and Orrin G. Judd’s quality, the court already included judges of superb caliber—Jacob Mishler, Jack B. Weinstein, and Eugene Nickerson, to name but a few. The jurisdiction of the district provided a rich docket: drug arrests were made at its airports; notorious organized crimes were committed within its five counties; lawsuits were frequently launched against its government; and complex commercial disputes were spawned in its boroughs. The government of the City of New York was an obvious target for lawsuits. Environmental actions in Nassau and Suffolk Counties, and copyright, trademark, and civil rights litigation were common. Trager had his share of these cases and much else.

I. TRAGER AND THE “CLASSIC” FEDERAL SPECIALTIES

The old specialties of the federal courts were well represented on the Eastern District’s docket and on Trager’s. Among those specialties were admiralty, bankruptcy, and criminal cases. Judge Trager had no “blockbuster” cases in these areas, but a brief discussion of the kinds of cases he adjudicated helps to remind us of the work federal judges have done and do.

A. Admiralty

Admiralty and maritime jurisdiction was a major reason for the creation of the federal district courts in 1789. After

---


5 Act of September 24, 1789, 1 Stat. 73. See, e.g., JEFFREY B. MORRIS, FEDERAL JUSTICE IN THE SECOND CIRCUIT 7 (1987).
Congress established the Eastern District of New York in 1865, the New York ports brought interesting and important questions arising out of collisions, groundings, and sinking of vessels, as well as suits brought by sailors and longshoremen for wages and physical injuries, to the federal courts.

While admiralty and maritime cases have declined in number—mirroring the decline of water shipping more generally—cases involving vessels, sailors, and longshoremen are still found in the work of the district courts. In spite of resemblances to the laws of torts and workers' compensation, admiralty and maritime law evoke a world and vocabulary of their own. Judge Trager was no stranger to the lingo:

The vessel contained four cargo holds (or hatches) with each hatch containing two levels—a tweendeck (upper cargo storage area) and a lower hold. . . .

Attached to the fore and front walls of each hold is a ladder leading up from the tweendeck to the main deck. About one meter to the left or the right of the tweendeck ladder, in the floor of the tweendeck, is an access door (or escape hatch). The access door cover can be lifted up by hand and opens toward the wall where it can be fastened to the wall to keep the access door open. . . .

Under the Longshore and Harbor Workers Compensation Act, a comprehensive federal workers' compensation program, “[t]he injured longshoreman is entitled to. . . benefits regardless of fault,” but “[t]he injured longshoreman's employer,” usually an “independent stevedore. . . is shielded from any further liability.” But if the injury was caused by the negligence of a vessel, the longshoreman can sue the vessel's owner.

Trager was quite sympathetic to seamen. In McMillan v. Tug Jane A. Bouchard, Judge Trager considered the law of maintenance and cure when a seaman claimed injury to his

---

6 MORRIS, supra note 4, at 7, 19.
9 Id. at *10-11 (citation omitted).
11 Dating back to at least the twelfth century, the law of maintenance and cure makes the ship owner responsible for paying maintenance and cure following any injury or sickness incurred by a seaman while in the owner's employ, whatever the cause.
back caused by lifting a shackle and line on the tug. As a result, the plaintiff missed some work. The defendant claimed that McMillan had an undisclosed, pre-existing back injury. His prior medical history also indicated neck and shoulder injuries and a history of valium use to control muscle spasms. Furthermore, rumor circulated that McMillan sought replacement on an upcoming fourteen-day hitch. When that request was turned down, McMillan said something equivalent to, “One way or another I am getting off the boat.”

Yet, employing a Court of Appeals decision favorable to the seaman plaintiff, Trager held that McMillan had a good-faith basis for withholding the information because McMillan missed only five days of work in 1976 and did not regularly use valium for his back injury. At worst, Trager wrote, McMillan’s belief was “an honest failure to reveal a prior medical condition.” Trager also indicated in a footnote that “it would be reaching to conclude from [McMillan’s] statement that some months earlier he had fraudulently concealed his pre-existing medical conditions.”

In the end, all doubts were resolved in favor of the seaman. As a result, Trager held that the defendant had terminated McMillan’s maintenance and cure too early. Even though McMillan stayed, rent-free, with family and friends during his recuperation, Trager did not deduct from McMillan’s judgment any savings or earnings accrued while off work (for McMillan had no other way of supporting himself). Trager held, “[A] shipowner cannot escape its liability for maintenance by forcing an injured seaman to involuntarily seek the financial support of family and friends during his or her convalescence.” But Trager did not award punitive damages and attorneys’ fees because the owner’s conduct was neither callous nor

“Maintenance” is the sum of money sufficient to provide food and lodging for the injured seaman during his or her convalescence and until he or she reaches the point of maximum medical cure. “Cure” consists of payments for all aspects of the seaman’s medical care until he or she reaches maximum medical cure.

Id. at 459 (citations omitted).

12 Id. at 456.
15 Id. at 461.
16 Id. at 461 n.11.
17 Id. at 465-66.
18 Id. at 465.
recalcitrant.\textsuperscript{19} The owner had made an erroneous decision, but one that had a "good faith basis."\textsuperscript{20}

B. Bankruptcy

The federal courts’ involvement in bankruptcies dates back to the beginning of the nineteenth century. The Bankruptcy Amendments and Federal Judgeship Act of 1984\textsuperscript{21} established the bankruptcy court as a unit of the district court; accordingly, the district court refers bankruptcy cases and proceedings to the bankruptcy court.

Trager handled a number of bankruptcy matters. In one, he gave short shrift to the debtor’s filing because the debtor failed to abide by the bankruptcy court’s repeated orders—first, to make adequate post-petition mortgage payments, and second, to comply with rental obligations. Also, the debtor could not propose a satisfactory repayment plan.\textsuperscript{22} In another case, Trager held against non-lawyer, fly-by-night bankruptcy preparers who engaged in the unauthorized practice of law and violated the bankruptcy law provision prohibiting fraudulent, unfair, or defective acts.\textsuperscript{23}

C. Criminal Cases

Considering Trager’s distinguished career as a U.S. Attorney, it could have been anticipated that he would make a mark as a judge in the field of criminal law. There was no question of Trager’s interest, ability, and comfort in the field. In turn, his opinions made the factual implications and legal stakes of each case clear. To reach these decisions, Trager used his ability to discern the tactics and strategies of attorneys.\textsuperscript{24}

Like his colleagues, Judge Trager heard many habeas corpus petitions from state prisoners. Generally, he denied these petitions with relatively brief opinions. Although usually cloaked

\textsuperscript{19} Id. at 466-67, 469.
\textsuperscript{20} Id. at 467.
\textsuperscript{24} Trager did not have many organized crime and political corruption cases as a judge. But see Russo v. United States, No. 04-CV-3871, 2007 U.S. Dist. LEXIS 62209 (E.D.N.Y. Aug. 23, 2007).
in a theory of ineffectiveness of counsel, the issues varied in habeas petitions brought before him. Judge Trager considered petitions concerning whether there was sufficient evidence to support the essential elements of a conviction;25 whether a line-up was too suggestive;26 and whether a petitioner's challenge rested on procedural or substantive grounds.27

Trager's ability to study a criminal case, determine its strengths and weaknesses from the point of view of both sides, and foresee and assess the tactics of the attorneys allowed him to decide more authoritatively than most judges. Vaknin v. United States28 is an excellent example. Vaknin was an action to vacate a sentence and conviction.29 The case dealt with a complex criminal scheme that involved obtaining commissions from a wireless carrier through the sale of customer information.30 The defendants used the information to fraudulently renew customer contracts and buy and sell wireless phones. Vaknin sought reversal because the government failed to disclose evidence favorable to the defendant under Brady v. Maryland31—specifically, that Vaknin was not aware of the sale of customer information.32 In assessing the motion to vacate defendant Vaknin's sentence, Trager ruled that the government's failure to disclose was not reversible error because the information would have simply confirmed what Vaknin already knew.33 Then, in dealing with the ineffectiveness of counsel claim, Trager demonstrated that there was no reasonable probability of Vaknin's case proceeding to trial: no competent counsel would have pursued that theory even if Vaknin's acquittal had been a possibility.34

Trager's effectiveness in handling facts is also evinced in an opinion he penned when he sat by designation on the

29 Id. at *1.
30 Id. at *5-9.
32 Vaknin, 2010 U.S. Dist. LEXIS 86254, at *41-42.
33 Id.
34 Even if this theory proved successful, the defendant's sentence would not have been mitigated. See id. at *53-54.
The defendant, Richardson, had been convicted of conspiracy to distribute five kilograms or more of cocaine. He argued that the evidence presented at trial had not proved a single overarching conspiracy but rather the existence of multiple distinct conspiracies for which he had not been charged. Writing for a panel, Trager lucidly and compellingly marshaled the evidence against Richardson, demolishing the theories raised by his defense. Richardson was, Trager wrote, “a ‘key man’ [who] direct[ed] and coordinate[d] the activities and individual efforts of various combinations of people.” Richardson “was not a spoke, but the hub of all the conspiratorial acts the government sought to prove at trial.” Judge Trager’s language is telling: this was not the kind of case where the government sought to convict a defendant who played a peripheral role in a vast conspiracy. Thus, the court, affirming the conviction, held that the government presented evidence sufficient to establish the common goal, underlying scheme, and overlap of participants. The evidence supported a reasonable conclusion that each co-conspirator worked with Richardson according to Richardson’s grand scheme. Trager’s ability to work with doctrine in criminal cases was also on display in Rivera v. Artus, a habeas case. In two footnotes, Judge Trager analyzed two lines of court of appeals decisions involving the suppression of statements made to police concerning both an unrepresented and a represented matter in the course of the same interrogation.

Trager also wrote several interesting opinions in criminal cases for the U.S. Court of Appeals for the Ninth Circuit. United States v. Paopao, for instance, posed the issue of a protective sweep. A police detective received a tip from a confidential informant that several suspected robbers of illegal gambling rooms were in a particular room in Honolulu. Arriving at the site and making an arrest, the officers performed a protective sweep.

---

35 United States v. Richardson, 532 F.3d 1279 (11th Cir.), reh’g en banc denied, 285 F. App’x 745 (11th Cir. 2008).
36 Id. at 1282.
37 Id.
38 Id. at 1286 (quoting United States v. Edouard, 485 F.3d 1324, 1347 (11th Cir. 2007)).
39 Id. at 1288.
40 Id. at 1285-86.
42 469 F.3d 760 (9th Cir. 2006).
43 Id. at 763.
sweep that took less than a minute.44 As the officers were leaving the apartment, one of them noticed an unzipped bag.45 Looking into the bag, one of the officers saw what he thought was the handle of a handgun and ammunition.46 He seized the bag, and the contents were later determined to be a gun, ammunition, a knife, and a black pouch of jewels.47

Writing for the court, Trager held that the defendant, Paopao, personally had no reasonable expectation of privacy in the game room and, therefore, had no standing to challenge the protective sweep.48 Indeed, if the judge ruled that Paopao had standing, the search would still have been held valid because the police had reasonable suspicion from the informant's tip.49 While Paopao could contest the seizure of his bag, the search of an object in plain view is constitutional.

In another Ninth Circuit criminal case for which Trager wrote the opinion, the panel affirmed the conviction of a thirty-five-year-old man who had attempted to persuade and entice a minor—an undercover FBI agent in an online chat room—to engage in sexual acts.50 The court held that the trial court had acted within its discretion in admitting evidence tending to prove the defendant's intent and modus operandi.51 Specifically, the trial court allowed evidence of the defendant's previous conviction for lewdness with two children, aged eleven and twelve, as well as evidence of complaints of similar behavior that had been made to America Online.52 Judge Trager held that its prejudicial impact was outweighed by its probative value.53

44 Id.
45 Id.
46 Id.
47 Id.
48 Id. at 764-65.
49 Id. at 765-67.
50 United States v. Cherer, 513 F.3d 1150, 1153 (9th Cir. 2008).
51 Id. at 1157.
52 Id. at 1156.
53 Id. at 1159. However, the distinguished philosopher-judge John Noonan disagreed with the sentence of almost twenty-four years in prison, which the court affirmed. The panel affirmed the sentence, finding that the trial judge had made no significant procedural error. In response to Judge Noonan's criticism, the panel opined on the district court's reasonableness "to conclude that a lengthy prison sentence was necessary to protect the public." Id. at 1160. Further, the panel held that the trial judge "was not required to consider the much more lenient sentences available for other violent and arguably more heinous sex offenses." Id. at 1161. Judge Noonan argued that the defendant's crime was "committed by words in the imaginary world of the chat room." Id. at 1162. In words reminiscent of Trager's colleague Jack B. Weinstein, see, e.g., United States v. Polizzi, 549 F. Supp. 2d 308 (E.D.N.Y. 2008), Noonan wrote: "The defendant is not a convenient abstraction such as 'a pedophile' but
Pro Se Cases

There are few federal judges who enjoy litigation brought pro se. There is no evidence that David Trager was such a judge. However, in only a few cases did he show exasperation.\textsuperscript{54} One such case seems to have begun when the pro se plaintiff allegedly forced his way into a sixteen-year-old tenant's home, which he owned.\textsuperscript{55} He was convicted of a misdemeanor, and he and his wife then filed a complaint, which was subsequently withdrawn. An amended complaint was then filed naming the original defendants—a state judge, Kings County District Attorney Charles Hynes, and two assistant district attorneys—along with the New York City Department of Corrections, Rent Guidelines Board, and Police Department, as well as the Criminal Court of Kings County and five other governmental offices.\textsuperscript{56} The plaintiff, George Pappanikolaou, alleged a conspiracy to convict him by tampering with the evidence.\textsuperscript{57} He also alleged abuse and discrimination claims against the Police Department, Department of Corrections, and the Rent Guidelines Board.\textsuperscript{58} Other deprivations were alleged in the thirty-one page complaint, which Trager described as “a lengthy narrative which rambled about conspiracy theories.”\textsuperscript{59}

Judge Jack Weinstein, among the most tolerant judges in dealing with pro se litigants,\textsuperscript{60} had dismissed the complaint.\textsuperscript{61}

\textsuperscript{54} One such case was brought by a homeless New Yorker who sought five hundred million dollars as well as other relief, though that was expressed incoherently. Belton v. City of New York, No. 05-CV-2937, 2005 WL 2133593 (E.D.N.Y. Sept. 1, 2005). The plaintiff was suing for five hundred million dollars because, he claimed, the New York City Transit Police at the Clark Street subway station (near to the main courthouse of the Eastern District) had “failed to keepsake his belongings” when he was arrested. Id. at *1. The five hundred million dollars of missing possessions was for an emergency crisis kit, one fifty dollar special comforter, and two twin size blankets. Id. at *2. Trager construed the complaint as a section 1983 action for deprivation of property without due process of law, but dismissed it because of the lack of any official police policy or custom. He did, though, point the plaintiff to state law causes of action for negligence, replevin, and conversion. Id. at *3.


\textsuperscript{56} Id. at *3.

\textsuperscript{57} Id. at *3-4.

\textsuperscript{58} Id.

\textsuperscript{59} Id. at *4.


\textsuperscript{61} Pappanikolaou, 2005 U.S. Dist. LEXIS 39201, at *5-6.
But the Second Circuit reversed to the extent that there were claims against the New York City defendants for deliberate indifference to the plaintiff’s medical needs and that there were claims under the Americans with Disabilities Act. Weinstein then recused himself from the case, but the plaintiff’s extensive litigation continued.

After a conference before a magistrate judge, Pappanikolaou filed a 118-page complaint which dealt largely with his arrest, incarceration, and prosecution, although the Second Circuit and the magistrate judge had already indicated that these issues were off limits. The plaintiff then filed a third complaint, described by Judge Trager as “a fifty-nine page, single-spaced document that was yet again filled with the same rambling and incomprehensible allegations that were present in the first two amended complaints.” This was followed by the fifth status conference, which, in turn, was followed by an order containing explicit instructions concerning how the complaint should be properly amended. Consequently, the plaintiff filed his fourth complaint. It was forty-three pages long and named numerous new defendants (among them, the New York Governor George E. Pataki, State Attorney General Eliot Spitzer and Chief Judge Judith Kaye). Despite the magistrate’s instructions, the plaintiff continued to allege that the Court of Appeals for the Second Circuit had already dismissed the case. Trager said that the new complaint “nonsensically ramble[d].” When the City and State moved to dismiss, the plaintiffs filed a 229-page opposition.

Yet, even after all this, Trager was unwilling to dismiss the case for failure to comply with a court order or for failure to comply with the pleading requirements of Federal Rule of Civil Procedure 10(b). Trager did, however, dismiss for failure to comply with Rule 8’s “short and plain statement” of claims requirement. The plaintiff had otherwise failed to state a valid claim for relief under section 1983, the Civil Rights statute, or the Americans with Disabilities Act. The dismissal was without leave to amend.

---

62 Id. at *6-7.
63 Id. at *7.
64 Id. at *8.
65 Id.
66 Id. at *10.
67 Id. at *10, *12 n.2.
68 Id. at *22.
69 Id. at *13 & n.3.
70 Id. at *15-17, *26-28.
There was also the action brought by Israel Valle, who, in 2004, Trager enjoined from filing any new action “seeking in forma pauperis status without obtaining leave of court.”

Valle sought leave to file five actions in 2002, two in 2006, nine in 2008, and four in 2009. In 2010, Valle filed three more actions alleging in the first two that numerous state court judges were involved in a criminal enterprise. In the third action, he sued district judges, a U.S. magistrate judge, the clerk of the court, and various other court personnel.

Judge Trager, one of the named defendants, noted that—as in prior attempts—he would not recuse himself, “as there [was] no basis to do so and it would only [have] provide[d] Valle with another person to sue should [the] case [have been] reassigned to another district judge.” He added: “In essence, Valle has formulated a template that he repackages with a different caption depending on which court has dismissed his latest attempts to file frivolous litigation.” As an order to show cause for sanctions was pending, Judge Trager simply denied leave to file the three actions and certified that any appeal of his order would not be in good faith.

Trager had a special connection to another pro se case, Leeds v. Meltz. In Leeds, the plaintiff was an attorney, an alumnus of the City University of New York School of Law.

Leeds sued the acting dean of CUNY School of Law and three student editors of the school's student newspaper because the newspaper had refused to publish a classified advertisement in which Leeds sought material to discredit faculty and administrators for a civil rights action against the school. Leeds claimed that his First and Fourteenth Amendment rights had been violated by the editors' refusal to publish the ad; the plaintiff argued that the student editors were acting under color of state law and were thus violating his constitutional rights.

Trager had no doubt that the student editors were not state actors. And in fact, Leeds's argument—reliant upon an

---

72 Id. at *1.
73 Id.
74 Id.
75 Id.
76 Id.
77 Id. at *2.
79 Id. at 147.
80 Id.
81 Id. at 147-48.
inference that the editors’ decision was influenced by the public school’s administration—was undermined by Leeds’s own complaint, which alleged that “the administration retaliated against the paper by cutting its budget and access to facilities.”

The fact that a publication is sponsored by a state agency, by itself, is insufficient to establish state action. The newspaper at issue was sufficiently independent, and its actions could not be characterized as state action. Indeed, Leeds had “alleged no facts from which it could plausibly be inferred that the editors’ actions were ‘fairly attributable’ to the law school administration.”

Former Dean Trager pointed out that it was “difficult to believe that all three student editors would have supinely accepted the alleged intimidation of the school administration.” “[S]tudents being students, more than likely they would have at least complained to some of their student colleagues about the administration or faculty pressure and the issue, in the natural course of events, would inevitably have become a subject of student discussion at the law school.”

The Leeds opinion also rested on a broad principle: “As most student publications are generally without substantial resources, baseless actions can impair the First Amendment rights of the publications and their student participants.” Further, if complaints could be sustained “on the flimsy basis present here,” students “would be unwilling to run the risk of having to pay substantial legal fees to defend themselves from unjustified legal actions and would forego the opportunity to participate in this activity.”

---

82 Id. at 148.
83 Id. at 148-49.
84 Id.
85 Id. at 149 (citations omitted).
86 Id. at 150.
87 Id.
88 Id. at 149.
89 Id. at 149-50. In Wasser v. New York State Office of Vocational & Educational Services, Trager dealt with a lawsuit brought by a Brooklyn Law School graduate, a quadriplegic suing pro se a New York State agency which provides services to those with disabilities (VESID). 683 F. Supp. 2d 201, 203-04 (E.D.N.Y. 2008), aff’d, 602 F.3d 476 (2d Cir. 2010). The relationship between Wasser and VESID had “been contentious from the start.” Id. at 204. VESID had reimbursed Wasser for part of his legal education at Brooklyn Law School, but only up to the cost he would have had if he had gone to CUNY Law School, which was much less than at Brooklyn. Id. at 219. Wasser argued that the one law school sponsored by the state, the State University of New York at Buffalo Law School, and CUNY were inappropriate for him because of his disability and his career goals. Id. at 220. Wasser had actually begun law school while Trager was still dean and he cited the “receptive atmosphere” at Brooklyn Law School. Trager ruled against Wasser on this claim (indeed, on all his claims). Id. at 220, 225.
E. Trager’s Decisions While Sitting with Courts of Appeal

The custom of designating district judges to sit with courts of appeal for short periods of time not only acclimatizes district judges to the different concerns of appellate judges, but it allows district judges the luxury to focus on more specific issues of law than they would normally be allowed. David Trager sat by designation on the Courts of Appeals for the Second, Ninth, and Eleventh Circuits. Appellate work was an important part of his legacy.

Cooper v. Meridian Yachts, Ltd. was an Eleventh Circuit case which raised difficult conflict-of-law issues. Following settlement of a ship captain’s maritime tort action for injuries he sustained as a result of an alleged defective ship food lift, a number of parties filed third-party complaints seeking to receive indemnity and contribution from the ship’s builder and designer and from an associated U.S. venture. The ship, the *Meduse*, was built in the Netherlands and registered in the Cayman Islands. The shipbuilder and ship designer principally operated in the Netherlands, but conducted business in the U.S in a joint venture. The ship’s purchaser was a business entity organized under the laws of the British Virgin Islands, and the captain was employed by a British Virgin Islands corporation. The ship’s management company (Vulcan Manager) and crew were incorporated in the State of...
Washington. The entity that constructed the yacht allegedly operated in the United States.

The district judge had dismissed the third-party plaintiffs’ claims using the Dutch Statute of Repose. As interpreted by the Eleventh Circuit, however, the case had two sets of issues raising distinct choice-of-law concerns over the third-party claims: one set of issues concerned a Dutch choice-of-law clause and Dutch limitation-of-liability provision; the second set of issues involved whether Dutch law, federal maritime law, or a third jurisdiction’s law governed the third-party claims. In his opinion, Trager indicated that the district court had been hampered by limited information about Dutch law and was completely uninformed as to the laws of the Cayman and British Virgin Islands.

This is not the place to trace the contours of Trager’s twenty-nine page opinion, which deals with, at least as calculated by this author, eleven issues and subissues. The opinion applies three different sets of law in different places: Dutch law, federal maritime law, and Florida law. The district judge, who apparently had not considered all the aspects of this complex puzzle, was affirmed in part and reversed in part. Trager’s opinion on the choice-of-law and conflict-of-laws issues suggests not only how comfortable he was in this arcane area, but also how much he enjoyed it.

Trager also wrote for the Eleventh Circuit Court of Appeals in an interesting civil rights suit, Young Apartments, Inc. v. Town of Jupiter. The case involved section 1983 and breach-of-contract claims brought by the owner of an apartment complex. The plaintiff alleged that the defendants, a town and town official, had harassed and discriminated against him and other Hispanics by refusing to provide them affordable housing.

---

95 Id. at 1159.
96 Id. at 1161.
97 Id. at 1180.
98 Trager handled an unusual number of conflict-of-laws and forum non conveniens issues. See, e.g., Giaguaro v. Amiglio, 257 F. Supp. 2d 529 (E.D.N.Y. 2003) (a complicated dispute over the sale of Italian canned, peeled tomatoes); see also Gerena v. Korb, 617 F.3d 197 (2d Cir. 2010) (where Trager wrote the opinion for the Second Circuit in a case involving a civil suit brought by a Yale student who claimed she was sexually assaulted in her dormitory room by another Yale student after a Yale sponsored back-to-school event).
99 529 F.3d 1027 (11th Cir. 2008).
100 Id. at 1032.
101 Id.
town had adopted an overcrowding ordinance in response to the increasing number of immigrant workers. The ordinance was purportedly enforced through “excessive and selective” housing inspections that targeted landlords housing Hispanics. The town inspected the plaintiff’s apartment complex and cited it for violations of the ordinance as well as physical defects on the property. The town condemned some units in the building which led to the cancellation of the contract to sell the building.

The district court dismissed most of Young’s complaint. The appeal centered on the holding that Young lacked standing to bring a race-based discrimination suit as well as the dismissal of the complaint against two town officials. The court of appeals reversed on the standing issue because Young Apartments had suffered financial injuries. That is, “Young could allege that it was injured by Jupiter’s discriminatory actions regardless of whether such claims might also vindicate the rights of its immigrant tenants.” Nor was Young barred by prudential principles of third-party standing from advocating on behalf of its customers against discriminatory actions that interfered with a business relationship. Indeed, “Young Apartments [was] uniquely positioned to assert claims on behalf of its Hispanic residents.” The court of appeals also reversed as to the proper level of review. The trial judge had used a rational basis test. However, “because Young Apartments [had] standing to attack the ordinance as racially discriminatory,” the proper test was strict scrutiny. The trial judge was affirmed in part and the case was remanded.

A third opinion Trager penned for the Eleventh Circuit also deserves mention. In Hanley v. Roy, Judge Trager encountered the Hague Convention on the Civil Aspects of International Child Abduction. Grandparents from Ireland, the Hanleys—who had been named testamentary guardians of their daughter’s children—argued that their son-in-law (Roy) had wrongfully removed their grandchildren to Florida.

---

102 Id. at 1033.
103 Id.
104 Id. at 1035.
105 Id.
106 Id. at 1038-39.
107 Id. at 1040.
108 Id. at 1044.
109 Id.
110 485 F.3d 641 (11th Cir. 2007).
111 Id. at 643.
district court dismissed the petition. The court of appeals, with Trager writing, reversed.

After their then-divorced daughter had been diagnosed with cancer, the Hanleys bought her a house in Ireland and moved in to help care for her children; Roy also moved into the house. Before her death, the daughter named her parents the testamentary guardians of the children. Roy and the children continued to live with the Hanleys for four-and-one-half years after the children’s mother died. Then he “suddenly moved the children from Ireland to Florida without the Hanley’s knowledge or consent, leaving only a note behind.”

The court of appeals held that under Irish law, “testamentary guardians” must act jointly with the father as guardians. If the father objects, the testamentary guardians are entitled to “seek a court determination enforcing their joint-guardianship rights.” The district court held that the status of testamentary guardians was not enough to accord the Hanley’s “rights of custody” under the Hague Convention and under the interpretation of the Convention by Irish courts. The grandparents were therefore entitled to seek a court determination enforcing their joint guardianship rights. The court of appeals then struck at the heart of the case: “[P]ermitting the very act which the Convention seeks to prevent—namely, flight—to constitute a construction to terminate the Hanley’s ‘rights of custody’ would make a mockery of the Convention.” The court of appeals directly ordered Roy to return the minor children to Ireland for proper proceedings, and, if he did not do so promptly, the district court was to order the minor children turned over to the Hanleys. Judge Trager’s opinions while sitting by designation on the courts of appeals suggest that he would have been comfortable and well qualified to sit fulltime as an appellate judge.

\[112\] Id. at 644.
\[113\] Id. at 643.
\[114\] Id. at 644.
\[115\] Id.
\[116\] Id. at 645-46.
\[117\] Id. at 646.
\[118\] Id. at 650.
\[119\] Id.
II. AREAS OF LAW WHERE JUDGE TRAGER MADE SPECIAL CONTRIBUTIONS

As judge, David Trager made significant contributions in two classic specialties of the federal courts—antitrust and copyright. He also handled two cases of broad interest—one relating to the events of September 11, 2001, the other an unusual criminal case involving terrorism, which deserves separate treatment.

A. Antitrust

Trager handled two important antitrust cases—one case involving the widely used antibiotic Cipro and another against Chinese manufacturers of Vitamin C. In the Cipro litigation, Trager supported the longstanding position of the pharmaceutical industry that settlement of suits between generic and branded pharmaceutical companies is not presumptively anticompetitive. At issue was a patent holder’s reverse payment to generic-drug manufacturers to avoid a patent challenge.\[120\]

In Cipro, Barr Laboratories and the Rugby Group challenged as anticompetitive agreements between Bayer AG, its American subsidiary, Bayer Corporation (the brand manufacturer of Cipro), and a group of generic manufacturers. Barr argued that the agreements, which were lawsuit settlements, violated antitrust law under the Drug Price Competition and Patent Term Restoration Act of 1984 (Hatch-Waxman Act).\[121\] In 1991, Barr filed a certification for generic Cipro claiming Bayer’s patent was invalid. In response, Bayer sued for infringement. The suit stayed approval of generic Cipro by the Food and Drug Administration for thirty months. Bayer and Barr ultimately settled their dispute to avert a costly trial. Under the settlement, Barr and Rugby acknowledged the validity of the Cipro patent, and Barr agreed to amend its certification so that it could market generic Cipro only after Bayer’s patent expired. For its part, Bayer agreed to, among other things, license Barr and Rugby to market a competing ciprofloxacin product six months before the Bayer

---

\[120\] See In re Tamoxifen, 277 F. Supp. 2d 121 (E.D.N.Y. 2003), aff’d, In re Tamoxifen Citrate Antitrust Litig., 466 F.3d 187 (2d Cir. 2006).

The agreement spawned multiple putative class actions in various state courts that were removed to federal court. The Judicial Panel on Multidistrict Litigation transferred those cases to Trager, who remanded some of them back to state courts and retained others with institutional plaintiffs.123

On May 20, 2003, Trager ruled that the Cipro settlement did not violate the antitrust laws. He noted that the challenged settlements resolved the entire patent dispute without creating a “bottle neck” for subsequent generics. The real question for antitrust purposes was whether the settlement would hinder lawful competition.124 Trager stressed that the American legal process encourages the settlement of lawsuits. A contrary rule might lead, he said, to less investment in research and development. He therefore denied the plaintiff’s motion for partial summary judgment. In 2005, Trager rendered judgment for the defendants on similar grounds.125 He held that it would be “inappropriate for an antitrust court, in determining the reasonableness of a patent settlement agreement to conduct an after-the-fact inquiry into the validity of the underlying patent. Such an inquiry would undermine any certainty for patent litigants seeking to settle their disputes.”126 Thus, Trager held, the settlement had not violated the antitrust laws because the settlement excluded no competition beyond the exclusionary scope of the patent. The appeal to the Court of Appeals for the Second Circuit was transferred to the Federal Circuit, which in 2008 affirmed Trager’s position.127

The second of the antitrust cases possessed significant international implications. The Vitamin C antitrust action was

---

126 Id. at 530.
127 In re Ciprofloxacin Antitrust Litig., 544 F.3d 1323, 1327 (Fed. Cir. 2008). The Second Circuit also backed Trager, although the panel invited the appellants to petition for rehearing en banc. The full court proved not to be interested. Ark. Carpenters Health & Welfare Fund v. Bayer AG, 604 F.3d 98 (2d Cir.), reh’g en banc denied, 625 F.3d 779 (2d Cir. 2010), cert. denied, 131 S. Ct. 1606 (2011).
notable because the Chinese government presented its views as an amicus.\textsuperscript{128} The multidistrict litigation consisted of separate class action suits brought by American individuals and entities that had purchased Vitamin C from Chinese manufacturers. The plaintiffs alleged that Chinese companies had met with the Association of Importer and Exporters of Medicines and Health Products of China and formed a cartel.\textsuperscript{129} As a result, the Chinese market share on Vitamin C increased and the price nearly tripled.\textsuperscript{130} Plaintiffs further alleged that the Chinese manufacturers met in June 2004 and agreed to raise prices by shutting down production and restricting exports to the United States.\textsuperscript{131} The Chinese defendants argued that comity and the act-of-state doctrine (a foreign government may not be questioned in another nation’s courts for actions within its borders) applied.\textsuperscript{132} Indeed, the AIEMC was associated with the Chamber of Commerce of Medicines and Health Products Importers and Exporters, an entity under the direct control of the Chinese Ministry of Commerce.\textsuperscript{133}

In June 2006, the Chinese manufacturers sought a stay of discovery pending their motions to dismiss. They argued that they were going to secure a dismissal on a theory akin to the act-of-state doctrine—that they could not be held liable for conduct in violation of U.S. antitrust laws because the Chinese government had compelled them to engage in their conduct.\textsuperscript{134} The Chinese government then filed an amicus curiae brief arguing that the Chinese manufacturers were compelled under Chinese law to collectively set a price for vitamin exports. Trager responded: “The Chinese government’s appearance as amicus curiae is unprecedented. It has never before come to the United States as amicus to present its views. This fact alone demonstrates the importance the Chinese government places on this case.”\textsuperscript{135}

Nevertheless, Trager reasoned that the documents the Chinese submitted as attachments to their brief, if credited, would present a complex interplay between the Chinese government and the defendants, where defendants’ independence

\textsuperscript{128} In re Vitamin C Antitrust Litig., 584 F. Supp. 2d 546, 552 (E.D.N.Y. 2008).
\textsuperscript{129} In re Vitamin C Antitrust Litig., 237 F.R.D. 35, 36 (E.D.N.Y. 2006).
\textsuperscript{130} Vitamin C Antitrust Litig., 584 F. Supp. 2d at 548-49.
\textsuperscript{131} Id. at 549.
\textsuperscript{132} Id. at 550.
\textsuperscript{133} Id. at 551.
\textsuperscript{134} Magistrate Judge James Orenstein denied the application to stay discovery. Vitamin C Antitrust Litig., 237 F.R.D. 35.
\textsuperscript{135} Vitamin C Antitrust Litig., 584 F. Supp. 2d at 552.
in making pricing decisions was difficult to determine at that stage in the litigation.\footnote{Id. at 556.}

But in November 2008, Trager rejected the claim by the Chinese companies to dismiss the case on the grounds that their price-fixing activities were compelled by the Chinese government. He held that the record was too ambiguous to foreclose further inquiry into the voluntariness of defendants' actions. While the Ministry brief was entitled to “substantial deference,” it would not be taken as conclusive evidence of compulsion.\footnote{Id. at 557.}

B. Copyright

Trager wrote a number of interesting opinions in the area of copyright law. One was Cosmetic Ideas, Inc. v. IAC/Interactive Corp.\footnote{606 F.3d 612 (9th Cir. 2010).} In that case, the U.S. Court of Appeals for the Ninth Circuit considered whether registration of a copyright takes place when the Copyright Office receives a completed registration application or only after the Copyright Office issues a certificate of registration. Four courts of appeals and several district courts had considered the issue and were divided equally.\footnote{Id. at 615 n.4.} In addition, there was a recent Supreme Court decision bearing on the issue.\footnote{Reed Elsevier, Inc. v. Muchnick, 130 S. Ct. 1237 (2010).}

The Ninth Circuit panel found no guidance from the language of the relevant clause in the statute or the law as a whole.\footnote{Cosmetic Ideas, 606 F.3d at 617-18.} However, looking at the purpose of the law, the panel concluded that the application approach better fulfilled Congress's purpose to provide broad copyright protection and maintain a “robust federal register.”\footnote{Id. at 618.} The Ninth Circuit thus sided with the Fifth and Seventh Circuits in holding that receipt by the Copyright Office of a complete application satisfies the registration requirement.\footnote{Id. at 616, 621.}

\footnote{Id. at 616, 621. One fine Trager opinion written for the Eleventh Circuit on the question as to whether the Copyright Act accords a magazine a privilege to produce a digital compilation containing exact images was vacated by an en banc court, although the latter court reached the same result. Greenberg v. Nat’l Geographic Soc’y, 488 F.3d 1331 (11th Cir.), reh’g en banc granted, 497 F.3d 1213 (11th Cir. 2007), rev’d en banc, 553 F.3d 1244 (11th Cir. 2008). At issue was whether the National Geographic
C. Extraordinary Rendition

One Trager decision that aroused considerable criticism involved the extraordinary rendition of a Canadian citizen who passed through John F. Kennedy Airport while attempting to catch a connecting flight to Canada.\footnote{144} Maher Arar, on his way home from vacationing in Tunisia, was intercepted by American officials who believed he was a terrorist.\footnote{146} Detained, Arar was interrogated, placed in solitary confinement, and given the opportunity to voluntarily return to Syria (which he refused).\footnote{147} Arar was permitted only a single meeting with counsel, and was then branded “clearly and unequivocally a member of al Qaeda,” flown to Jordan, and turned over to Syrian authorities.\footnote{148} Neither the Canadian Consulate nor Arar’s attorney was informed before he was taken from the United States.\footnote{149}

During his ten-month detention in Syria, Arar was tortured\footnote{149} (as Arar warned American officials he would be\footnote{150}). He was forced to sign a confession stating that he had participated in terrorist training in Afghanistan, which he later denied.\footnote{151} His statements were apparently shared with the United States government. Ultimately, Syria released Arar, who returned to Canada.\footnote{152} Later, Arar—a telecommunications engineer who held both Syrian and Canadian citizenship—was completely exonerated by the government of Canada. Seven

\textit{Society} could release an unedited CD-ROM collection of its back issues without paying photographers royalties for their work. Trager’s panel and the en banc court held that the addition of new material to the work—an introductory sequence—did not take the revised collective work outside the privilege, for other portions were privileged (the reproduced issues of the magazines themselves (“Replica”) and the computer program (“program”)).

Trager also handled a large number of copyright-infringement suits brought by record companies throughout the United States in response to online distribution. The suits were brought against individuals in an attempt to combat and deter what record labels perceived as massive copyright infringement over the Internet. See generally UMG Recordings v. Lindor, CV-05-1095, 2010 U.S. Dist. LEXIS 20952 (E.D.N.Y. Mar. 8, 2010); Elektra Entm’t Grp., Inc. v. Torres, No. CV-07-640, 2007 U.S. Dist. LEXIS 92774 (E.D.N.Y. Dec. 18, 2007); UMG Recordings, Inc. v. Lindor, 531 F. Supp. 2d 453 (E.D.N.Y. 2007); UMG Recordings v. Lindor, No. CV-05-1095, 2006 U.S. Dist. LEXIS 83486 (E.D.N.Y. Nov. 9, 2006).

\footnote{145} Id. at 253.
\footnote{146} Id.
\footnote{147} Id. at 254.
\footnote{148} Id. at 252-54.
\footnote{149} Id. at 254.
\footnote{150} Arar v. Ashcroft, 585 F.3d 559, 586 (2d Cir. 2009) (dissenting opinion).
\footnote{151} Id.
\footnote{152} Arar, 414 F. Supp. 2d at 255.
months after Trager’s decision on rendition (February 2006), the Canadian government concluded that Canadian intelligence, under pressure to find terrorists, had passed on false warnings about Arar to the United States. The 822-page report of the commission established by the Canadian government found that Arar had no involvement in Islamic extremism and was not a security risk. The Prime Minister of Canada sent a letter of apology to Arar and his family along with payment of approximately $9.75 million. In June 2008, the Inspector General of the Department of Homeland Security told a congressional committee that he would not rule out that U.S. officials had violated U.S. laws.

In the meantime, Arar brought suit seeking declaratory relief and compensatory and punitive damages. Having been removed to Syria under the covert U.S. policy of “extraordinary rendition,” through which non-U.S. citizens were “sent to foreign countries to undergo methods of interrogation not permitted in the United States,” Arar alleged that the U.S. officials had violated (1) the Torture Victim Prevention Act (TVPA); (2) his Fifth Amendment rights when the defendants knowingly subjected him to torture and coercive interrogation in Syria; (3) his Fifth Amendment rights through his arbitrary and indefinite detention in Syria, including denial of access to counsel, the courts, and his consulate; and (4) his Fifth Amendment Rights because he had suffered “outrageous, excessive, cruel, inhumane and degrading conditions of confinement in the United States where he had been subjected to coercive and involuntary custodial interrogation and deprived of access to lawyers and courts.”

In the course of a lengthy opinion that resolved many (but not the most important) issues in a manner favorable to the plaintiff, Trager awarded no relief. He first held that Arar lacked standing for declaratory relief because the activity he was challenging was neither ongoing nor likely to impact him in the

---

153 Apparently Arar had acquaintances in Canada who were being investigated. Id. at 256 n.1.
154 The United States refused to cooperate with the Canadian inquiry. Arar v. Ashcroft, 532 F.3d 157, 162 (2d Cir. 2008).
156 Arar, 414 F. Supp. 2d at 256.
157 Id. at 257-58.
future. Then, after making a series of rulings for Arar regarding the application of the TVPA, Trager held that because Congress intended for the Act to be used as a remedy for U.S. citizens subject to torture overseas, it did not apply in Arar’s case.\textsuperscript{160}

With respect to compensatory and punitive damages, the claim for relief was based on the so-called Bivens remedy, one created by the Supreme Court for federal officials’ violations of the Fourth, Fifth, and Eighth Amendments.\textsuperscript{161} Once again, Trager resolved the preliminary questions favorably to Arar but refused to extend Bivens to overseas conduct because of Congress’s power over aliens and the “national security and foreign policy decisions at the heart of this case.”\textsuperscript{162} Extending Bivens in this sort of case, wrote Trager, “could significantly disrupt the ability of the political branches to respond to foreign situations involving our national interest.”\textsuperscript{163} Further, Trager believed that most if not all judges lacked the experience or the background “to adequately and competently define and adjudge the rights of an individual vis-à-vis the needs of officials acting to defend the sovereign interests of the United States.”\textsuperscript{164} The task of balancing individual rights against national security concerns was one, he felt, courts should not take on without the guidance or authority of the coordinate branches.\textsuperscript{165} The claim that Arar was deprived of due process rights during his period of domestic detention still might, Trager wrote, potentially raise Bivens claims, but Arar would have to redraft a complaint excluding the rendition claim and naming the defendants that were personally involved in the alleged unconstitutional treatment. That Arar never did.\textsuperscript{166}

Trager’s opinion in Arar was subjected to considerable criticism. He may have escaped some obloquy had he resolved

\textsuperscript{159} Id. at 258-59.
\textsuperscript{160} Id. at 263.
\textsuperscript{162} Arar, 414 F. Supp. 2d at 281.
\textsuperscript{163} Id. (quoting United States v. Verdugo-Urquidez, 494 U.S. 259, 273-74 (1990)).
\textsuperscript{164} Id. at 282.
\textsuperscript{165} Trager denied Arar’s claim of denial of access to counsel and he dismissed all claims against the individuals without prejudice to repleading Count 4. Id. at 286.
\textsuperscript{166} The United States had moved for summary judgment in the Arar case, invoking the state-secrets privilege. Having held for the government on the statutory and constitutional claims, Trager found the issue involving state secrets moot. Id. at 287.
the case on the basis of the state-secret doctrine, but he possibly did some good in resolving so many subissues favorably to Arar. The federal judiciary as a whole did not cover itself with glory in litigation involving 9/11. This author wonders about the claim that executive officials, often with very little experience in dealing with national security matters, are more up to the task of balancing individual rights against the claims of national security than experienced judges. Barbara Olshansky, Deputy Legal Director of the Center for Constitutional Rights, has put it this way: “There can be little doubt that every official of the United States [involved in Arar’s torture] knew that sending him to Syria was a clear violation of the U.S. Constitution, federal statutes, and international law.... This is a dark day indeed.⁴⁶⁷ A ruling for Arar would have flown in the face of most of the 9/11 cases and perhaps such a step would not have been appropriate for a district judge. At least Trager’s rulings for Arar on the subissues might have operated as a brake on the executive.

A panel of the Court of Appeals for the Second Circuit affirmed Trager’s decision over one dissent.⁴⁶⁸ Without either party requesting it, the Second Circuit vacated the panel opinion, and then, sitting en banc affirmed Trager by a vote of seven to four.⁴⁶⁹ Writing for the court, Chief Judge Dennis Jacobs affirmed Trager as to standing. He dispatched Arar’s claim under the TVPA more easily than Trager, however. Judge Jacobs held that to state a claim under the TVPA, there must be an allegation that U.S. officials possess power under foreign law and the offending actions must derive from an exercise of that power.⁴⁷⁰ On the Bivens claim, the court also affirmed. While not precluding judicial review and oversight, it held that if there was to be a civil remedy in damages suffered in the context of extraordinary rendition, Congress would need to create such a remedy.⁴⁷¹ The Supreme Court denied

---

⁴⁶⁸ Arar v. Ashcroft, 532 F.3d 157 (2d Cir. 2008).
⁴⁶⁹ Arar v. Ashcroft, 585 F.3d 559 (2d Cir. 2009) (en banc).
⁴⁷⁰ Id. at 568.
⁴⁷¹ Id. at 576. In his dissent, Judge Robert A. Sack found the majority’s recitation of the facts as “generally accurate, but anodyne.” He then stated that “[t]he
The treatment of the Arar case by all three tiers of the federal judiciary was of a piece with the approach of all three branches of the U.S. government to refuse to be held accountable for torture after 9/11.

D. United States v. Nelson

The most publicized decision of Judge Trager's career was connected to boiling racial tensions in his home borough, Brooklyn. On August 19, 1991, an automobile driven by a Hasidic Jew struck two black children in Brooklyn's Crown Heights; one of the children was killed. Rumors swiftly spread that the ambulance driver at the scene had treated the Hasid ahead of the children. A riot occurred. Eleven African-Americans—including a defendant, Lemrick Nelson Jr.—chased Yankel Rosenbaum, an orthodox Jew, and stabbed him to death. Nelson, then sixteen, was found with a bloody knife and was positively identified by the victim before he died. Nelson and Charles Price, who had harangued the crowd, were charged with second-degree murder and acquitted in state court. Bitter feelings between the Jewish and African-American communities festered.

Nelson moved to Georgia where he got into more trouble. He pleaded guilty, as an adult, to aggravated assault and carrying a concealed weapon. The U.S. Attorney for the
district court's opinion carefully and fully sets forth Arar's allegations." Id. at 582, 584 (Sack, J., dissenting).

Judge Trager also had before him a rather different case involving terrorism. This was a suit brought by a family who were dual citizens of the United States and Israel, who sued a Swiss financial institution with offices in the United States and Israel over the death of their husband/father who was killed in Israel in a bus blown up by terrorists. The plaintiffs claimed that the bank (UBS AG) had provided financial services for the alleged terrorist organization, Hamas. In September 2009, Judge Charles P. Sifton dealt with issues of standing, forum non conveniens, and motions to dismiss. Goldberg v. UBS AG, 660 F. Supp. 2d 410 (E.D.N.Y. 2009). After Sifton's death in November 2009, the case was assigned to Trager who denied motions for reconsideration of the forum non conveniens issue and to certify Sifton's order for interlocutory appeal. Goldberg v. UBS AG, 690 F. Supp. 2d 92 (E.D.N.Y. 2010).

174 Id. at 637-38.
175 Id. at 638.
176 Id.
177 Id.
178 Id. at 640.
Eastern District of New York charged Nelson, then nineteen, with juvenile delinquency over Rosenbaum’s death. Five days after the information was filed, the government moved to transfer Nelson to criminally prosecute him as an adult for violation of Rosenbaum’s civil rights. Trager denied the order but was reversed by the Second Circuit. The appellate court held that Trager had improperly evaluated the strength of the government’s evidence and erred by not considering Nelson’s age at the time of the transfer (minimizing the seriousness of the offense), by using a “glimmer-of-hope” test to determine the possibility of Nelson’s rehabilitation, and by not making any inquiry into juvenile programs available for someone of Nelson’s age.  

The Court of Appeals vacated and remanded to Trager for further findings and reconsideration.

Before the hearing on remand, Nelson had been charged with resisting arrest and criminal trespass. The latter had occurred when he refused to leave the federal courthouse. In his opinion on remand, Trager granted the motion to transfer because of the seriousness of the crime charged as well as the finding that “it [was] not ‘likely’ that Nelson would be rehabilitated.” This time the decision was upheld by the Court of Appeals. Trager wanted to avoid turning the Nelson trial into a proceeding as racially divisive as Rodney King’s had been in Los Angeles. He intended to empanel “a moral jury that render[ed] a verdict that ha[d] moral integrity.” However, an important Supreme Court decision, Batson v. Kentucky, stood in his way. Batson held that peremptory strikes on the basis of race violate the Equal Protection Clause. That case and its progeny were primarily aimed at lawyers, but presumably bound judges as well.

Presiding over jury selection in the Nelson case, Trager denied the defendant’s for-cause challenge of a Jewish juror (Juror 108) who had doubted his ability to be objective. Then, when an African-American juror was excused, the Judge did not replace the juror with the first alternate juror but rather

---

182 Id.
185 Id.
186 Nelson, 277 F.3d at 172.
sua sponte removed a second white juror from the panel and filled the two spaces with an African-American juror and Juror 108. The resulting jury, which included three African-Americans and two Jews, convicted both men. Trager sentenced Nelson to 235 months and Price to 262 months.

On appeal, much of Judge Guido Calabrisi’s lengthy opinion revolved around the application of the Civil Rights Act to the Nelson case. But in dealing with Trager’s race- and religion-based shuffling, the court held that Juror 108 had been improperly seated because he had revealed sufficient bias during the voir dire. Further, the consent given to his selection by the attorneys was invalid because it was obtained in exchange for the improper empanelling of a jury chosen partly on the basis of race. Even though the defendants had improperly consented to the scheme,

where the trier of fact in a criminal trial is a biased jury that resulted from a district court’s erroneous failure to grant a for-cause challenge to an actually biased juror whose bias was revealed at the voir dire, we question whether a defendant can subsequently waive his claim . . . to be tried before an impartial fact finder.

Even though “the motives behind the district court’s race- and religion-based jury selection procedures were undoubtedly meant to be tolerant and inclusive... that fact cannot justify the district court’s race-conscious actions.”

Dissenting in part, Judge Chester J. Straub, though troubled about the jury selection, would have affirmed, but noted the court’s willingness to vacate such efforts in the future. As for the Nelson case, Straub said, “When one considers the overall circumstances and conditions of this trial, we can have overwhelming confidence in the fairness and validity of its verdict.

Thus, Trager was reprimanded for an “activist” solution intended to avoid divisiveness and reach a “moral” result.

---

187 Id.
188 Id. at 173.
189 Id. at 203-04.
190 Id. at 203-04.
191 Id. at 208, 210.
192 Id. at 206.
193 Id. at 207.
III. INDIVIDUAL RIGHTS CASES

A. Section 1983 and More Modern Statutory Civil Rights Cases

While Trager was not a notable enthusiast for workplace-based claims under sections 1981\(^{194}\) and 1983,\(^{195}\) plaintiffs did not fare too badly in cases before him.\(^{196}\) Trager granted summary judgment against British Airways in a Title VII action. The plaintiff, Elizabeth de Chanval Pellier, claimed that after she was promoted to duty-maintenance manager (a middle-level management position) at John F. Kennedy Airport, other employees engaged in inappropriate sexual conduct, including the posting of pornographic material directed specifically at Pellier.\(^{197}\) Because the airline did not effectively respond to her complaints, Pellier ultimately had to accept a position with fewer material responsibilities in order to escape the intolerable conditions.

Trager found that Pellier's transfer was not an "adverse employment action because she [had] voluntarily requested and accepted it."\(^{198}\) However, he did find that Pellier had a triable claim of hostile work environment.\(^{199}\) British Airways argued that Pellier had engaged in conduct similar to that of which she was complaining: "over-exuberant hugging and kissing," sexually explicit conversations, and—during her work history at the airlines—"intimate relations" with two British Airways employees (one of whom she later became engaged to).\(^{200}\) Yet, the judge found the consensual relations "irrelevant."\(^{201}\) "Even if... Pellier was comfortable with certain

\(^{194}\) 42 U.S.C. § 1981 (2006) (protecting the rights of all persons within the jurisdiction of the United States to, among other things, make and enforce contracts and entitling them to the full and equal benefit of laws for the security of persons and properties as enjoyed by white citizens).

\(^{195}\) Id. § 1983 (providing the remedy for deprivation of federal constitutional and statutory rights when violated under color of state law).


\(^{198}\) Id. at *4.

\(^{199}\) Id. at *15.

\(^{200}\) Id. at *7.

\(^{201}\) Id.
sexual behavior in the workplace, a reasonable jury need not conclude that she was thereby comfortable...[with] being singled out by an all-male staff as the lone target of sexually explicit materials.” He later granted summary judgment to the defendant for the sex-discrimination claim, but denied it as to the retaliation and hostile-environment claims. Trager evidenced no lack of sympathy for the defendant in this situation.

Trager also held for the plaintiff in a case involving the New York City Transit Authority. For promotion to the position of subway station supervisor, the authority required an EKG test if the candidate had a problematic medical history and was over forty. The plaintiff challenged the policy under the Age and Discrimination in Employment Act. Trager refused to accept the defendant’s bona fide occupational-qualification defense. As he wrote, practically and with wit,

To put the question sharply, what exactly would happen if a Station Supervisor, Level I were to suffer an unexpected heart attack? While such an event would unquestionably be more than a small inconvenience to the individual himself...there is no indication in this record that a great danger would befall either the TA’s operations or the general public.

In another case, Trager clearly was greatly disturbed by the narrative of Lawrence Hardy’s section 1983 suit. Hardy, an ex-convict who had served a four-year sentence for robbery, was arrested for violating parole. He then sued under section 1983. The heart of the section 1983 action, discussed by Trager in thirty-four pages of the Federal Supplement, was the New York City Department of Corrections’ deliberate indifference to Hardy’s very serious ear infection.

As Trager told the story, Hardy’s ear infection had been causing him pain and dizziness prior to his arrest. The defendants’ gross negligence magnified the symptoms, and a golf-ball-sized swelling grew from Hardy’s neck. From there, the symptoms developed into extreme pain, blurred vision,

---

202 Id. at *8 (citation omitted).
203 Epter v. N.Y.C. Transit Auth., 127 F. Supp. 2d 384, 385 (E.D.N.Y. 2001). After winning the lawsuit and getting the transit authority to change its policy, Epter was promoted to station supervisor, but then asked to be returned to his old position. Judge Trager awarded back pay damages and attorneys’ fees to Epter but held that liquidated damages were not appropriate. Epter v. N.Y.C. Transit Auth., 216 F. Supp. 2d 131, 139 (E.D.N.Y. 2002).
204 Id., 127 F. Supp. 2d at 391-92.
206 Id. at 118.
discharge from the ear, and hearing loss.\footnote{207} Within eight weeks, Hardy had difficulty walking and standing as the swelling in his neck grew to the size of a tennis ball with pus emanating from his ear in a constant flow.\footnote{208} Only then was Hardy seen by a specialist and a mastoidectomy performed.\footnote{209} The operation was but a partial success as a staph infection developed.\footnote{210} A little more than two months after the operation, Hardy was released from custody and admitted immediately to Manhattan Eye and Ear Hospital.\footnote{211} Diagnosed with a life-threatening infection, Hardy survived but ended up in a nursing home where he eventually lost hearing in one ear.\footnote{212}

Trager dismissed Hardy's action against the New York State defendants on Eleventh Amendment grounds and dismissed claims of deliberate indifference against some of the officers.\footnote{213} However, Trager preserved claims against some of the defendants at the Downstate Correctional Center, as well as a nurse, a physician, and correction officers at the Willard Drug Treatment Center.\footnote{214}

B. Social Security Disability Appeals

By the time David Trager ascended to the bench, the “war” the federal judiciary had fought with the Social Security Administration was over. Under political pressure, bureaucrats and administrative law judges had in the 1980s denied large numbers of applicants’ benefits and had thrown many recipients of benefits off the payroll. Federal judges of all judicial persuasions had been remanding cases to the agency.\footnote{215} The battle had waned by the time Trager was appointed. Judge Trager generally upheld agency determinations without publication of opinions. There was, however, one case where he granted judgment on the pleadings and awarded ten years of disability retroactively.

\footnote{207} {Id. at 119.}
\footnote{208} {Id. at 120.}
\footnote{209} {Id. at 122.}
\footnote{210} {Id. at 123.}
\footnote{211} {Id. at 124.}
\footnote{212} {Id.}
\footnote{213} {Id. at 127-28.}
\footnote{214} {Id. at 145-46.}
\footnote{215} See Jeffrey B. Morris, Establishing Justice in Middle America 271-72 (2007); Morris, supra note 60, at 188-89.
Irene Rooney was denied benefits for her disabilities three times.\footnote{Rooney v. Shalala, 879 F. Supp. 252, 254 (E.D.N.Y. 1995).} Fifty-nine years old, she had been a reliable employee for thirty-three years but suffered from progressive asthmatic disorder, chronic obstructive pulmonary disease, bronchitis, and arrhythmia.\footnote{Id.} She was also in considerable pain due to degenerative disc disease and chronic spinal strain. She was also blind in one eye.\footnote{Id.}

In 1987, Rooney filed an application for disability benefits.\footnote{Id.} Denied benefits initially and on reconsideration, she received a notice from the Department of Health and Human Services informing her that she had sixty days to request a hearing if she believed that the determination was incorrect.\footnote{Id. at 255.} The notice stated that if she did not request a hearing she could still file another application at any time.\footnote{Id.} Nothing in the notice informed her that not requesting a hearing would mean permanent loss of her opportunity to claim substantial benefits up to the time of the new claim. There was nothing giving her notice of the consequences of electing not to have a hearing.

From 1987 until 1991, Rooney, unaware that reapplication was not the same as requesting a hearing, filed two applications—both denied.\footnote{Id. at 256.} However, by 1991, the agency’s notice of reconsideration had changed and warned about the adverse effect of failing to appeal.\footnote{Id. at 257.} Proceeding with counsel, she finally requested a hearing during which she sought to open the previous denials.\footnote{Id.} When the administrative law judge refused to pursue the issue, Rooney appealed to the district court.\footnote{Id.}

Even though the decision not to reopen an adjudicated claim is generally not reviewable by the courts, Trager found a violation of procedural due process. Writing that the record was “somewhat muddied,”\footnote{Id. at 257.} Trager stated that the notice Rooney had been given was “almost deliberately crafted to divert the lay person from a timely appeal.”\footnote{Id. at 256.} The record in Rooney’s case,
Trager wrote, was “a testament to inattention and disregard.” Rather than remand the case, Trager sympathized with Rooney, “a fifty-year old woman with a litany of impairments who had worked hard all her life.” He went on to describe the record as devoid of “malingering” and with a “remarkable number of procedural and evidentiary errors.” In turn, Judge Trager directly awarded ten years of disability benefits to her without a remand.

C. Immigration, Deportation, and Extradition Matters

Judge Trager held for the government in most of the immigration, deportation, and extradition matters before him. One case, though, is worth brief scrutiny partly because of its lengthy, thorough opinion holding that an indictment for illegally entering the country was unlawful because the underlying deportation had been unlawful. The case is worth further scrutiny because of what this observer reads as a particular sympathy for the defendant.

Richard Garcia-Jurado, a Colombian, had come to the United States as a legal permanent resident to join his mother, who was then a legal permanent resident and later became a U.S. citizen. Garcia-Jurado’s family in the U.S. included his brother and sister, stepfather, and half-sister (who were U.S. citizens), and his brother and older sister who were legal permanent residents. Further, he had a long-term relationship with a legal permanent resident with whom he had a daughter. In 1993, Garcia-Jurado pleaded guilty to possession

---

228 Id. at 261.
229 Id. at 258, 261-62.
230 Id. at 262.
231 See, e.g., Guang v. Immigration & Naturalization Serv., No. CV-02-5916, 2005 WL 465436 (E.D.N.Y. Feb. 28, 2005) (rejecting habeas corpus petition attempting to block removal to China where rejection of asylum claim was correct); Echendu v. United States, No. CV-02-1255, 3003 WL 21653370, at *1, *8 (E.D.N.Y. July 14, 2003) (denying writ of coram nobis to prospective deportee who alleged ineffectiveness of counsel where he was convicted of knowingly failing to appear in court in a case where he was acquitted); Lo Duca v. United States, No. CV-95-713, 1995 U.S. Dist. LEXIS 21155 (E.D.N.Y. July 12, 1995) (upholding legality of extradition order by Magistrate Judge; extradition request and supporting documentation timely; questions of sufficiency of extradition documents both beyond the scope of review of petition for habeas corpus, but also sufficient on the merits).
233 Id. at 500.
234 Id.
235 Id.
of a controlled substance.\textsuperscript{236} For that, he was sentenced to four to twelve years in prison—which he used to complete his GED and earn two vocational training certificates.\textsuperscript{237} At his deportation hearing, Garcia-Jurado indicated that he would seek discretionary relief under section 212(c) of the Immigration and Naturalization Law.\textsuperscript{238} However, the immigration judge inaccurately told him that he was not entitled to seek a discretionary waiver of deportation.\textsuperscript{239} Somehow—and Judge Trager ignored this part of the story—Garcia-Jurado returned to the United States four years later and was arrested for criminal possession of marijuana.\textsuperscript{240} Charged with illegally reentering the United States after deportation, an aggravated felony, the defendant came before Judge Trager.\textsuperscript{241}

Trager’s reading of the facts led him to the position that Garcia-Jurado had been deported before his administrative proceedings had ended and “before the deadline for judicial review had passed.”\textsuperscript{242} After considerable effort, Trager found that the immigration judge’s error in failing to provide Garcia-Jurado with a section 212(c) hearing “was a procedural error that rendered the [previous] proceedings fundamentally unfair.”\textsuperscript{243} The loss of a 212(c) hearing supported a collateral attack.\textsuperscript{244} That, in turn, led to Trager’s holding, “Because Garcia-Jurado was deprived of [his right to] judicial review . . . his deportation was fundamentally unfair.”\textsuperscript{245} Trager concluded the opinion with what could be viewed as a letter of recommendation:

Garcia-Jurado claims that he would have been a good candidate for a § 212(c) waiver of deportation. Indeed, he would. Garcia-Jurado had lived here since he was a teenager. He has strong family ties in this country: a mother, stepfather and two half-sisters who are United States citizens. He also has two other siblings who are legal permanent residents, and, importantly, a daughter who is a citizen. He attended high school here, was employed for a year prior to his arrest. He also received a GED and vocational training in prison,

\begin{footnotes}
\item[236] Id.
\item[237] Id.
\item[238] Id. at 501.
\item[239] Id.
\item[240] Id.
\item[241] Id.
\item[242] Id. at 506-07.
\item[243] Id. at 512.
\item[244] Id. (citation omitted).
\item[245] Id. at 514.
\end{footnotes}
and was employed as a machine operator after his release from prison up to the time he was deported.\textsuperscript{246}

D. Dean v. United States

In the case of Dean v. United States, Judge Trager found a clear injustice and ordered the government to correct an arrest record—an unusual remedy.\textsuperscript{247} The case involved the arrest of a school-bus driver for public lewdness in a national park.\textsuperscript{248} Dean had been arrested, read his Miranda rights, photographed, and fingerprinted.\textsuperscript{249} Dean alleged that, at the time of his arrest, one of the officers at the police station had advised him not to involve an attorney. Rather, if he just paid the fine, the officer stated, "the incident would fall off [his] record in a few years."\textsuperscript{250} Dean had promptly mailed in the form he was given and paid the eighty-dollar fine.\textsuperscript{251}

Twelve years later Dean's arrest showed up in a background check for renewal of his commercial driver's license.\textsuperscript{252} The payment of the fine was treated as a guilty plea. As a result, under state law, Dean was unable to continue his employment as a school-bus driver.\textsuperscript{253}

Dean attempted first to have his fingerprints record expunged on the ground that his criminal record had made it difficult for him to secure employment. Trager held that "such relief [could] not be granted under the circumstances."\textsuperscript{254} Dean tried using habeas corpus to achieve the same end. Trager treated the petition as one for coram nobis, a challenge to invalidate convictions after the sentence has been served, and appeared ready to grant that writ.\textsuperscript{255} Instead, though, he directed the government to, within thirty days, produce the form upon which Dean allegedly pleaded guilty.\textsuperscript{256}

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{246} Id. at 515.
\item \textsuperscript{247} 418 F. Supp. 2d 149, 157-58 (E.D.N.Y. 2006).
\item \textsuperscript{248} Id. at 151.
\item \textsuperscript{249} Dean v. United States, No. 05-CV-1496, 2005 U.S. Dist. LEXIS 40903, at *1 (E.D.N.Y. Aug. 31, 2005).
\item \textsuperscript{250} Id. (internal quotation marks omitted).
\item \textsuperscript{251} Id. at *2.
\item \textsuperscript{252} Id.
\item \textsuperscript{253} Id.
\item \textsuperscript{254} Dean v. United States, No. 04-MC-299, 2004 U.S. Dist. LEXIS 30062, at *6 (E.D.N.Y. Nov. 8, 2004).
\item \textsuperscript{255} Dean, 2005 U.S. Dist. LEXIS 40903, at *4.
\item \textsuperscript{256} Id.
\end{itemize}
\end{footnotesize}
The government was unable to locate the document but, attempting to comply with the order, submitted the “violation notice” form used by the U.S. Park Police at the time of Dean’s arrest. “The word ‘guilty’ did not appear on the face of [that] notice.” Trager was troubled because Dean had not in any way been informed of the legal consequences he might suffer if he paid the fine. Trager held that Dean should have been warned of the consequences of his guilty plea—that payment of the fine was acceptance of a federal conviction. Without that warning, Dean could not have knowingly waived his rights.

Trager rejected the government’s argument that “allowing a litigant to challenge a petty offense conviction” in this manner “would potentially call into question every petty offense and misdemeanor.” “It appears,” Trager wrote, “that the government—and perhaps the courts are complicit—wants to have the ‘benefit’ of a criminal conviction for a large number of petty offenses without the burden of providing appropriate procedural protection.” Trager made it clear that, if the government wanted to treat this sort of collateral forfeiture as a criminal penalty, then to satisfy constitutional concerns, “the individual must be given clear notice that payment of the fine constitutes a guilty plea resulting in a conviction of a petty federal offense and be informed of the right to retain counsel.”

With regard to the coram nobis petition, Trager held for the plaintiff even though there was “an extremely stringent standard applied to orders to change arrest and conviction records.” He granted coram nobis relief, directing the government to complete Dean’s record with “a clarification reflecting that he was not convicted of a crime.”

---

258 Id. at 155-56.
259 Id.
260 Id. at 156.
261 Id.
262 Id. at 157.
263 Id. at 152.
264 Id. at 157. In a later opinion, Trager granted the government the authority to depose Dean on whether he had been aware of the conviction so that it could show that the petition was not timely. Otherwise he rejected the government’s petition for reconsideration. Dean v. United States, 436 F. Supp. 2d 485 (E.D.N.Y. 2006). But see Grunberg v. Bd. of Educ., No. CV-00-4124, 2006 U.S. Dist. LEXIS 22424 (E.D.N.Y. Mar. 30, 2006).
E. Funding for the Profoundly Disabled and Medically Fragile

David Trager was no “wooly headed dreamer.” Throughout his career, Trager sought solutions that were practical and just. That was, perhaps, why he pursued the course he did in the Lemrick Nelson case. But Judge Trager’s passion was closest to the surface in a litigation he handled over the residential placement of severely disabled young adults.

To some, this litigation was just a lawsuit that attempted to pry money from the overburdened fiscs of the State and its cities and counties. New York incurred a large financial burden from treating and housing the severely disabled; residential placements were absolutely necessary for them. When no satisfactory placements existed within the state, out-of-state placements had to be made, and New York split expenses with its cities and counties. Federal funds had eased the financial burden, but individuals were not eligible for funding after the age of twenty one.

In 1992, arrangements were made wherein New York agreed to share 50 percent of the expense of out-of-state placement until suitable in-state institutions were located. In 1994, New York enacted a statute that increased state payments to localities and scheduled the phase-out of out-of-state placements and the out-of-state recompensation provision.

In 1994, New York City pulled out of the arrangements because of budgetary pressure. During the first half of 1995, the State attempted to place recipients of Transitional Care Funding (TCF) in state facilities with some success. In February of that year, lawsuits were brought by parents and guardians of TCF recipients to prohibit New York City and New York State from terminating transitional care. The suits were unsuccessful. Then, over the Fourth of July weekend, the State attempted to transfer some of the deeply disturbed

---

266 Id.
267 Id.
individuals without the permission, or even notification, of their parents/guardians. The result was somewhat of a fiasco.

A federal action was brought by eighteen of the disabled young adults, and eight others sought to intervene. They were asking for a preliminary injunction requiring the state to take all necessary steps to maintain the placements until an orderly transition to permanent, state-approved placements was accomplished.

As is often the case with judicial opinions, the manner in which the facts are stated presage the ultimate conclusion of law and give a clue to the impact the case has had upon the judge. This is the first paragraph of Judge Trager’s opinion in Brooks v. Pataki, the first lawsuit involving the funding: “This case concerns the care and treatment of about fifty profoundly disabled and medically fragile individuals whose rights under the federal Constitution have been gravely imperiled as the result of an unfortunate funding dispute between the City and State of New York.”

In the following paragraph, Trager vividly illustrates the “victims” in the dispute:

[O]ne is a twenty-six old woman who has epilepsy and an IQ of about 70-72 as well as other disabilities. Another is a twenty-five year old, profoundly retarded . . . woman. Seven of the eight proposed intervenors are profoundly autistic and pose potential danger, certainly, to themselves and, possibly to others.

Six months later in a different lawsuit over the same issues with similar plaintiffs, Trager wrote:

It is important to understand just how profoundly disabled plaintiffs are. TCF recipients are not likely candidates for de-institutionalization. For instance, plaintiff Lora Hoops, aged twenty-five, diagnosed as functioning in the profound range of mental retardation, with cerebral palsy and a seizure disorder has been at The Woods School, in Langhorne, Pennsylvania, since she was placed there by a Suffolk County local school district nineteen years ago, at the age of six.

In the suit involving Westchester County funding, Trager quoted from a description of plaintiff Jason Goodhue, age

269 Brooks, 908 F. Supp. at 1160 & n.2.
270 Id. at 1143.
271 Id.
272 Id. at 1143-44 (citations omitted).
twenty three, who had “a diagnosis of severe mental retardation, cerebral palsy, spastic diplegia, functional scoliosis, . . . encephalopathy and a seizure disorder. He uses a wheel chair which [he] is able to push independently, though his control is poor. He is nonverbal, understands simple commands and usually responds yes/no verbally to questions.”

In considering whether to grant the preliminary injunction in the New York City case, Trager focused on claims of due process—claims which had not been finally adjudicated in the State's case. Trager found the constitutional standard in a Second Circuit decision, Society for Good Will to Retarded Children, Inc. v. Cuomo, and deemed the State's procedure “manifestly unprofessional.” According to Trager, the episode on the Fourth of July was “an unconstitutional violation of the rights of these individuals that threatened them with irreparable harm.”

Trager offered an analogy to a State's or Congress's decision to end support for a dialysis program:

Such a program is not an entitlement; there is no custody of the patient for state action purposes; and the program's elimination is within the discretion of a State or Congress. Still, no one would seriously argue that the Due Process Clause does not impose an obligation upon a State or Congress to provide reasonable notice to dialysis recipients before terminating funding for those who have relied on that life-sustaining program, so as to allow them the opportunity to obtain alternate access to treatment.

Trager denied the State's motion to stay the judgment. In that opinion, he made clear that the funding was not an entitlement. Rather, once the State undertook to provide residential care for individuals, it was obligated to provide “necessary safe conditions and freedom from undue restraint determined by the exercise of professional judgment.” Having failed to make appropriate in-state placements, the State could not abruptly leave these severely disabled individuals “stranded in placements when the City [terminated its] funding.”

A few months later, Suffolk County followed New York City in withdrawing from the State program. Now Trager had

---

275 737 F.2d 1239, 1250 (2d Cir. 1984).
276 Brooks, 908 F. Supp. at 1151.
277 Id. at 1152.
278 Id.
279 Id. at 1150-51 (citation omitted).
280 Id. at 1154.
a companion case. In Suffolk Parents of Handicapped Adults v. Pataki, Trager granted a preliminary injunction ordering Suffolk County to resume funding for six months “so as to provide...the opportunity to arrange alternative care in an orderly manner.” The State indicated it would “reimburse the county for sixty percent of the cost” and was ordered to “assume the burden of funding any remaining TCF placements at the end of the six-month period.” Just about one month later he denied the County’s request for a stay pending appeal.

Westchester County then withdrew from the state program, providing Trager with a third case. The Westchester case posed the same issues as the Suffolk case. However, the Westchester case came down after the U.S. Court of Appeals had remanded the New York City case. As will be seen shortly, Trager was unable to rest on the Due Process Clause any longer. Instead, he based his decision in the Westchester case on the Equal Protection Clause, holding that the County had violated equal protection by granting protections to persons institutionalized in-state that the TCF recipients (out-of-state) did not have. Trager held that the TCF statute lacked a rational basis. Trager also held that the State failed to offer procedural due process. As a result, three weeks after the Court of Appeals had vacated Trager’s decision in the New York City case, Trager granted the Westchester plaintiffs’ motion for a preliminary injunction against the County defendants. The State was ordered to, among other things, continue to pay for the out-of-state facilities where plaintiffs were still residing. However, Trager stayed the injunction based upon the stay the Court of Appeals had ordered in the Suffolk case.

However, as has already been intimated, three weeks before the first Westchester decision the Court of Appeals, by a two to one vote, rained on Trager’s parade. The panel majority vacated the injunction in the New York City case and remanded

282 Id. at 986.
283 Id.
285 With one exception—a venue issue. The court held that venue was properly exercised in the Eastern District. Id. at 1004-05.
286 Id. at 1006.
287 Id. at 1010.
288 Id. at 1010-11.
289 Id. at 1013-14.
it to Trager.\textsuperscript{290} The appellate court held that the plaintiffs’
claims were barred by res judicata. The court also held that the
defendants had no duty under the Due Process Clause to
provide professional care to plaintiffs and no duty to resume
payments for the plaintiffs’ out-of-state placements.\textsuperscript{291} The
Court of Appeals also held that the “July Fourth weekend
episode amounted to a breach of the State defendants’ duty to
ensure that the involuntary transfers were constitutionally
appropriate, but that this injunction [was] not the appropriate
remedy for this breach.”\textsuperscript{292} Barrington Parker dissented from
the Court of Appeals decision. He would not have “disturbed
Trager’s conclusions that there [were] sufficiently serious
questions regarding the risk of further abuse of the plaintiffs’
substantive due process rights, and that the balance of
hardships tip[ped] decidedly in plaintiffs’ favor.”\textsuperscript{293}

The parties in Brooks v. Pataki reached an agreement
whereby the State would continue to provide the requested funds
and the case was administratively closed on May 11, 2000.\textsuperscript{294}

In the handling of the litigation involving the mentally
disabled, David Trager sounded and acted more like his
colleague, Jack B. Weinstein, than one might expect. In a letter
written while the litigation was still alive, Trager commented,
“As a non-believer in entitlements, I found myself developing a
purely procedural due process theory to justify judicial
intervention in a fact pattern that cried out for relief.”\textsuperscript{295}

One of the opinions in the Suffolk County case indicated
how far Judge Trager, no judicial activist by philosophy, traveled
in the mental health litigation: “The role of the judiciary to review
executive and legislative actions and to protect the rights of
persons unable to protect themselves from unconstitutional
governmental intrusion has long been recognized.”\textsuperscript{296}

\textsuperscript{290} Brooks v. Pataki, 84 F.3d 1454, 1468 (2d Cir. 1996).
\textsuperscript{291} Id. at 1465-66. The court of appeals held that the Supreme Court decision,
DeShaney v. Winnebago County Department of Social Services, 489 U.S. 189 (1989),
had altered the reach of the Second Circuit decision, Society for Goodwill to Retarded
Children, Inc. v. Cuomo, 737 F.2d 1239 (2d Cir. 1984).
\textsuperscript{292} Brooks, 84 F.3d at 1465.
\textsuperscript{293} Id. at 1470. Five months later, the U.S. Court of Appeals for the Second
Circuit vacated the injunction in the Suffolk case. Suffolk Parents of Handicapped
Adults v. Wingate, 101 F.3d 818 (2d Cir. 1996).
\textsuperscript{294} See Sanchez v. Pataki, No. CV-98-6282, 2008 U.S. Dist. LEXIS 21218, at
*2-3 (E.D.N.Y. 2008), aff’d, Sanchez v. Patterson, 328 F. App’x 689 (2d Cir. 2009).
\textsuperscript{295} Letter from David G. Trager to Jeffrey Morris (Jan. 9, 1996) (on file with author).
\textsuperscript{296} Suffolk Parents of Handicapped Adults v. Pataki, 924 F. Supp. 431, 435
(E.D.N.Y. 1996).
CONCLUSION

It is premature to assess a judicial career of seventeen years that only ended this year. Much of what a district judge does is not found in the judge's opinions. One would, for example, have loved to have been a fly on the wall observing this larger-than-life personality conduct settlement negotiations. Unfortunately, there is limited coverage of Trager's work in newspapers. However, the potential availability of the oral history Trager created during the last year of his life will enrich our knowledge of his work and—hopefully over time—interviews of his clerks, colleagues, and friends will provide a more definitive account of the work of not just a very able judge, but a man whose career as a whole stands as a model of how a single lawyer can benefit society.

But some observations of Trager, as judge, may be permitted at this time. Although Trager was a strong and unforgettable personality, as a judge, he clearly did not seek the spotlight. His opinions could be analytically complex, but his prose is almost always straightforward. A man, who in person could be extraordinarily amusing, rejected wit in judicial opinions.

Trager's career certainly demonstrates the breadth of the work of contemporary federal judges. Indeed, this author lacked the time and space to discuss Trager's handling of threshold matters (standing, ripeness, etc.). Nor did I discuss interesting Trager opinions in trademarks, commercial law, torts, and attorney dealings. The particular federal specialties that offered Trager the richest opportunities were antitrust and copyright. As one might have expected, Trager demonstrated marked ability in criminal cases both as a district and a court of appeals judge. Indeed, Trager penned for three U.S. Courts of Appeals; he would have distinguished himself as an appellate judge, had that job fallen to him. On the other hand, the district bench offered him constant interactions with others, something Trager loved.

A moderate and careful judge who abjured grandiloquence, Trager's emotions could be deeply engaged by his cases. That clearly was true in the mental health cases, the Dean case, and, probably, in other cases of injustice, such as befell Irene Rooney and Lawrence Hardy. In the case involving Lemrick Nelson, Trager was engaged in attempting to achieve in a different way a result which would ease rather than raise the simmering tensions of a divided city. David Trager was no “bleeding heart,” but he cared deeply for the community to which he had devoted his public life. And for them, he was a judge who was willing to take risks.
NOTES

Resolving the Conflict Between the Stolen Valor Act of 2005 and the First Amendment

INTRODUCTION

In May 2010, the New York Times unveiled a series of misstatements regarding the military career of Connecticut’s Democratic Senate candidate, Richard Blumenthal. As early as 2000, Mr. Blumenthal was believed to be a Vietnam War veteran. Not only did he expressly state, “I served in Vietnam,” but he also described the anguish he suffered as a result of the criticism and cynicism he and his “fellow” veterans endured when they came home. Although he made these misstatements for years, Mr. Blumenthal never served overseas. From 1965 through 1970, Mr. Blumenthal reportedly received five draft deferments; three were educational deferments and two were occupational. Once Mr. Blumenthal exhausted his potential deferments and drew a very low number in the draft lottery, he secured a position with the Marine Corps Reserve and avoided the battlefield. Although Mr. Blumenthal took “full responsibility” for these false claims once they were uncovered,

2 Hernandez, supra note 1, at A1.
3 Id.
4 Id.; Halbfinger & Barron, supra note 1, at A3.
5 Occupational deferments were very rare, especially after President Lyndon B. Johnson’s administration drastically reduced graduate school deferments in 1968. Hernandez, supra note 1, at A1.
6 Id.
he emphasized that he simply “misspoke” and did not intentionally lie to the American public.  

Shortly thereafter, the Washington Post began investigating the military history of Illinois Republican Senate candidate Mark Kirk. In his official biography and during a speech at a House committee hearing in March 2002, Mr. Kirk claimed that he was once the Navy’s Intelligence Officer of the Year. Aware that the Post investigation would disclose the inaccuracy of that statement and mindful of the Blumenthal debacle, Mr. Kirk acted. During the week of the media frenzy surrounding Blumenthal, Mr. Kirk blogged about the misrepresentation he had made in his biography. In actuality, the National Military Intelligence Association, a professional organization, gave the award to not just him, but his entire service unit, which had been based in Aviano, Italy, in 2000. Mr. Kirk had also embellished his military career by claiming that he “served in the Gulf War,” “commanded the Pentagon war room,” and flew “intelligence missions over Iraq” while “under fire.” Though Mr. Kirk was in fact a member of the Navy Reserve beginning in 1989, these stories are simply not a part of his service record. Like Mr. Blumenthal, Mr. Kirk came clean in the beginning of June and apologized for his misstatements. Despite lying about their military valor, both Mr. Blumenthal and Mr. Kirk were elected to the United States Senate in November 2010.

7 Halbfinger & Barron, supra note 1, at A3.
9 Id.
11 Id.
12 Id.
13 Id.
14 Id.; Smith, supra note 8.
15 Lighty, supra note 10. In a meeting with The Chicago Tribune, Kirk admitted that portions of his résumé regarding his military experience were embellishments and not as precise as they should be. He apologized for the “misstatements” and repeatedly indicated that they were “mistakes” due to an effort to translate technical terms to the voters and the carelessness of his campaign staff. Id.
In addition to public officials, private citizens are lying about their military valor as well. In many instances, individuals have gone further than Blumenthal or Kirk, claiming to have received some of the most noteworthy and honorable military awards, such as the Congressional Medal of Honor and the Purple Heart.

In light of a substantial increase in this behavior, Congress enacted the Stolen Valor Act of 2005. The Act criminalizes the unauthorized wearing, selling, manufacturing, and distributing of military awards and decorations, as well as false representations regarding receipt of the awards made in written or spoken word.

Despite this congressional effort, courts have not consistently upheld the Act under the First Amendment. Section 704(b) of the Act is particularly problematic; it criminalizes false representations made “verbally or in writing.” Both the U.S. Court of Appeals for the Ninth Circuit and the U.S. District Court for the District of Colorado have held that section 704(b) is a facially invalid, content-based

---

17 For example, an individual, Andrew Alexander Diabo, claimed to be “a wounded Marine helicopter pilot” in Afghanistan and Iraq and a West Point Cadet. He supported his untruths with a West Point uniform hanging in his closet and military medals such as a Purple Heart and Silver Star framed and hanging on the walls of his home. He used these lies about his military history to defraud numerous people, accumulating over a half million dollars in debt to these individuals. After the Marine Corps Inspector General’s Office in Washington warned Diabo to stop, Diabo disappeared before any federal officials could properly charge him with any crimes. Larry King, How a Local ‘War Hero’ Went AWOL, PHILA. INQUIRER, Apr. 25, 2010, at A01; see also United States v. Strandlof, 746 F. Supp. 2d 1183, 1185 (D. Colo. 2010) (order granting defendant’s motion to dismiss information) (The defendant, Rick Glen Strandlof, was charged for falsely claiming to have received a Purple Heart on four separate occasions and a Silver Star on another occasion.); United States v. McGuinn, No. 07 Cr. 471(KNF), 2007 WL 3050502, at *1-5 (S.D.N.Y. Oct. 18, 2007) (memorandum and order) (The defendant, Louis Lowell McGuinn, after being discharged from the Army as a private, claimed that he was a lieutenant colonel and actually wore military medals such as the Silver Cross, Purple Heart, and Silver Star without ever having received the medals for his service. The state charged him under the Stolen Valor Act.); Christian Davenport, One Man’s Database Helps Uncover Cases of Falsified Valor, WASH. POST, May 10, 2010, available at http://www.washingtonpost.com/wp-dyn/content/article/2010/05/09/AR2010050903363.html (“The FBI investigated 200 stolen valor cases...and typically receives about 50 tips a month, triple the number that came in before the September 2001 terrorist attacks.”).  
18 See sources cited supra note 17.  
19 Davenport, supra note 17.  
21 Id.  
22 United States v. Alvarez, 617 F.3d 1198, 1217 (9th Cir. 2010) (holding the Stolen Valor Act of 2005 facially invalid under the First Amendment), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210); see also Strandlof, 746 F. Supp. 2d at 1192 (granting defendant’s motion to dismiss information because the Stolen Valor Act is unconstitutional under the First Amendment).
restriction under the First Amendment. In United States v. Alvarez, the Ninth Circuit concentrated on the question of whether “false statements of fact,” targeted by section 704(b), are protected by the First Amendment. The majority held that false statements are protected and, therefore, cannot be the predicate of criminal sanction unless the Act passes strict scrutiny. As a result of this conclusion, the court struck down section 704(b) because less restrictive alternative means were available to Congress. Similarly, in United States v. Strandlof, the District Court of Colorado found that the Act failed strict scrutiny. According to the court, the government failed to provide a compelling government interest in support of the Act.

By contrast, the U.S. District Court for the Western District of Virginia found section 704(b) constitutional in United States v. Robbins. Directly conflicting with the Ninth Circuit, the court held that false statements of fact are not protected by the First Amendment. As a result of this finding, the Western District of Virginia did not apply strict scrutiny, but rather upheld the Act's constitutionality under the overbreadth doctrine.

Overall, the constitutionality of section 704(b) has been inconsistently rejected among the lower courts. In response to the uncertainty presented by these courts' decisions, on October, 17, 2011, the Supreme Court granted certiorari in

---

23 See, e.g., Alvarez, 617 F.3d at 1217; Strandlof, 746 F. Supp. 2d at 1192. Under First Amendment jurisprudence, if the statute at issue is a content-based restriction on protected speech it is subject to the strict scrutiny test, see e.g., Citizens United v. Fed. Election Comm'n, 130 S. Ct. 876, 898 (2010); Reno v. ACLU, 521 U.S. 844, 874 (1997); Gooding v. Wilson, 405 U.S. 518, 522 (1972), which requires that the government show that the statute is “narrowly tailored to serve a compelling government interest.” Boos v. Barry, 485 U.S. 312, 321 (1988). By contrast, if the statute is a content-based restriction on a type of speech within a traditionally unprotected category, then it is not subject to strict scrutiny, but rather to less stringent analyses, such as the overbreadth and void-for-vagueness doctrines. See, e.g., Vill. of Hoffman Estates v. Flipside, Hoffman Estates, Inc., 455 U.S. 489, 494-504 (1982); Broadrick v. Oklahoma, 413 U.S. 601, 615 (1973); Grayned v. City of Rockford, 408 U.S. 104, 108-22 (1972). For a more detailed analysis of the overbreadth and void-for-vagueness doctrines, see infra Part II.C.

24 Alvarez, 617 F.3d at 1201-17.
25 Id. at 1217.
26 Id. at 1216-17.
27 Strandlof, 746 F. Supp. 2d at 1192.
28 Id. at 1189.
30 Id. at 817.
31 Id. at 818-19. See Part II.C.1 for a discussion of the overbreadth doctrine.
32 Compare United States v. Alvarez, 617 F.3d 1198, 1217 (9th Cir. 2010), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210), and Strandlof, 746 F. Supp. 2d at 1192, with Robbins, 759 F. Supp. 2d at 821.
United States v. Alvarez to determine whether the statute is facially invalid under the First Amendment.33

Before certiorari was granted, Congress sought to uphold the initiative of the Stolen Valor Act of 2005 and proposed a new statute entitled the Stolen Valor Act of 2011.34 The bill seeks to eliminate subsection (b) of 18 U.S.C. § 704 and remedy the section’s flaws.35 The bill also expands the reach of section 704(b) by criminalizing misrepresentations about one’s “military service,” rather than just misrepresentations about the receipt of a military award or decoration.36

This note examines section 704(b) of the Stolen Valor Act of 2005 and the proposed Stolen Valor Act of 2011 under the First Amendment. It argues that both Acts are unconstitutional. Part I recounts the history of protecting military valor in America leading up to the enactment of the Stolen Valor Act of 2005 and the introduction of the 2011 bill. Part II argues for section 704(b)’s unconstitutionality under First Amendment jurisprudence, and Part III discusses the unconstitutionality of the Stolen Valor Act of 2011. Additionally, this note suggests further change to the Stolen Valor Act of 2005 and calls upon state and local legislatures, in Part IV, to supplement these federal remedies. Specifically, Part IV suggests that Congress should restructure the Stolen Valor Act of 2005 to resemble a fraud statute. And furthermore, state and local legislatures, to hold public officials accountable, should impose an eligibility requirement prohibiting false claims of military valor and service by individuals seeking an elected position.

I. PROTECTING MILITARY VALOR AND THE STOLEN VALOR ACT OF 2005

Valor is defined as the “strength of mind or spirit that enables a person to encounter danger with firmness.”37 Derived from Middle English, Anglo-French, and Medieval Latin, valor has historically become synonymous with strength, worthiness, and

---

34 See Stolen Valor Act of 2011, H.R. 1775, 112th Cong. § 1 (2011); see also Alvarez, 617 F.3d at 1217 (9th Cir. 2010); Strandlof, 746 F. Supp. 2d at 1192; Robbins, 759 F. Supp. 2d at 821.
35 See H.R. 1775 § 2.
bravery. Moreover, valor has traditionally been used to characterize the distinguished bravery of members of the military.

In an effort to encourage, honor, and recognize military valor, numerous military awards and decorations have been created and awarded. Honoring members of the military is a tradition in the United States that dates back to the 1780s, when George Washington was President. As a former general, President Washington identified with a “desir[e] to cherish [the] virtuous ambition in his soldiers” exemplified through “instances of unusual gallantry...extraordinary fidelity, and essential service.” In 1782, President Washington created “honorary badges of distinction” to carry out his objective to “meet [valor] with [] due reward.” Included in these honorary badges was the predecessor to one of today’s most admirable military awards—the Purple Heart. While introducing these badges and describing their purposes and importance, Washington also strongly

---

38 Id.
39 See, e.g., id. (As an example of the use of the word valor, the site uses, “The soldiers received the nation’s highest award for valor.”).
40 See, e.g., Medal of Honor, 10 U.S.C. § 3741 (2006) (awarded to a member of the Army who “distinguished himself conspicuously by gallantry and intrepidity at the risk of his life above and beyond the call of duty”); Distinguished-Service Cross, id. § 3742 (awarded to a member of the Army who “distinguishes himself by extraordinary heroism not justifying the award of a medal of honor”); Distinguished-Service Medal, id. § 3743 (awarded to a member of the Army who “distinguishes himself by exceptionally meritorious service to the United States in a duty of great responsibility”); Silver Star, id. § 3746 (awarded to a member of the Army who “is cited for gallantry in action that does not warrant a medal of honor or distinguished-service cross”); Distinguished Flying Cross, id. § 3749 (awarded to a member of the Army who “distinguishes himself by heroism or extraordinary achievement while participating in an aerial flight”); Soldier’s Medal, id. § 3750 (awarded to a member of the Army who “distinguishes himself by heroism not involving actual conflict with an enemy”); Civil War Battle Streamers, id. § 3753 (authorizes the wearing of these streamers to units and regiments in the Army entitled by the Secretary of the Army); Korea Defense Service Medal, id. § 3756 (awarded to qualifying members of the Army who served in the Republic of Korea or adjacent waters during the conflict in Korea); Purple Heart, id. § 1129 (awarded to a member of the armed forces “who is killed or wounded in action as the result of an act of an enemy of the United States”); see also Purple Heart, id. § 1131.
42 Id. at 30.
43 Id.
asserted that individuals who falsely represented themselves as recipients of the awards should "be severely punished."\textsuperscript{45}

As the passage of the Stolen Valor Act of 2005 demonstrates, Washington was not the only government official who sought to protect the reputation of the nation's military awards.\textsuperscript{46} Originally, the Stolen Valor Act only criminalized the unauthorized wearing, manufacturing, or selling of military awards and decorations.\textsuperscript{47} In late 2005, however, Congress introduced an amendment to 18 U.S.C. § 704 that enhanced the section's protection.\textsuperscript{48} The amendment allows the prosecution of those who falsely claim receipt of military awards.\textsuperscript{49} The proposed amendment's sponsor, former U.S. Congressman John Salazar, was inspired by the college thesis of a constituent, Pamela Sterner. In the thesis, Mrs. Sterner advocated for a criminal statute to police false claims of military valor.\textsuperscript{50} Upon

\begin{itemize}
\item \textsuperscript{45} \textsc{General Orders of George Washington}, supra note 41, at 30-31.
\item \textsuperscript{47} Stolen Valor Act, 18 U.S.C. § 704 (1994), amended by 18 U.S.C. § 704 (2006) (“Whoever knowingly wears, manufactures, or sells any decoration or medal authorized by Congress for the armed forces of the United States, or any of the service medals or badges awarded to the members of such forces, or the ribbon, button, or rosette of any such badge, decoration or medal, or any colorable imitation thereof, except when authorized under regulations made pursuant to law, shall be fined under this title or imprisoned not more than six months, or both.”).
\item \textsuperscript{48} See Statements on Introduced Bills and Joint Resolutions, 151 Cong. Rec. S12,684, S12,688 (Nov. 10, 2005) (statement of Sen. Kent Conrad) (“Recipients of the Medal of Honor, Distinguished Service Awards, Silver Star, or Purple Heart have made incredible sacrifices for our country. They deserve our thanks and respect. Unfortunately, however, there are some individuals who diminish the accomplishments of award recipients by using medals they have not earned. These imposters use fake medals—or claim to have medals that they have not earned—to gain credibility in their communities. These fraudulent acts can often lead to the perpetration of very serious crimes. Currently, Federal law enforcement officials are only able to prosecute those who wear counterfeit medals. . . . My legislation will allow law enforcement officials to prosecute those who falsely claim, either verbally or in writing, to be medal recipients.”).
\item \textsuperscript{49} See id.
\end{itemize}
receipt of Mrs. Sterner’s submission, Mr. Salazar, a Vietnam veteran, sought to carry out her request. 51

This objective was accomplished by enacting 18 U.S.C. § 704(b), which states,

Whoever falsely represents himself or herself, verbally or in writing, to have been awarded any decoration or medal authorized by Congress for the Armed Forces of the United States, any of the service medals or badges awarded to the members of such forces, the ribbon, button, or rosette of any such badge, decoration, or medal, or any colorable imitation of such item shall be fined under this title, imprisoned not more than six months, or both. 52

Additionally, while the original Act only included an enhanced punishment for unauthorized use of a Congressional Medal of Honor, 53 the amended Act includes enhanced punishment for misrepresentations regarding a Navy cross, an Air Force cross, a Silver Star, or a Purple Heart. 54 Instead of a fine and/or incarceration for up to six months, a misdemeanor of this type is punishable by a fine and/or incarceration for up to one year. 55

Within a short time, the bill had 111 cosponsors in the House of Representatives and twenty-seven cosponsors in the Senate. 56 On December 20, 2006, the amendments were enacted

Currently, military records are not up to par to aid in these investigations so much so that private individuals, such as Mr. Sterner, have taken it upon themselves to investigate and unveil these imposters. See id. Congress is, however, attempting to pass a bill to produce a public online database listing recipients of awards and all individuals who have served in the armed forces. See Military Valor Roll of Honor Act, H.R. 666, 111th Cong. § 1136 (2009).


Stolen Valor Act, 18 U.S.C. § 704(b) (1994), amended by 18 U.S.C. § 704(c) (2006) (“If a decoration or medal involved in an offense under subsection (a) is a Congressional Medal of Honor, in lieu of the punishment provided in that subsection, the offender shall be fined under this title, imprisoned not more than 1 year, or both.”); see also id. (defining a “Congressional Medal of Honor” as a medal awarded under 10 U.S.C. § 3741 (1994), 10 U.S.C. § 6241 (1994), or 10 U.S.C. § 8741 (1994)).

Stolen Valor Act of 2005, 18 U.S.C. § 704(d) (2006) (“If a decoration or medal involved in an offense described in subsection (a) or (b) is a distinguished-service cross awarded under section 3742 of title 10, a Navy cross awarded under section 6242 of title 10, an Air Force cross awarded under section 8742 of title 10, a silver star awarded under section 3746, 6244, or 8746 of title 10, a Purple Heart awarded under section 1129 of title 10, or any replacement or duplicate medal for such medal as authorized by law, in lieu of the punishment provided in the applicable subsection, the offender shall be fined under this title, imprisoned not more than 1 year, or both.”).

Compare 18 U.S.C. § 704(a), and § 704(b), with § 704(c), and § 704(d).

into law and received widespread support. In fact, the Senate unanimously voted in favor of the Act. Two states, using the Stolen Valor Act of 2005 as a model, adopted similar legislation shortly thereafter. Utah adopted the exact language of 18 U.S.C. § 704(b). Similarly, California enacted a statute criminalizing misrepresentations about one's military history if the misrepresentations were made with the intent to defraud. California's legislature also enacted a separate provision requiring public officials of the state to resign if they are convicted under the federal Stolen Valor Act of 2005.

Despite broad support in Congress, some courts have held the Act unconstitutional. As a result, Congress has already begun to revise it. On May 5, 2011, the House of Representatives introduced HR 1775, which repeals section 704(b) and creates an entirely new statute entitled the Stolen Valor Act of 2011.


(2) Any person who intentionally makes a false representation, verbally or in writing, that the person has been awarded a service medal is guilty of a class C misdemeanor.

(3) Any person who wears, purchases, attempts to purchase, solicits for purchase, mails, ships, imports, exports, produces blank certificates of receipt for, manufactures, sells, attempts to sell, advertises for sale, trades, barters, or exchanges for anything of value a service medal, or any colorable imitation thereof, except when authorized by federal law, or under regulations made pursuant to federal law, with the intent to defraud, or with the intent to falsely represent that the person or another person has been awarded a service medal, is guilty of a class C misdemeanor.

60 CAL. PENAL CODE § 532b(c)(1) (West 2011) states, “any person who, orally, in writing, or by wearing any military decoration, falsely represents himself or herself to have been awarded any military decoration, with the intent to defraud, is guilty of a misdemeanor.” While this statute is similar to the federal criminal sanction, the federal version does not require an intent to defraud. See supra notes 52-54 for the language of the federal statute. The absence of such language was a problem addressed by the Ninth Circuit. See United States v. Alvarez, 617 F.3d 1198, 1212 (9th Cir. 2010), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210).
61 CAL. GOV'T. CODE § 3003 (West 2010) (“An elected officer of the state or a city, county, city and county, or district in this state forfeits his or her office upon the conviction of a crime pursuant to the federal Stolen Valor Act of 2005 . . . .”).
63 Id. §§ 1-2.
The proposed statute provides,

Whoever, with intent to obtain anything of value, knowingly makes a misrepresentation regarding his or her military service, shall—(1) if the misrepresentation is that such individual served in a combat zone, served in a special operations force, or was awarded the Congressional Medal of Honor, be fined under this title, imprisoned not more than 1 year, or both; and (2) in any other case, be fined under this title, imprisoned not more than 6 months, or both.  

The proposed legislation provides an exception to prosecution if an individual falsely denies military service. Furthermore, the statute creates a defense “if the thing of value is de minimis” and criminalizes not just false statements about the receipt of military awards but also misrepresentations about one’s military service in general. The bill defines military service as follows:

(A) service in the Armed forces of the United States;

(B) service in a combat zone as a member of the Armed Forces of the United States;

(C) attainment of a specific rank in the Armed Forces of the United States; and

(D) receipt of—

   (i) any decoration or medal authorized by Congress for the Armed Forces of the United States;

   (ii) any of the service medals or badges awarded to members of such forces; or

   (iii) the ribbon, button, or rosette of any such badge, decoration, or medal.

Due to this change in statutory language, the reach of the proposed legislation differs from section 704(b) of the Stolen Valor Act of 2005. For example, expansion of the crime as proposed would enable law enforcement to reach individuals like Blumenthal, who may not have the audacity to claim

---

64 Id. § 2.

65 Id.

66 Id.

67 Id.

68 Id.

falsely that they received a Congressional Medal of Honor, but who falsely claim to have served, for example, in Vietnam. Further, instead of generally criminalizing misrepresentations regarding the receipt of military awards “verbally or in writing,” the proposed bill criminalizes conduct with mens rea components of knowledge and “intent to obtain anything of value.” The absence of a mens rea element in the Stolen Valor Act of 2005 was one of the main concerns of the courts that found it facially invalid under the First Amendment. Interestingly, the 2011 bill does not provide a harm element either, which the Ninth Circuit and District of Colorado both suggested would be necessary for the Act to be constitutional.

II. Application of First Amendment Jurisprudence to the Stolen Valor Act of 2005

The Stolen Valor Act of 2005 has been analyzed under a variety of tests pursuant to First Amendment case law, which has contributed to inconsistent decisions as to the constitutionality of the Act. In anticipation of the Supreme Court’s review of the Act’s constitutionality, this section first demonstrates that the Act regulates protected speech based on its content. Then, this section explicates each of the applicable tests under the First Amendment and applies them to section 704(b). Ultimately, the Act is unconstitutional because it runs afoul of each of these tests and the Supreme Court should strike it down.

---

70 Xavier Alvarez claimed that he had received the Congressional Medal of Honor. United States v. Alvarez, 617 F.3d 1198, 1199 (9th Cir. 2010), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210).
71 Senate candidate Richard Blumenthal claims he misspoke about serving in Vietnam as opposed to in the Military Reserves. Hernandez, supra note 1, at A1. Any service in the military is noteworthy. But when a person claims he served in one (usually higher-ranking) capacity when he actually served in another, it is difficult to fathom that such a misrepresentation would be made unknowingly or unintentionally—unless the serviceman truly “misspoke” and later corrected himself. As a note, courts should not determine which service outweighs another; rather, courts should use the system the military has already established. For the Institute of Heraldry’s order of precedence for military ribbons, see Institute of Heraldry, supra note 44.
74 Alvarez, 617 F.3d at 1216; Strandlof, 746 F. Supp. 2d at 1188.
A. The Stolen Valor Act of 2005 & False Statements of Fact: An Unconstitutional Restriction on Protected Speech

A constitutional analysis of the Stolen Valor Act of 2005 must begin with a threshold determination of whether the First Amendment protects the speech the Act seeks to regulate. In United States v. Stevens, the Supreme Court reaffirmed a list of categories of speech that the First Amendment does not protect: “obscenity, defamation, fraud, incitement, and speech integral to criminal conduct.” Even though the Court acknowledged that this list is not exhaustive, the Court is generally reluctant to create a new category because it does not want to broaden the reach of the government over public discourse. Thus, in determining whether the speech at issue in a government regulation should be afforded an exception to First Amendment protection, the Court first tries to fit the speech into one of the existing categories. If there is no recognized classification, then the Court requires the government to proffer evidence that the speech has not traditionally been protected by the First Amendment—and therefore should not be protected today.

The Stolen Valor Act concerns “false statement[s] of fact,” or lies. Surely “there is no unbridled constitutional right to lie.” But mere lies do not fall under the list of traditionally unprotected categories, nor are they worthy of a newly recognized one. Therefore, the Stolen Valor Act of 2005 regulates protected speech and is not presumptively constitutional under this analysis.

1. Failing to Fit Within Unprotected Categories of Speech

False statements of fact are most closely related to defamation and fraud, but are insufficiently identifiable with either to constitute unprotected speech.

78 Id. at 1586.
79 See id. at 1585.
80 See id. at 1584.
81 See id. at 1585.
83 Id. at 1205.
84 See Stevens, 130 S. Ct. at 1585-86.
85 See id. at 1584 (listing unprotected categories). The Act’s false statements of fact would not fall under Stevens’ remaining three categories. To be obscene, the
Defamatory speech must inflict harm to another’s reputation. If the statements concern a “matter of public concern” or a public official, the plaintiff must show that the statements were made with “actual malice.” Even though false statements can be defamatory, the Stolen Valor Act does not require that false statements be made with a specific intent or a resulting harm. Additionally, despite the Act’s regulation of speech that would, in effect, “defame” the military, the speech does not constitute defamation because the military is a government agency, not an individual. Therefore, false statements of fact under the Act do not fall within the category of defamation.

The fraud category is equally inapplicable. The prosecution in a case of fraud must prove a bona fide harm; fraudulent statements or conduct must “induce another to act to his or her detriment.” As indicated, prosecution under the Stolen Valor Act of 2005 requires no evidence of harm.

Even when comparing the Act to other types of fraud statutes, it still does not fit neatly. For example, 18 U.S.C. § 912, which criminalizes the impersonation of government employees or officers, attaches when the individual “perform[s]. . . acts under the guise of [the] assumed identity.”

speech would have to do with sexuality or sexual desire. Miller v. California, 413 U.S. 15, 24 (1973) (In determining whether a work is obscene, “the trier of fact must [ask] (a) whether the average person, applying contemporary community standards would find that the work, taken as a whole, appeals to the prurient interest; (b) whether the work depicts or describes, in a patently offensive way, sexual conduct specifically defined by the applicable state law; and (c) whether the work, taken as a whole, lacks serious literary, artistic, political, or scientific value.”) To constitute incitement, the speech would have to “persuade[e] another person to commit a crime.” BLACK’S LAW DICTIONARY 347 (3d pocket ed. 2006). To amount to speech integral to criminal conduct, the speech must be “intrinsically related” to a crime. See New York v. Ferber, 458 U.S. 747, 761-62 (1982).

86 BLACK’S LAW DICTIONARY 188 (3d pocket ed. 2006).
87 Id.
89 See Stolen Valor Act of 2005, 18 U.S.C. § 704(b) (2006). While it can be argued that injury to the public may be implied if the speaker is a public official, the Stolen Valor Act does not specifically address public officials. For a more detailed analysis of the harms resulting from false statements made by public officials, see infra Part IV.B.
90 BLACK’S LAW DICTIONARY 300 (3d pocket ed. 2006).
91 18 U.S.C. § 912 (“Whoever falsely assumes or pretends to be an officer or employee acting under the authority of the United States or any agency, department, or officer thereof, and acts as such, or in such pretended character demands or obtains any money, paper, document, or thing of value, shall be fined under this title or imprisoned not more than three years, or both.”).
in an effort to obtain a benefit to which he is not entitled.\footnote{18 U.S.C. § 912.} The Stolen Valor Act of 2005 does not require any specific acts in furtherance of any motive; it requires the mere utterance of words. Additionally, the Act is not comparable to perjury, requiring a “willful” false statement made under oath, or even criminally fraudulent administrative filings, requiring: (1) scienter and (2) the objective to interfere with the proper functioning, or the economic interests, of the government or a private party.\footnote{United States v. Alvarez, 617 F.3d 1198, 1211-12 (9th Cir. 2010) (citing United States v. Dunnigan, 507 U.S. 87, 94 (1993) (invoking a prosecution for perjury under 18 U.S.C. § 1621 (2006), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210); and 18 U.S.C. § 1035 (an example of a federal statute criminalizing falsities when trying to obtain health care benefits)).} Without these extra elements, the false statements of fact criminalized under the Stolen Valor Act of 2005 fail to fall within any of the categories of unprotected speech articulated in Stevens.

2. False Statements of Fact—Their Own Category?

Even though “false statements” under section 704(b) do not fall within any of the currently unprotected categories of speech, the Court could recognize false statements of fact as a new category.\footnote{See United States v. Stevens, 130 S. Ct. 1577, 1584 (2010).} But when analyzed in terms of history and tradition, false statements of fact do not fall within the type of speech that has traditionally remained unprotected by the First Amendment and therefore do not constitute their own category of unprotected speech.

In New York Times v. Sullivan, the Supreme Court noted its tolerance of some “erroneous statement[s].”\footnote{N.Y. Times Co. v. Sullivan, 376 U.S. 254, 271-72 (1964) (stating that “erroneous statement is inevitable in free debate” (citing NAACP v. Button, 371 U.S. 415, 433 (1963))).} Protection of some false speech is required to preserve the “breathing space that [freedoms of expression] need to survive.”\footnote{Id.} In other words, as the Ninth Circuit articulated, “the First Amendment requires that we protect some falsehood in order to protect speech that matters.”\footnote{Alvarez, 617 F.3d at 1203 (quoting Gertz v. Robert Welch, Inc., 418 U.S. 323, 341 (1974)) (internal quotation marks omitted).} Therefore, a category of unprotected speech for all false statements should not be created.

Despite this seemingly supported rationale, the Western District of Virginia, in United States v. Robbins, relied upon...
Gertz v. Robert Welch, Inc. to support its conclusion that false statements of fact constitute their own category of unprotected speech. The Court in Gertz described false statements of fact as “belonging] to that category of utterances... of such slight social value... that any benefit that may be derived from them is clearly outweighed by the social interest in order and morality.” However, more recently in Stevens, the Court seemed to reject the approach utilized in Gertz, stating that “[t]he First Amendment’s guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits.” According to the Supreme Court, the language articulated in Gertz was not a test to be applied to speech when determining whether it is an exception to First Amendment protection; rather, it was dictum. Even if the “speech is not very important” or lacks societal value, the Court seems to say that it cannot be influenced by these factors when determining whether government regulation has stepped too far.

Moreover, the Ninth Circuit, in Alvarez, explained that while some categorical exclusions do comprise false-statement speech, the falsity of speech alone is not enough to take it beyond the scope of First Amendment protection. Despite relying on Supreme Court precedent, Judge Bybee, in his dissent, did not appear to fully refute the majority’s assertion. For example, when arguing that false statements of fact fall outside of First Amendment protection, Judge Bybee partially relied upon Garrison v. Louisiana, which states that “knowingly false statement[s]... do not enjoy constitutional protection.” Yet the false statement in Garrison was defamatory and therefore had to be made with “actual malice.”

---

100 Gertz, 418 U.S. at 340 (quoting Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942)).
102 Id. at 1585-86.
105 Id. at 1219 (Bybee, J., dissenting) (quoting Garrison v. Louisiana, 379 U.S. 64, 75 (1964)).
106 Id.
107 Compare Garrison, 379 U.S. at 75, with Alvarez, 617 F.3d at 1219 (Bybee, J., dissenting).
Overall, false statements of fact, without more, have historically been protected by the First Amendment and should not be classified as their own category of unprotected speech.

B. The Stolen Valor Act of 2005—An Unconstitutional Content-Based Restriction

Since the Stolen Valor Act regulates protected speech, the Act is subject to further constitutional analysis. In addition to the distinction between protected and unprotected speech, the Supreme Court has differentiated between content-neutral and content-based regulations. If the regulation is content-based, then the regulation is presumptively invalid; the government may rebut this presumption by showing that the law passes strict scrutiny, “the most demanding test known to constitutional law.” Under this standard, the Stolen Valor Act of 2005 is unconstitutional because it sets forth a content-based restriction in section 704(b) and fails to satisfy strict scrutiny.

1. Content-Based Versus Content-Neutral Restrictions on Speech

Content-based restrictions prohibit or inhibit a type of speech based solely on its topic, whereas content-neutral restrictions regulate speech in general, regardless of its subject matter. The Supreme Court has confirmed this distinction in a number of cases. In United States v. Stevens, the Supreme Court found a statute that prohibited “visual and audio depictions” in which “a living animal [was] intentionally harmed” to be content-based. In Police Department of Chicago v. Mosley, the Court also found the City’s regulation content-based where the statute prohibited all peaceful picketing next to schools except if the picketing related to a “school’s labor-management


112 See, e.g., Police Dep’t. of Chi. v. Mosley, 408 U.S. 92, 95 (1972); Schact, 398 U.S. at 63; see also Stone, supra note 109, at 278.

113 Stevens, 130 S. Ct. at 1584.
dispute.” The Court in Mosley noted, however, that had the city ordinance prohibited all peaceful picketing without exception, the statute would have been content-neutral.115

The Stolen Valor Act of 2005 prohibits false statements of fact only when the statements misrepresent the individual’s receipt of a military award.116 Similar to the statutes at issue in Stevens and Mosley, the Act requires the government to look at the content of the expression to determine whether the speech is prohibited. Therefore, the Stolen Valor Act of 2005 is a content-based restriction on speech.

2. Applying Strict Scrutiny

As a content-based restriction on protected speech, the Act is subject to strict scrutiny. To pass constitutional muster under strict scrutiny, the government must prove that the statute at issue “serve[s] a compelling state interest and that it is narrowly drawn to achieve that end.”117 Ultimately, the Stolen Valor Act of 2005 fails to satisfy either element of this test.

a. Government Interest

In order to pass strict scrutiny, the government first bears the burden of showing that the statute is supported by a “compelling government interest.”118 A compelling government interest is one “of the highest order”119 and very few government interests are able to meet this standard.120 If there is no “compelling government interest” for the Stolen Valor Act of 2005, the Act will not pass constitutional muster.121 Even though the government may conceivably present a few purposes that support the Act, none is compelling.

The primary purpose of the Act is to protect the “sacrifice, history, reputation, honor and meaning associated with military

114 Mosley, 408 U.S. at 95.
115 See id.
118 Id.
121 Yoder, 406 U.S. at 215.
The Act is also meant to incentivize acts of meritorious bravery. In other words, if anyone can claim that he or she has a military award, the incentive value of the award diminishes. While the Ninth Circuit and the District of Colorado recognized that these are important government interests, they are not compelling. Due to congressional authority to “raise and support armies,” courts generally give more deference to regulations with military purposes, especially when those purposes are “unrelated to the suppression of speech.” However, as the District of Colorado noted, there is no precedent that the Act’s interests are compelling.

National security is the closest recognized compelling interest because incentivizing acts of valor and maintaining the reputation of the military strengthens the nation’s defense. Nevertheless, the absence of the Act’s protection would not create disarray or malfunction great enough to threaten national security. The Act simply seeks to protect a rewards system, albeit a very important and honorable system, but not something that upholds the security of the nation. Additionally, precedent suggests that the putative interests behind the Stolen Valor Act of 2005 are not compelling. In Texas v. Johnson, the Supreme Court struck down a statute that criminalized flag burning because the interest put forward by the government, “preserving the flag as a symbol of national unity,” was insufficiently compelling. According to the Supreme Court, “To conclude that the government may permit designated symbols to be used to communicate only a limited

---

123 Id. at 1190.
124 Id.
125 United States v. Alvarez, 617 F.3d 1198, 1217 (9th Cir. 2010), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210); see Strandlof, 746 F. Supp. 2d at 1189-91.
set of messages would be to enter a territory having no discernible or defensible boundaries.\textsuperscript{131}

Furthermore, the judiciary treats the secondary interest regarding incentives with equal, if not more, disfavor. Both the District of Colorado and the Ninth Circuit were skeptical that a soldier’s act of valor is solely motivated by a medal.\textsuperscript{132} After all, soldiers performed acts of bravery well before the existence of military awards.\textsuperscript{133} By arguing that military decorations incentivize valor, the government in fact diminishes the value of the valorous acts it claims to protect.\textsuperscript{134}

Though not all courts have addressed whether the government’s interests in issuing awards of commendation are compelling, not a single court has held that they are. While the Ninth Circuit avoided a clear holding on this issue, the District Court of Colorado clearly indicated that the government interests are not compelling.\textsuperscript{135} In conclusion, the Stolen Valor Act of 2005 does not serve any “compelling government interest” and therefore fails the first prong of strict scrutiny.

b. Narrow Tailoring

Even if a court finds an asserted government interest compelling, the Stolen Valor Act of 2005 would still fail strict scrutiny because it is not narrowly tailored.\textsuperscript{136} In order to meet this standard, the Act must utilize the “least restrictive means” to further the government interest.\textsuperscript{137} The Stolen Valor Act of 2005 fails to meet this standard due to the vast number of alternative, less restrictive measures the government could have implemented in lieu of the Act.

For example, in Alvarez, the Ninth Circuit suggested that if the Act were restructured to include elements of mens rea and injury, it may pass constitutional muster.\textsuperscript{138} In fact, if the Act were amended to attach only when another individual’s reliance on falsities were to his or her detriment, the Act would more closely resemble a fraud statute (a category of unprotected speech).\textsuperscript{139}
Aside from amending the Act, the court also suggested that there may be no need for a criminal statute at all.\(^{140}\) As declared by the Supreme Court in *Texas v. Johnson*, “If there be time to expose through discussion the falsehood and fallacies, to overt the evil by the processes of education, the remedy to be applied is more speech, not enforced silence.”\(^{141}\) Since the Stolen Valor Act of 2005 punishes the statements solely because they are false, more speech that unveils the falsehood could be sufficient to deter the lies. In turn, law enforcement would not need to waste its time prosecuting.

The Ninth Circuit also suggested that the following were less restrictive means to achieve the government’s goals: “publicizing the names of legitimate recipients or false claimants, creating an educational program, [or] prohibiting the act of posing as a veteran to obtain certain benefits.”\(^{142}\)

Altogether, the government could have used less restrictive means that would not have burdened individuals’ First Amendment rights. Although not all courts have analyzed the tailoring of the Act, the Ninth Circuit’s suggestions appear reasonably valid.\(^{143}\) Overall, the Stolen Valor Act of 2005 fails both prongs of strict scrutiny and is therefore unconstitutional.

C. The Stolen Valor Act of 2005—Unconstitutionally Overbroad and Void-for-Vagueness

Even though the Stolen Valor Act of 2005 should be struck down as a content-based restriction of constitutionally protected speech, the Act may also fail scrutiny under the overbreadth and void-for-vagueness doctrines.\(^{144}\) Under both of these doctrines, the Stolen Valor Act of 2005 is unconstitutional.

1. Overbreadth Doctrine

The First Amendment’s overbreadth doctrine may be used to facially challenge a statute if the statute is either a content-neutral or a content-based restriction on protected speech or if it regulates unprotected speech.\(^{145}\) According to the

\(^{140}\) *Alvarez*, 617 F.3d at 1216.


\(^{142}\) *Alvarez*, 617 F.3d at 1210.

\(^{143}\) Id. at 1217.


Supreme Court, “the party challenging the law must demonstrate not just from the text of the statute, but also from actual fact that a substantial number of instances exist in which the law cannot be applied constitutionally.” If the facial challenge is successful, “the prosecution fails regardless of the nature of the defendant’s own conduct, because [it] renders a statute unconstitutional and invalid in all its applications.”

Because an overbreadth challenge will render an entire statute invalid even if the statute has no unconstitutional effect as applied, the Court is generally reluctant to strike down legislation under this doctrine. Despite this reluctance, the Stolen Valor Act of 2005 may be struck down as facially overbroad.

As recognized by the dissent in Alvarez, penned by Judge Bybee, there are at least two ways section 704(b) may be overbroad. First, Judge Bybee conceded that the Act may reach “inadvertent violations of the act” due to the lack of a scienter requirement. Second, the Act gives no exception for “satire or imaginative expression,” both historically protected speech. Whether these mistaken and theatrical statements are substantial enough to create a realistic threat is a difficult question to answer. However, it appears that the Act under these circumstances would cut into private conversations and chill speech historically protected by the First Amendment.

To determine whether either of these examples of overreaching is sufficient to support an overbreadth challenge, the first step is to ask “whether the Stolen Valor Act actually covers” mistaken or theatrical statements. When interpreting statutes under this doctrine, courts do “not lightly assume that Congress intended to infringe constitutionally protected liberties” and therefore look for a “limiting construction” that

---

147 Alvarez, 617 F.3d at 1236 (Bybee, J., dissenting) (quoting Wurtz v. Risley, 719 F.2d 1438, 1440 (9th Cir. 1983); Bd. of Trs. of State Univ. of N.Y. v. Fox, 492 U.S. 469, 483 (1989)).
148 United States v. Esparza-Ponce, 193 F.3d 1133, 1137 (9th Cir. 1999) (quoting Broadrick, 413 U.S. at 613).
149 Alvarez, 617 F.3d at 1236 (Bybee, J., dissenting).
150 Id. at 1236-37.
151 Id. at 1222.
154 Alvarez, 617 F.3d at 1237 (Bybee, J., dissenting).
could avoid overbreadth. Judge Bybee, in his Alvarez dissent, and Judge Jones, in United States v. Robbins, used two different approaches to limit the Stolen Valor Act’s construction. Neither of the constructions seems sufficient to overcome an overbreadth challenge.

In Robbins, Judge Jones of the Western District of Virginia asserted that the Stolen Valor Act of 2005 should be read to only address “knowingly false statements” where “the defendant intended to deceive.” Judge Jones concluded that this limiting construction would not suppress ambiguous, mistaken, or misunderstood statements, nor punish “false statement in fictions, in parody, or as rhetorical hyperbole;” “only outright lies, not ideas, [would be] punishable.”

While this approach does limit the Act to avoid overbreadth, the construction also seems to be impermissible. Perhaps interpreting the statute to include a “knowing” element would be permissible, but assuming an “intent to deceive” does not seem appropriate when no actual words in the statute support that notion. In support of the specific intent, Judge Jones cited two cases. Each case involved facial challenges against a statute criminalizing false representations of United States citizenship. In the first case, the Ninth Circuit in United States v. Esparza-Ponce considered statutory language “mak[ing] it a crime for anyone to knowingly and falsely represent himself to be a citizen of the United States without regard” to whom the statement was made. The statute should be read, the Court held, to require proof that “the person to whom [the] false statement was made had good reason to inquire into the nationality status.” In the second case, the Second Circuit held in United States v. Achtner that “representation of citizenship must still be made to a person having some right to inquire or adequate reason for

157 Robbins, 759 F. Supp. 2d at 819.
158 Id.
159 Id. (citing United States v. Esparza-Ponce, 193 F.3d 1133, 1137-38 (9th Cir. 1999); United States v. Achtner, 144 F.2d 49, 52 (2d Cir. 1944)).
160 Id.
161 Esparza-Ponce, 193 F.3d at 1137 (quoting Smiley v. United States, 181 F.2d 505, 507-08 (9th Cir. 1950)).
ascertaining a defendant's citizenship. In both of these cases, the courts pointed out that the limiting construction they applied to the statute already existed prior to the facial challenges brought before them. Using the Supreme Court's analysis in Reno v. ACLU, the Ninth Circuit noted that even though courts may impose limiting constructions if there is not one already in place, to impose a new one, the statute must be “readily susceptible to [the] construction.”

No court had implemented a limiting construction of the Stolen Valor Act of 2005 prior to Judge Jones's attempt in Robbins. Further, section 704(b) does not appear to be “readily susceptible” to the limiting construction Judge Jones sought to employ. To begin with, neither of the cases Judge Jones cited limits the construction of the statute to include a specific intent of the perpetrator. Additionally, no words or phrases in the statute imply that Congress meant to require the specific intent to deceive.

In ACLU of Georgia v. Miller, the Northern District of Georgia rejected the government's suggestion to engraft onto a statute the specific intent to deceive or defraud. The statute made it a crime for

any person . . . knowingly to transmit any data through a computer network . . . for the purpose of setting up, maintaining, operating, or exchanging data with an electronic mailbox, home page, or any other electronic information storage bank or point of access to electronic information if such data uses any individual name . . . to falsely identify the person . . .

The district court noted that “[b]y its plain language the criminal prohibition applies regardless of whether a speaker has any intent to deceive or whether deception actually occurs.” Additionally, the phrases intent to deceive and intent to defraud appear nowhere in the language of the statute, despite the “express[] indu[slion] [of such phrases] in other Georgia criminal
statutes which require proof of specific intent." Moreover, the district court asserted that although the word falsely does appear in the statute, it is not synonymous with intent to deceive or intent to defraud. "Falsely’ means merely ‘wrongly,’ ‘incorrectly,’ or ‘not truthfully.’

Here, the plain language of the Stolen Valor Act of 2005 applies “regardless of whether a speaker has any intent to deceive." The Act omits the phrases intent to deceive and intent to defraud—phrases that Congress expressly uses in other federal criminal statutes. Furthermore, while the Act uses the word falsely, the Act cannot be construed to require a specific intent based on the use of this word because falsely is by no means synonymous with an intent to deceive. Thus, the Stolen Valor Act of 2005 does not appear to be “readily susceptible” to Judge Jones’s limiting construction requiring a specific intent.

In Alvarez, Judge Bybee applied a different limiting construction. But the construction is still insufficient to satisfy the overbreadth doctrine. Judge Bybee based his limiting construction upon the interpretation of the word represents in the Act, calling upon Webster’s Dictionary’s “first definition of the word . . . ‘[t]o bring clearly before the mind; [to] cause to be known . . . [to] present esp. by description.’” With this definition in mind, Judge Bybee asserted that “an ambiguous statement that could conceivably be misinterpreted to claim receipt of a military award could not be punished under the Act because such a statement would not ‘bring clearly before the mind’ of the listener that the speaker has described himself as having won the award.” He also argued that the Act would not apply under his construction to “satirical or theatrical
statements claiming receipt of a military award” because any such statements would be “entirely untrue.”

Concededly, Judge Bybee’s limiting construction is compelling; it could easily be implemented to restrict the reach of the Stolen Valor Act of 2005. However, Judge Bybee seemed to focus on a very narrow portion of the meaning of represent and creates a misimpression as to how effective this limiting construction would be. For example, Webster’s Dictionary also indicates that represent is defined as “to serve as a sign or symbol of,” “to portray or exhibit in art,” “to serve as the counterpart image of,” “to produce on the stage,” “to act the part or role of,” and “to describe as having a specified character or quality.” Additionally, while Webster’s Dictionary’s first definition is “to bring clearly before the mind,” Oxford Dictionaries lists first, to “be entitled or appointed to act or speak for (someone), especially in an official capacity.” That first definition is followed by “constitute; amount to” and “depict (a particular subject) in a picture or other work of art.” In the legal context, the Federal Circuit has found the definition of represent to include, “to be an accredited deputy or substitute for (a number of persons) in a legislative or deliberative assembly,” “to describe as having a specified character or quality; to give out assert or declare to be of a certain kind,” and “to symbolize, to serve as a visible or concrete embodiment.” With this fuller understanding of the meaning of represent, the possibility that the Act would still impermissibly punish works of satire, fiction, and parody seems markedly clear.

In addition, the Act would still apply to mistaken remarks regarding receipt of military awards because the definition Judge Bybee relies upon focuses on the point of view of the listener, not the speaker. An individual prosecuted under the Act need not even realize that he or she created a misimpression,

---

179 Id. at 1240-41.
182 Id.
but the statement he or she made still may have “bring[ught] [that
misimpression] clearly before the mind” of the listener.\footnote{Alvarez, 617 F.3d at 1238 (Bybee, J., dissenting). Perhaps if this limiting construction were combined with a requirement that the individual knowingly make the false statements, then the Act would pass constitutional muster under the overbreadth doctrine. However, while knowingly may be a construction the Act is readily susceptible to, Congress has expressly included a knowledge requirement in other statutes when it intended for knowledge to be an element of the crime.}

Overall, the two limiting constructions offered by Judge Bybee and Judge Jones do not appear to prevent the Act’s application to mistaken or theatrical statements. Even though it is difficult to determine whether these applications of the Act are substantial enough, it appears that no reasonable limiting construction would be able to refute a facial challenge of the overbreadth of the Stolen Valor Act of 2005. Although no court has so held, the Act is facially invalid under the First Amendment because it is overbroad.

2. Void-for-Vagueness Doctrine

In general, when the Court applies the overbreadth doctrine, it also analyzes the regulation in terms of void-for-vagueness.\footnote{See Vill. of Hoffman Estates v. Flipside, Hoffman Estates, Inc., 455 U.S. 489, 494-504 (1982); Grayned v. City of Rockford, 408 U.S. 104, 108-14 (1972).} Under this analysis, a law must “state explicitly and definitely what acts are prohibited, so as to provide fair warning and preclude arbitrary enforcement.”\footnote{BLACK’S LAW DICTIONARY 754 (3d pocket ed. 2006).} The Supreme Court has asserted that the more important element of the two is the principle regarding enforcement guidelines proffered by the legislature.\footnote{Kolender v. Lawson, 461 U.S. 352, 357-58 (1983).} “Where the legislature fails to provide such minimal guidelines, a criminal statute may permit ‘a standardless sweep [that] allows policemen, prosecutors, and juries to pursue their personal predilections.’”\footnote{Id. (quoting Smith v. Goguen, 415 U.S. 566, 575 (1974)).}

In his overbreadth analysis of the Act, Judge Bybee contended that the term represents guides law enforcement in restricting the sweep of the statute.\footnote{See Alvarez, 617 F.3d at 1237-38 (Bybee, J., dissenting).} However, reliance upon law enforcement’s interpretation of a word (especially a word that is defined in a plethora of ways) does not completely prevent arbitrary enforcement. Thus, in light of the way the Stolen Valor Act of 2005 is written, there is nothing to stop its arbitrary enforcement without “explicit[]” or “definite[]” guidelines.
Furthermore, the Act does not provide citizens “actual notice” of what statements are truly punishable within the ambiguities of the word represents.\textsuperscript{190} Because of this lack of notice, the Act has the ability to “chill” speech in contravention of First Amendment purposes. Only Judge Bybee, in his dissent, has analyzed the Act under the void-for-vagueness doctrine. Despite this, it seems apparent that the Act is void-for-vagueness and unconstitutional.

III. CONGRESS’S IMPERFECT PROPOSAL—THE STOLEN VALOR ACT OF 2011

Although the foregoing constitutional analyses provide a basis to strike down the Stolen Valor Act of 2005,\textsuperscript{191} in an attempt to fix the flaws in section 704(b), Congress proposed a new statute entitled the Stolen Valor Act of 2011.\textsuperscript{192} While the proposed bill provides a better means to police speech that steals valor, Congress has not addressed all of the flaws the courts have raised and may still be unable to receive a constitutional consensus among the courts.

A. The Stolen Valor Act of 2011 Would Still Regulate Protected Speech

Even though Congress attempted to distinguish the proposed bill from section 704(b) by criminalizing “misrepresentations” rather than “false claims,”\textsuperscript{193} “misrepresentations” like “mere lies” or “false statements of fact” would likely not be deemed to hold their own category of unprotected speech under Stevens.\textsuperscript{194} The “misrepresentations” in the proposed bill, however, do not completely stand alone; the bill requires specific “intent to obtain anything of value.”\textsuperscript{195} While this added element may aid in avoiding First Amendment scrutiny, intent alone would likely not be sufficient without a requisite harm.

In his amicus curiae brief in United States v. Strandlof, Eugene Volokh argued that the Stolen Valor Act of 2005, if

\textsuperscript{191} See supra Part II.
\textsuperscript{194} See supra Part II A.
\textsuperscript{195} H.R. 1775.
Professor Volokh asserted that people who lie about decorations generally do so for a reason: They may want to get elected to public office, or to get more credibility for their own statements in another's election campaign, or to get more credibility in some nonelectoral political debate, or even just to get more respect from neighbors, acquaintances, and potential business associates. They are thus trying to manipulate people's behavior through falsehood, and their false claims are quite likely to indeed affect others' behavior (especially since having a military decoration is often seen as an especially important mark of good character).196

Professor Volokh suggests that manipulating "private citizens' behavior through falsehoods is a significant enough harm" on its own.197 Therefore, there is no need to eliminate an express harm or injury requirement in the statute.198 While Professor Volokh's argument is consistent with the perceived unsavory nature of these misrepresentations, in striking the law down, the Ninth Circuit and the District Court of Colorado both found it persuasive that the Stolen Valor Act of 2005 does not require harm to be proven.199

This contradiction may be due to Clay Calvert and Rebekah Rich's suggestion that, after Stevens, the courts have shifted their constitutional analysis under the First Amendment from a more "value-based methodology"—which examines the societal value of the speech at issue and only protects speech that matters—to a "causation-of-harm-based methodology," which focuses upon "the direct nexus (if any) between that speech and the alleged harms to humans that it causes."200

This understanding of Stevens cuts against an argument that the misrepresentations contemplated by the Act should be given their own category of unprotected speech without a requisite harm. Further, there is no historical evidence that misrepresentations with specific intent but without requisite harm have been unprotected by the First Amendment.201

---

197 Id. at 354.
198 See id.
201 See supra Part II.A.2.
courts’ recent shift to a “causation-of-harm-based methodology”\textsuperscript{202} coupled with the reluctance of lower courts to carve out new categories of unprotected speech\textsuperscript{203} suggest that courts will not find that the Stolen Valor Act of 2011 regulates unprotected speech.

B. The Proposed Bill Would Still Be a Content-Based Restriction Subject to Strict Scrutiny

In addition to regulating protected speech, just as section 704(b) was a content-based restriction,\textsuperscript{204} the proposed legislation, which regulates misrepresentations regarding one’s military service, would also be a content-based restriction. Therefore, the proposed bill would likely be subject to strict scrutiny.\textsuperscript{205}

The government’s purported interest behind the Stolen Valor Act of 2011 would likely go further than the Stolen Valor Act of 2005. Potentially, the government could argue that the proposed bill protects not just the reputation of the awards but the reputation of the entire military. This expanded government interest may be found compelling under a strict scrutiny analysis, despite a lack of precedent. First, the interest does not run afoul of \textit{Texas v. Johnson}, which took issue with “preserving the flag as a symbol of national unity,”\textsuperscript{206} because “symbols” such as the military awards and decorations of section 704(b) are not the only content regulated by the proposed legislation. Additionally, because the government interest is a broader military purpose, it may receive the deference that courts generally give to military purposes in line with the congressional authority to “raise and support armies.”\textsuperscript{207}

Despite the seemingly compelling nature of this government interest, preserving the reputation of the military, a government entity, presents a conflict with a traditional First Amendment policy of preventing statutes providing for “government . . . self-preservation.”\textsuperscript{208} As Geoffrey Stone

\textsuperscript{202} Calvert & Rich, supra note 50, at 4.
\textsuperscript{203} See supra Part II.A.2.
\textsuperscript{204} See supra Part II.B.1.
\textsuperscript{205} See United States v. Stevens, 130 S. Ct. 1577, 1584 (2010).
contends, in striking down self-preserving laws, the Court not only wishes to prevent legislators from using “the power of government to intimidat[e] [and] silence its critics,” but the Court also does not want the government “to dominate and manipulate public debate.” The government may breach this principle outwardly, or it may do so pretextually, that is, by stating a permissible purpose while trying, in fact, to suppress opposition. Either way, the Court has illustrated that such regulations are unconstitutional.

For example, in Schact v. United States, the Court struck down a law that prohibited actors from wearing accurate military uniforms in theatrical productions that negatively portrayed the military. This law was not only an attempt to censor critics of the Vietnam War, but was also an attempt to dominate the public arena with only positive associations between soldiers and the war.

While the statute in Schact seemed to be enacted with the express purpose to preserve the government, the Stolen Valor Act of 2011 would not seem to have the same purpose. The proposed legislation is a response to inconsistent court rulings regarding the Stolen Valor Act of 2005’s constitutionality. Further, the Stolen Valor Act of 2005 was a response to the increased crime perpetuated by the false representations. Moreover, unlike the statute in Schact, the misrepresentations of the Stolen Valor Act of 2011 do not depict the military in a negative light. Thus, the bill does not aid in “silencing [government] critics,” nor does it “manipulate public

---

209 Stone, supra note 109, at 277.
210 See, e.g., Stevens, 130 S. Ct. at 1582, 1592 (holding 18 U.S.C. § 48, which criminalized the creation, sale, or possession of certain depictions of animal cruelty, unconstitutional).
211 See Stone, supra note 109, at 277-78 (discussing the “pretext effect’s strong suspicion of any government regulation that is consistent with . . . an impermissible motive such as suppressing particular ideas because they don’t want citizens to accept those ideas in the political process”).
212 See, e.g., Stevens, 130 S. Ct. at 1582, 1592 (holding 18 U.S.C. § 48, which criminalized creation, sale, or possession of certain depictions of animal cruelty, unconstitutional).
214 See id.
217 Davenport, supra note 17.
218 Schact, 398 U.S. at 63.
debate.\textsuperscript{220} Even though the Stolen Valor Act of 2011 could be perceived, as a matter of policy, as an impermissible regulation, policing misrepresentations regarding one’s military service does not seem to amount to actual government self-preservation.

Even if the Court deems the government’s interest compelling, the bill would still be subject to a narrow tailoring analysis under strict scrutiny. With the specific intent requirement, the proposed legislation is a less restrictive measure than the currently enacted section 704(b). However, without requisite harm the legislation does not seem to provide the “least restrictive means” to achieve the compelling government interest at stake.

As already indicated, Calvert and Rich suggest that harm caused by speech is now the central focus of First Amendment analysis.\textsuperscript{221} Even though Volokh would likely suggest that misrepresentations themselves produce a significant enough harm to pass constitutional muster,\textsuperscript{222} under a strict scrutiny analysis it is unlikely that a court invoking the “causation-of-harm-based methodology” would uphold the proposed legislation without a more express harm indicated in the statute.\textsuperscript{223} Therefore, the Stolen Valor Act of 2011 would likely be held unconstitutional under strict scrutiny and would not completely resolve the conflicts between the Stolen Valor Act of 2005 and the First Amendment.

C. The Stolen Valor Act of 2011 Would Not Be Overbroad or Void-for-Vagueness

In the event that the Stolen Valor Act of 2011 is deemed a regulation of unprotected speech, the Act would be subject to the overbreadth and void-for-vagueness doctrines instead of strict scrutiny.\textsuperscript{224} Under these two analyses, however, the Court would likely hold the proposed legislation constitutional, unlike the currently enacted section 704(b).

The proposed legislation not only requires that the misrepresentations be made knowingly but also that the misrepresentations be made with the specific intent to obtain

\textsuperscript{220} Stone, supra note 109, at 277.
\textsuperscript{221} Calvert & Rich, supra note 50, at 4.
\textsuperscript{222} See Volokh, supra note 196, at 353.
\textsuperscript{223} Calvert & Rich, supra note 50, at 4.
anything of value. In interpreting the Stolen Valor Act of 2011, a limiting construction would not be necessary because the specific intent already restricts the reach of the proposed legislation to reduce its scope. The proposed act does not regulate speech in a manner that is overbroad like the Stolen Valor Act of 2005. Thus, the proposed legislation would not be unconstitutionally overbroad.

Similarly, with the addition of the specific intent requirement, more specific guidelines would be in place to prevent law enforcement from “pursu[ing] their personal predilections” and arbitrarily enforcing the criminal statute. Therefore, the proposed legislation would pass constitutional muster under the void-for-vagueness doctrine.

IV. Suggested Legislative Measures to Combat Speech that Steals Valor

In proposing the new legislation, Congress recognized the value in policing speech that steals valor and the need for reconstructing the Stolen Valor Act of 2005 to avoid running afoul of the First Amendment. However, Congress’s proposed legislation does not appear to sufficiently amend the Act to pass constitutional muster. This note suggests two measures that will better police speech that steals valor. First, Congress should redraft the proposed legislation to resemble a fraud statute. Second, state and local legislatures should impose an eligibility requirement for individuals seeking to run for public office.

A. Fraudulent Misrepresentations of Military Content—A Better Resolution

To transform the Stolen Valor Act of 2011 into a fraud statute, Congress would need to add a harm element to the statute. Calvert and Rich suggest that Congress should draft a fraud statute that includes a “monetary harm” requirement. But the statute may be more effective if it were to include a less specific form of harm (such as obtaining a benefit that the perpetrator is

---

226 See supra Part II.C.1.
227 See supra Part II.C.2.
228 See H.R. 1775, § 2.
229 BLACK’S LAW DICTIONARY 300 (3d pocket ed. 2006).
230 Calvert & Rich, supra note 50, at 34.
not entitled to). While a monetary harm may provide stricter guidelines for law enforcement to follow, as Volokh points out, these military lies are made for a variety of purposes, not just for money. Additionally, Congress should eliminate the specific intent it has imposed in the 2011 Act and only cover knowing misrepresentations. The fraud statute would also be able to invoke the Stolen Valor Act of 2011’s broader application to not just false representations about receiving military awards, but also misrepresentations about one’s military service.

By restructuring the statute to resemble fraud, Congress would be able to regulate unprotected speech and would only be subject to the overbreadth and void-for-vagueness doctrines, not strict scrutiny. Requiring knowledge and harm would aid in satisfying these doctrines because the statute would not punish inadvertent or mistaken misrepresentations—or satirical or imaginative expression. The special fraud statute would also avoid granting law enforcement too much discretion, which would lead to arbitrary enforcement.

There are a few conceivable objections to creating a special fraud statute, but they do not appear to be compelling. First, some may argue that criminal sanctions are ineffective or too harsh. As already indicated, in Texas v. Johnson, the Supreme Court asserted that “[i]f there be time to expose through discussion the falsehood and fallacies, to overt the evil by the processes of education, the remedy to be applied is more speech, not enforced silence.” However, “more speech” is evidently an insufficient deterrent due to the increased incidence relating to these misrepresentations.

Second, implementing a special fraud statute may be moot because there are already general fraud statutes in

---

232 See Volokh, supra note 196, at 353-54.
233 See Part II.A.1 (describing fraud as requiring knowledge and a bona fide harm).
236 See supra Part II.C.1 (analyzing the Stolen Valor Act of 2005 under the overbreadth doctrine).
237 See supra Part II.C.2.
239 Id.
240 Davenport, supra note 17.
A separate and distinct fraud statute addressing misrepresentations of military content would, however, ensure that the problems arising from such statements would without question be addressed. Moreover, Congress can impose an enhanced punishment for committing this specialized fraudulent act. In turn, prosecutors can not only charge a separate count on an indictment (and perhaps an extra conviction), but also incur greater deterrence. Furthermore, a fraud statute would be more effective than relying upon existing impersonation statutes. That is, fraud does not require the same requisite degree of conduct in furtherance of the misrepresentations.

Third, as a matter of policy, some may argue that this specialized fraud statute is inconsistent with First Amendment principles. The Supreme Court has specifically tried to reveal and quash government interference with speech that attempts government self-preservation, government suppression of the democratic “marketplace of ideas,” and government repression of speech in times of crisis. In terms of government self-preservation, the Court is mainly concerned with legislators using “the... power of government to intimidat[e] [and] silence its critics.” Here, the purpose behind the specialized fraud statute is concededly to protect the reputation of the military, a government entity. However, the misrepresentations at issue do not portray or criticize the military in a negative light but rather serve as a tool to use military stature to one’s benefit. As for suppression of the “marketplace of ideas,” the Court seeks to uphold the people’s democratic self-governance by maintaining a free and open public forum for speech. When the government seeks to restrict speech that “convey[s] a political message, ... a matter of public concern, or ... a viewpoint or opinion” regardless of any political underpinnings, the

241 See supra Part II.A.1. The implementation of a fraud statute in this context may also be moot if the Supreme Court deems the Stolen Valor Act of 2005 constitutional in its review of United States v. Alvarez. United States v. Alvarez, 617 F.3d 1198 (9th Cir. 2010), cert. granted, 80 U.S.L.W. 3098 (U.S. Oct. 17, 2011) (No. 11-210).
243 See supra Part II.A.1.
244 Stone, supra note 109, at 277-78.
245 Id. at 277.
246 For example, in Schacht v. United States, 398 U.S. 58, 63 (1970), the Court struck down a law prohibiting actors from wearing accurate military uniforms in theatrical productions that negatively portrayed the military.
Court seeks to protect it. While misrepresentations about one's military service could conceivably be considered a “matter of public concern,” a special fraud statute would not chill or discourage speech because of the statutory limitations upon the punishable speech. Perhaps the Stolen Valor Act of 2005 chills speech because it reaches a broader range of lies including mistaken or satirical false statements, but the special fraud statute would not. An additional type of regulation the Court generally tries to prevent as a matter of policy arises “in times of crisis, real or imagined, [because] citizens and government officials tend to panic, to grow desperately intolerant, and to rush headlong to suppress speech they can demonize as dangerous, subversive, disloyal, or unpatriotic.” This type of government regulation was present, for example, during World War I, when states attempted to promote national unity in response to xenophobic panic. While the special fraud statute would be enacted partly in response to the increased amount of incidence relating to false statements of military commendation, the statute would not be the type of poorly thought-out and excessive regulation the Court has sought to strike down. State and federal legislatures respond to increases in crime rates all the time. Thus, increased crime rates do not always constitute the type of crisis regulations the Supreme Court seeks to prevent.

Therefore, restructuring the proposed legislation to resemble a fraud statute is a less restrictive and a more likely constitutional alternative to the existing Act. Further, a fraud statute would still be an effective means of meeting the government interests at stake.

---

250 Stone, supra note 109, at 278.
252 Davenport, supra note 17.
253 See Stone, supra note 109, at 277-78.
254 See, e.g., City of Los Angeles v. Alameda Books, Inc., 535 U.S. 425, 435 (2002) (holding that the government could reasonably rely on a study indicating a significant crime rate relating to the adult entertainment industry in order to regulate the industry by restricting speech).
B. Holding Public Officials Accountable

In addition to drafting a specific fraud statute, this note calls upon state and local legislatures to enhance eligibility requirements for individuals seeking to serve as elected officials. Public officials, such as Blumenthal and Kirk, have used their military status to further their political goals even when that status, in reality, is not what they made it out to be. Despite the exposure of their lies, these individuals are still being elected into office, which seems unsavory. After all, these individuals are seeking a position of power in a government while at the same time diminishing the honor of the military that serves to protect it.

The Constitution currently sets forth the minimum requirements to run for an elected position in the federal government. The Constitution requires of candidates a minimum age, a minimum time of citizenship, and legal residence within the area for which the candidate seeks to serve. In addition to these bare essentials, each state has general eligibility requirements for federal, state, and local level positions. For example, some states require that a candidate not be a convicted felon. In light of the recent political scandals, this note suggests that state and local legislatures should impose an additional eligibility requirement that individuals not misrepresent their military service. For example, the eligibility requirement could provide the following:

No person shall qualify as a candidate for elective public office or maintain a public office if already elected in the state of XX who has knowingly made misrepresentations regarding his or her military service during the course of any campaign for nomination or election to public office, by means of campaign materials, including an advertisement on radio, television, or the Internet, or in a newspaper or periodical, a public speech, or press release, with the intent to promote the election of that person.

See, e.g., Hernandez, supra note 1, at A1; Smith, supra note 9.
256 Haigh, supra note 16; Lighty & Secter, supra note 16.
257 U.S. CONST. art. I, § 2, cl. 2; U.S. CONST. art. I, § 3, cl. 3.
258 U.S. CONST. art. I, § 2, cl. 2; U.S. CONST. art. I, § 3, cl. 3.
259 See, e.g., N.J. ADMIN. CODE § 19:3-5 (2009); Wis. CONST. art. XIII, § 3.
260 See, e.g., Wis. CONST. art. XIII, § 3, cl. 2 ("No person convicted of a felony, in any court within the United States, no person convicted in federal court of a crime designated, at the time of commission, under federal law as a misdemeanor involving a violation of public trust and no person convicted, in a court of state, of a crime designated, at the time of commission, under the law of the state as a misdemeanor involving a violation of public trust shall be eligible to any office of trust, profit or honor in this state unless pardoned of the conviction.").
This eligibility requirement is similar to a criminal statute upheld by the Ohio Court of Appeals in State v. Davis:

(B) No person, during the course of any campaign for nomination or election to public office or office of a political party, by means of campaign materials, including sample ballots, an advertisement on radio or television or in a newspaper or periodical, a public speech, press release, or otherwise, shall purposely do any of the following:

. . . .

(10) post, publish, circulate, or distribute a written or printed false statement, either knowing the same to be false or with reckless disregard of whether it was false or not, concerning a candidate that is designed to promote the election, nomination, or defeat of the candidate.

In Davis, the court examined this criminal sanction under strict scrutiny and found that the state had a compelling interest “to promote honesty in the election of public officers.” According to the court, “[f]reedom of speech does not include a right to purposely, with knowledge of its falsity, publish a false statement about a candidate for public office with the intent to promote the election or defeat of such candidate.” The court also found the criminal statute narrowly tailored because it “expressly limits a conviction to cases where there is proof that the statements were known to be false or were made in reckless disregard of their falsity.”

Despite this supportive precedent, holding public officials accountable for these misstatements, concededly, would be difficult. Other cases do not uphold a state’s interest in criminalizing false political campaign speech. In fact, one court found that “[t]he constitutional guarantee of free speech has its ‘fullest and most urgent application in political campaigns.’” For example, in Brown v. Hartlage, the Supreme Court struck down a criminal statute that sanctioned an elected official’s unfulfilled and false campaign promises. Even though the Court found that the State had legitimate interests in ensuring “that its governing political institutions
and officials properly discharge public responsibilities and maintain public trust...[and] in upholding the integrity of the electoral process itself,” the Court did not find the interest to be compelling. Additionally, in Washington ex. rel. Public Disclosure Commission v. 119 Vote No! Committee, the Washington Supreme Court analyzed a criminal statute that sanctioned political candidates who “sponsor[ed] with actual malice...[p]olitical advertising that contain[ed] a false statement of material fact.” This statute was similar to the statute at issue in Garrison v. Louisiana, which penalized statements criticizing a public official’s conduct when made with actual malice. In both cases the statutes were struck down. In Vote No!, the court vigorously argued against the criminalization of false campaign speech, asserting that “[i]n political campaigns the grossest misstatements, deceptions, and defamations are immune from legal sanction unless they violate private rights—that is, unless individuals are defamed.” In addition to adverse precedent, Congress has exempted false political campaign advertisements from the Federal Trade Commission sanctions upon other types of false advertising. Moreover, numerous scholars suggest that sanctioning false campaign speech would be an ineffective method because it would fuel voter alienation; afford yet another avenue for one candidate to attack her opponent; and lead to extensive litigation, civil or criminal, that would continue long after the election is over.

267 Id. at 52.
268 119 Vote No!, 957 P.2d at 693 (quoting WASH. REV. CODE ANN. § 42.17.530(1)(a) (West 1998)).
270 Id.; 119 Vote No!, 957 P.2d at 699.
272 See Undertaking Spam, Spyware, and Fraud Enforcement With Enforcers Beyond Borders Act of 2006, 15 U.S.C. § 45 (2006). Interestingly, these sanctions are imposed upon false advertising regarding commercial goods in an effort to protect consumers from making an ill-informed transaction. See F.T.C. v. Cinderella Career & Finishing Schs., Inc., 404 F.2d 1308, 1313 (D.C. Cir. 1968); Slough v. F.T.C., 396 F.2d 870, 872 (5th Cir. 1968). While a voter is not making an economic transaction, a voter is making a choice; a choice regarding who should be elected into a position of power over them. Even though Congress did not seem to think that the same protections need to be afforded to American voters, it seems as though false advertising in a political campaign can have the same effect of inducing an ill-informed decision. The counterargument, however, is that the American people are supposed to be the truth-finders during an election in order to uphold democratic principles, not the government. See Eu v. S.F. Cnty. Democratic Cent. Comm., 489 U.S. 214, 223-24 (1989).
However, the courts, Congress, and scholars have generally addressed false campaign speech that one candidate or political organization puts forth about the opponent. Here, the suggested eligibility requirement targets false statements about the individual's own résumé. In this situation, public officials are certainly able to bear this burden. Further, targeting this type of speech does not raise the same concerns of voter alienation and extensive litigation among opponents. The eligibility requirement would also circumvent the constitutional problems implicated by criminally sanctioning campaign speech because the “integrity of the electoral process” would be preserved in a less restrictive manner.

In addition to a lack of criminalization, the eligibility requirement is further distinguishable from all of the statutes at issue in Brown, Garrison and Vote No!. In Brown, the statute sanctions speech that operates prospectively, whereas the misrepresentations in the eligibility requirement would target speech about one's past. In Garrison, the statute targeted statements made with actual malice, without regard to whether those false statements were made knowingly or were in fact false. Thus, the statute criminalized truthful statements as well as inadvertent or mistaken false statements. In Vote No!, the statute sanctioned false statements made with actual malice. While the statute did not criminalize truthful statements, it still had the ability to sanction inadvertent or mistaken false statements.

Thus, the imposition of this eligibility requirement, which expressly prohibits individuals from obtaining or maintaining a government position if they misrepresent their military history, would likely provide a permissible and effective source of accountability.

Interestingly, if the eligibility requirement and the specialized fraud statute were enacted together, the eligibility requirement would enhance the effectiveness of the fraud statute. Prosecutors could not charge political candidates under the

---

274 See, e.g., 119 Vote No!, 957 P.2d 691; Marshall, supra note 273; Kimmel, supra note 273; Calvert, supra note 273.
276 Id. at 54.
279 119 Vote No!, 957 P.2d at 693.
280 Id. at 696.
specific fraud statute due to the difficulty in establishing individualized detrimental reliance amongst voters. But the eligibility requirement would aid in expanding the reach of the fraud statute by holding politicians, such as Blumenthal or Kirk, accountable for their misrepresentations. Additionally, if public officials are held accountable for their misrepresentations, the publicity such political scandals receive will aid in enhancing the deterrence of the same behavior in private citizens by illustrating the ramifications of this behavior and exemplifying equality of treatment amongst the rulers and the people.

CONCLUSION

In sum, the Stolen Valor Act of 2005 is unconstitutional under the First Amendment because it is a content-based restriction on speech that fails to satisfy strict scrutiny. Under First Amendment principles, however, the type of speech the Stolen Valor Act seeks to prevent is not the type of speech the framers sought to protect. Due to the history of protecting military valor in the United States and the increased number of false representations of military service, legislators must take another path to deter this type of behavior. While Congress has begun an attempt at revising the Act, the proposed legislation is still unlikely to pass constitutional muster. Thus, Congress should instead restructure the Act to resemble a fraud statute. In addition, state and local governments should hold public officials accountable for such misrepresentations. An effective method of doing so would be to impose an eligibility requirement that prohibits, for individuals seeking an elected position, false claims of military valor and service. Overall, false claims of military valor are worth policing in some way. Legislatures should revitalize their original initiative behind the Stolen Valor Act of 2005 and invoke the aforementioned constitutional alternatives.

Stephanie L. Gal

† B.A., Government and Psychology, Georgetown University, 2009; J.D. Candidate, Brooklyn Law School, 2012. I wish to thank Professor Nelson Tebbe for his supportive guidance, expertise, and assistance through the note writing process, and Professor Aliza Kaplan for her encouragement that gave me the necessary confidence in my work. I would also like to thank the Brooklyn Law Review staff for their help with the editing process, especially Shawna MacLeod, T. Daris Isbell, Philip Weiss, and Christopher Sevier. Finally, a special note of thanks to my family and friends for their endless love and support throughout my time in law school.
Cutting the Baby in Half

AN ECONOMIC CRITIQUE OF INDIVISIBLE RESOURCE PARTITION

INTRODUCTION

The laws of partition\(^1\) are inadequate to solve most modern co-ownership disputes fairly and efficiently because there are only two generally recognized judicial methods of partition: partition in kind and partition by sale.\(^2\) With partition in kind, the resource at issue is physically divided;\(^3\) with partition by sale, the resource is sold and the proceeds are distributed to the parties.\(^4\) Partition thus presumes that all property can either be physically divided or sold to effectuate a fair distribution.

However, one vexing partition problem courts and scholars have struggled to solve, though have generally avoided, is how to partition property that can neither be sold nor divided physically.\(^5\) Take for instance a situation where two

\(^1\) 59A AM. JUR. 2d Partition § 1 (2003) (defining partition as “the dividing of lands held by joint tenants, coparceners, or tenants in common, into distinct portions, so that they may hold them in severalty” or “any division of real or personal property between co-owners, resulting in individual ownership”).

\(^2\) See id. § 2 ("Partition takes two forms: (1) partition in kind . . . and (2) a partition by sale . . . ."); 68 C.J.S. Partition § 1 (2009) (“Partition’ is a division . . . of . . . property . . . effected by the setting apart of [joint] interests so that [the owners] can enjoy and possess it in severalty or by a sale of the whole and the awarding to each of his or her share of the proceeds.”); WILLIAM B. STOEBUCK & DALE A. WHITMAN, THE LAW OF PROPERTY § 5.13 (3d ed. 2000) (recognizing "physical division or sale and division of proceeds" as the two available methods); Abraham Bell & Gideon Parchomovsky, A Theory of Property, 90 CORNELL L. REV. 531, 600 (2005) [hereinafter Bell & Parchomovsky, Theory of Property] (“Co-tenancies may be partitioned in two ways: either by sale or in kind.”). An overwhelming majority of sources limits the two available methods of partition to partition in kind and partition by sale. Among the main aims of this article is to show that courts in many jurisdictions also engage in a third method of partition: partition by allotment.

\(^3\) See 59A AM. JUR. 2d Partition § 2 (2003) (noting that partition in kind is sometimes known as “actual partition”).

\(^4\) See id. (noting that partition by sale is known as “partition by licitation” in Louisiana).

\(^5\) See DUKEMINIER ET AL., PROPERTY 300 (6th ed. 2006) (posing the question and citing In re McDowell, 345 N.Y.S.2d 828 (Sur. Ct. 1973)); see also Pugh v. NPC Servs., Inc., 721 So. 2d 1056, 1058 (La. Ct. App. 1998) (where, because of hazardous waste contamination, the property either had “no value or a negative value,” the court
siblings are fighting over ownership of their late grandfather’s old rocking chair. Using only partition in kind and partition by sale, a court cannot adequately resolve the matter. This struggle to partition indivisible resources underscores one manifest limitation of partition. More importantly, however, it reveals a fundamental problem with the general approach of partition law: the disregard of subjective value.

In property law, courts generally apply the so-called “property rule”: an entitlement can only be removed if the holder of the entitlement voluntarily sells it (for a price determined by the holder). A property rule thereby protects the value that an owner subjectively attaches to his property. Conversely in partition, owners of concurrent interests are often forced to relinquish their entitlement in exchange for the entitlement’s objective value. In using this “liability rule” courts assume that (as in eminent domain) owners cannot set the price they will be paid when they are forced to sell their property.

By using economic principles of game theory, allocations in partition can be fairly and efficiently determined from the owners’ subjective values, rather than from the objective view of the market. This note argues that when fair and efficient to do so, courts in partition should protect each co-owner’s entitlement by using neither a property rule nor a liability rule, but rather a hybrid rule, whereby the court would require a co-owner to sell his interest at a price determined subjectively by the co-owners themselves. So, in two-party indivisible resource partitions,

asked “[h]ow can property that is susceptible to neither division in kind nor judicial sale be partitioned?”.

See generally In re McDowell, 345 N.Y.S.2d 828 (presenting substantially similar facts).

See id. at 830; see also Pugh, 721 So. 2d at 1058.


See Calabresi & Melamed, supra note 8, at 1108.

See Lucas J. Asper, The Fair Market Value Method of Property Valuation in Eminent Domain: “Just Compensation” or Just Barely Compensating?, 58 S.C. L. REV. 489 (2007). In this article, Asper suggests that owners’ subjective evaluation of their property should be considered in determining “just compensation” in eminent domain cases, instead of simply paying owners the fair market value of their property. To do so, Asper suggests “courts should utilize a hybrid rule that makes use of the characteristics of both property rules and liability rules.” Id. at 502. However, this “hybrid rule” requires courts to balance “all of the factors affecting the property value, including subjective values such as sentimental attachment,” rather than have the owners themselves set the value of their entitlement. Id.
courts should eschew both partition in kind and partition by sale, and instead allot the entire property to the co-owner who values it the most, paying to the other co-owner an amount determined by the parties’ subjective evaluation of the property.

This note analyzes established modes of partition and criticizes them for failing to conform to normative economic criteria. Following the introduction, Part I will track the development of the modern rules of partition in a historical context. Part II will outline the normative criteria and economic theories necessary to analyze different partition methods. Part III will apply these criteria to various methods of partition, including popular methods such as chance and rotation, as well as the judicial methods of in kind, sale, and allotment. Part IV suggests a new method for partitioning indivisible resources, referred to here as equitable allotment. Finally, this note concludes by discussing unresolved problems and encouraging further debate.

I. HISTORY AND DEVELOPMENT OF PARTITION LAW

Partition is a classic—problem appearing in many contexts—from disputes as mundane as two children fighting over how to divide a cupcake, to divorces, to corporate bankruptcies.

Perhaps the most celebrated story of partition is the biblical tale of King Solomon decreeing that a baby—claimed

---

12 See Abraham Bell & Gideon Parchomovsky, Reconfiguring Property in Three Dimensions, 75 U. CHI. L. REV. 1015, 1021 (2008) [hereinafter Bell & Parchomovsky, Reconfiguring Property] (noting that “one of the most basic problems of property law” is “what to do when owners of property in common decide to part ways”).
13 See Hervé Moulin, Fair Division and Collective Welfare 235 (2003) (describing the applicability of fair division to cake-cutting, bankruptcy, divorce, inheritance, and dividing disputed territory between countries); see also Michael J. Meurer, Fair Division, 47 BUFF. L. REV. 937, 937 & nn.1-6 (1999) (book review) (describing the importance of fair division to remedies, probate, family law, partnership law, bankruptcy, and other fields).
14 The cake-cutting problem may seem trivial, but it has occupied mathematicians and economists for well over half a century. See, e.g., Lee Anne Fennell, Revealing Options, 118 HARV. L. REV. 1399, 1401-02 (2005); Hugo Steinhaus, The Problem of Fair Division, 16 ECONOMETRICA 101 (1948).
15 See, e.g., Jeremy A. Matz, Note, We’re All Winners: Game Theory, the Adjusted Winner Procedure and Property Division at Divorce, 66 BROOK. L. REV. 1339 (2001) (criticizing the applicability of a game-theoretic property division method to divorce cases).
16 See MOULIN, supra note 13, at 235.
by two women each as her son—be cut in half. Solomon, of course, did not actually cut the baby in half but instead gave him to the woman who had protested the partitioning, declaring that she was the boy’s true mother, rather than the other woman who seemed satisfied that physically dividing the baby was a just method. Solomon’s judgment, however, is not truly a partitioning scheme because Solomon did not intend to divide the baby, but sought instead to discern the parties’ subjective evaluation of the resource at issue and allocate the resource to the party that valued it the most.

Ancient history contains several other treatments of partition. The Greek myth of Zeus and Prometheus sharing meat describes perhaps the oldest allusion to the divide-and-choose method. In that story, Prometheus separated the meat into two portions, and Zeus selected his share. The Talmud (the paramount text of Jewish rabbinic law) discusses a few partitioning cases, beginning with the famous garment problem. In this case, where two men grab on to a garment that each claims as exclusively his, the Talmud decrees that each be given half. Roman law allowed for physical division and payment


18 See 1 Kings 3:25.
19 See 1 Kings 3:27.
20 See BRAMS & TAYLOR, WIN-WIN SOLUTION, supra note 17, at 8.
21 See BRAMS & TAYLOR, FAIR DIVISION, supra note 17, at 6 (explaining that Solomon’s solution was really a game to discern the women’s preferences and strategies).
22 Thus achieving a more efficient distribution. See infra notes 100-01 and accompanying text.
23 See BRAMS & TAYLOR, FAIR DIVISION, supra note 17, at 10.
24 See BABYLONIAN TALMUD, BABA MEZIA 2a (I. Epstein ed., Salis Daiches & H. Freedman trans., 1986). Besides the garment problem, the Talmud goes on to discuss a similar situation: where “one rides [on an animal] and the other leads it” both claiming it as all theirs, each will get half. In a variation, one person holding the garment claims “it is all mine” but the other claims “half of it is mine.” Here the Talmud decrees the first gets three-quarters, and the other one quarter, on the theory that only half of the garment was in dispute, and so only that portion was divided. See H. PEYTON YOUNG, EQUITY IN THEORY AND PRACTICE 65 (1994) (discussing this Talmudic rule in the context of equality and proportionality). For other Jewish laws of partition, see 5 EMMANUEL QUINT, A RESTATEMENT OF RABBINIC CIVIL LAW, Part IV (Laws of Partition of Realty) (1994).
25 See Baba Mezia, supra note 24, at 2a (“Two hold a garment. One of them says, ‘I found it,’ and the other says, ‘I found it’; one of them says, ‘it is all mine,’ and the other says, ‘it is all mine.’”).
of an owelty. But, if the property was not capable of division, the Corpus Juris Civilis of Emperor Justinian allowed allotment of the entire property to one of the parties. English laws of partition can be traced at least as far back as Roman law.

Under the early common law, physical partition was greatly limited, and partition by "sale was out of [the] question." In medieval England, laws of primogeniture left inheritance to the eldest son, and thus made the need for partition infrequent. However, the common law limitation had one major exception: where all of a decedent’s heirs were female, the law required partition. The law referred to these heiresses as coparceners, precisely because partition was required of them.

Blackstone briefly discusses partition among coparceners and notes that partition occurred either by consent or by compulsion. Consensual partition included partition in kind, where sisters chose parcels in order of seniority, or by random chance. Blackstone also makes an early reference to the divide-and-choose method, allowing the eldest sister to divide the land and choose the last parcel. Compulsory partition, notes Blackstone, was had under a writ of partition, whereby the sheriff would make a partition of the land on the verdict of a jury and assign the parcels to the respective parties.

Blackstone acknowledges that “there are some things which are in their nature impartible,” and so, in situations involving a “mansion-house” for instance, Blackstone forbids physical partition. As an alternative to physical partition, Blackstone writes that in these situations “the eldest sister, if she pleases, shall have [the property], and make the others a reasonable satisfaction.” However, if that is not possible, Blackstone holds

---

28 See Dig. 10.2.55 (Ulpian, Ad Edictum 2) (S.P. Scott trans., 1932); see also Loyd, supra note 27, at 163 (referencing the same).
29 See Loyd, supra note 27, at 163.
30 See id. at 166.
31 Id. at 167.
32 See id. at 164.
33 See id.
34 See id.
35 See 2 WILLIAM BLACKSTONE, COMMENTARIES *189.
36 See id.
37 See id. (“[W]here the eldest divides, and then she shall chose last; for the rule of law is, cujus est divisio, alterius est election. [S]he who makes the division has the last choice.”) (English translation in original).
38 See id.; see also Loyd, supra note 27, at 167 (“[T]he sheriff with a jury of twelve went upon the land, made a division of it and allotted the shares.”).
39 2 WILLIAM BLACKSTONE, COMMENTARIES *189-90.
40 Id.
that the sisters “shall have the profits of the thing by turns.” This
taking turns—or rotation—was also noted by Sir Edward Coke,
who wrote that if a mill were to be partitioned, coparceners would
use it for alternating periods of time.\footnote{Id.}

With the statute of Henry VIII,\footnote{See Loyd, supra note 27, at 167.} the right to physical
partition was greatly expanded. First, partition became
available to tenants in common and joint tenants, rather than
just coparceners.\footnote{See id. at 168-69; 2 AMERICAN LAW OF PROPERTY § 6.21 (1952); STOEKBURK & WHITMAN, supra note 2, § 5.11.} Second, the right to partition became
considered absolute,\footnote{See 2 AMERICAN LAW OF PROPERTY, supra note 43, § 6.21. Apparently,
castles “necessary for defence of the realm” were exempted on policy grounds. Turner v.
Morgan, (1803) 32 Eng. Rep. 307 (Ch.) 308; 8 Ves. Jun. 143, 144; see also Loyd, supra note 27, at 167 (noting that unlike halls, which were sometimes physically divided,
castles used for defense were to remain with the oldest son, his brothers receiving
payment for their shares).} and thus, a “court could not refuse it or
order a sale, although the result might be inconvenient or even
absurd.”\footnote{Loyd, supra note 27, at 173.} For instance, in Turner v. Morgan, the English Court
of Chancery held that, in a partition of a house, it was not error to “allot[] to the Plaintiff the whole stack of chimneys, all the
fire-places, the only staircase in the house, and all the
conveniences in the yard.”\footnote{See Turner, 32 Eng. Rep. at 308; Crocker v. Cotting, 48 N.E. 1023, 1024
(Mass. 1898) (Holmes, J.) (citing Turner, 32 Eng. Rep. 307); Heldt v. Heldt, 193 N.E.2d 7, 9 (Ill. 1963) ("Motive for partition is immaterial . . . and . . . the absolute right to
partition yields to no consideration of hardship, inconvenience or difficulty."); cf. ARK.
CODE ANN. § 18-60-404 (2003) (providing that minority shareholder party not related to
other co-owners by four degrees of consanguinity cannot petition for partition within
three years of the purchase of that interest).}

In response to increased urbanization and the occasionally ridiculous results of compulsory physical partition\footnote{32 Eng. Rep. at 308; see also Scovil v. Kennedy, 14 Conn. 349, 360-61
(1841) (exhibiting the difficulty in partitioning a stream of water).} (as in Turner v. Morgan), American states enacted
legislation allowing courts to sell the property and divide the
Apr. 22, 1992) (opining that the Connecticut legislature may have enacted the state's
partition statute in response to Scovil, 14 Conn. 349 (partitioning a stream)).} Every state now governs partition
by statute.\footnote{See Wilk v. Wilk, 795 A.2d 1191, 1193 (Vt. 2002).} However, under these schemes, statutes in almost\footnote{Plus the District of Columbia, Guam, and the Virgin Islands. See, e.g., CAL.
CIV. PROC. CODE § 874.010 (Deering 2011); CONN. GEN. STAT. ANN. § 52-495 (West
2005); MISS. CODE ANN. § 11-21-1 (LexisNexis 2010); N.J. STAT. ANN. § 2A:56-2. (West
2000); N.Y. REAL PROP. ACTS. § 901 (McKinney 2010); S.C. CODE ANN. § 15-61-10 (West
2005); VA. CODE ANN. § 8.01-81 (2007).}
every jurisdiction still anachronistically\(^\text{52}\) favor partition in
kind,\(^\text{53}\) and allow sale only if physical partition would result in
“great prejudice”\(^\text{54}\) or “manifest injury”\(^\text{55}\) to a party in interest.\(^\text{56}\)
Nevertheless, evidence shows that courts’ loyalty to the
statutes is only nominal,\(^\text{57}\) and judges are now more likely to
order partition by sale despite the stated preference.\(^\text{58}\)
Aside from partition in kind and by sale, courts in at least fifteen
states also participate in a third generally unacknowledged\(^\text{59}\)
method of judicial partition called partition by allotment,\(^\text{60}\)

---

\(^{51}\) See IOWA CODE ANN. § 1.1201(2) (LexisNexis 2010) (“Property shall be
partitioned by sale . . . unless a party prays for partition in kind . . . .”); Fannin v.
Fannin, 75 S.W.2d 1042, 1043 (Ky. 1934) (“[T]he court will presume . . . that a town
lot is not susceptible of advantageous division.”). Maine statutes do not provide
However, Maine courts will allow partition by sale under the court’s equity

\(^{52}\) As early as 1919, William H. Loyd wrote, “American statutes permitting a
sale when great prejudice would result from a division were passed at a time when
interests were mainly agricultural and neither legislator nor judge would willingly
disturb the traditional attitude toward ownership.” Loyd, supra note 27, at 189; see
also Manel Baucells & Steven A. Lippman, Justice Delayed Is Justice Denied: A
Cooperative Game Theoretic Analysis of Hold-Up in Co-Ownership, 22 CARDOZO L.
REV. 1191, 1195 (2001) (“The law ostensibly favors partition in kind for historical
reasons . . . .”).

\(^{53}\) See Loyd, supra note 27, at 188; Delfino v. Vealencis, 436 A.2d 27, 30
(Conn. 1980); Thomas J. Miceli and C. F. Sirmans, Partition of Real Estate; Or,
Breaking Up is (Not) Hard to Do, 29 J. LEGAL STUD. 183, 787 (2000).


\(^{55}\) The statutory standard varies by jurisdiction. See 4 THOMPSON ON REAL
PROPERTY, supra note 53, § 38.04 & nn.108-09; see also John G. Casagrande Jr., Note,
Acquiring Property Through Forced Partition Sales: Abuses and Remedies, 27 B.C. L.
REV. 755, 760 n.47 (1986) (listing each state’s statutory conditions for allowing sale).

\(^{56}\) See 4 THOMPSON ON REAL PROPERTY, supra note 5453, § 38.04; UNIF.
PARTITION OF HEIRS PROP. ACT prefatory note (2010), available at
UPHPA] (“Despite the overwhelming statutory preference for partition in kind, courts in a
large number of states typically resolve partition actions by ordering partition by sale . . . .”);
see also Bell & Parchomovsky, Theory of Property, supra note 2, at 601 (noting that
despite the “[r]hetorical] . . . preference for partition in kind . . . courts have often
favored tilting the balance toward partition by sale”); Candice Reid, Note, Partitions in

\(^{57}\) See 2 AMERICAN LAW OF PROPERTY, supra note 43, § 6.26 (writing in 1952
that “[a]s a practical matter the modern practice is to decree a sale in partition actions
in the great majority of cases, which usually involve single parcels of improved
property”); Phyliss Craig-Taylor, Through a Colored Looking Glass: A View of Judicial
(“[I]n practice . . . . partition sales are the rule rather than the exception.”).

\(^{58}\) See infra notes 205-07 and accompanying text.

\(^{59}\) See infra Part III.E.
where the entire property is allotted to one party who pays the others an objective amount for their shares.\textsuperscript{61}

Courts still maintain that concurrent owners have an “absolute right to partition.”\textsuperscript{62} Therefore, as with eminent domain, the law may force an owner to unwillingly sell his property. Indeed, the right is so strong that courts will rarely deny partition even when partition would be unduly oppressive to a party\textsuperscript{63} or would displace a family with minor children.\textsuperscript{64} One author noted that although partition actions are brought in equity, the courts have been strangely reluctant to deny partition on the basis of [equitable] defenses. Attempts to defeat partition by claims of hardship have almost always been unsuccessful and frequently have elicited statements by the court that the right to partition is absolute.\textsuperscript{65}

In 2010, the National Conference of Commissioners of State Laws promulgated the Uniform Partition of Heirs Property Act (UPHPA)\textsuperscript{66} in order to address many of the problems associated with forced sale under partition statutes. Under the UPHPA, the cotenants who did not request partition by sale have the opportunity to buy the shares of those cotenants who did request partition\textsuperscript{67} at a price based on an initial appraisal value of the property.\textsuperscript{68} If the shares of the co-owners who requested partition by sale are not all bought-out, the UPHPA directs the court to order partition in kind, unless doing so would prejudice the cotenants.\textsuperscript{69} If the court orders partition by sale, the UPHPA requires that a real estate broker (paid a “reasonable commission”) conduct an open-market sale for a

\textsuperscript{61} See generally infra Part III.E. Variations on partition by allotment may avoid its characterization as something separate from partition by sale or in kind. For instance, in an in kind division with an owelty payment, it is conceivable that the property be physically divided into one part containing the entire property and the other containing nothing. Furthermore, the winning bidder at a partition sale may be one of the co-owners. The sale may also be conducted privately.

\textsuperscript{62} Heldt v. Heldt, 193 N.E.2d 7, 9 (Ill. 1963) (explaining that “motive for partition is immaterial and that the absolute right to partition yields to no consideration of hardship, inconvenience or difficulty” (citations omitted)).

\textsuperscript{63} See 4 THOMPSON ON REAL PROPERTY, supra note 54, § 38.03(a)(2)(iii) (citing Hassel v. Workman, 260 P.2d 1081 (Okla. 1953) and Condrey v. Condrey, 92 So. 2d 423 (Fla. 1957) as rare examples of the exception).

\textsuperscript{64} See Heldt, 193 N.E.2d at 9; see also 4 THOMPSON ON REAL PROPERTY, supra note 54, § 38.03(a)(1) (citing Heldt).

\textsuperscript{65} Leonard A. Girard, Equitable and Contractual Defenses to Partition, 18 STAN. L. REV. 1428, 1433 (1966).

\textsuperscript{66} UPHPA, supra note 57.

\textsuperscript{67} See id. § 7 & accompanying cmts.

\textsuperscript{68} See id. § 6.

\textsuperscript{69} See id. § 8.
value not less than the original appraisal value.\textsuperscript{70} The UPHPA also allows the property to be sold at auction or by sealed bids\textsuperscript{71} and allows one of the cotenants to be the winning purchaser.\textsuperscript{72} The UPHPA is limited only to certain types of tenancy in common property,\textsuperscript{73} and has only been adopted by one state.\textsuperscript{74}

Courts have occasionally transferred principles of partition to other situations. For instance, in 2001, a dispute arose between two baseball fans over Barry Bonds’s record-setting homerun ball,\textsuperscript{75} when, after Bonds hit the ball into the stands, each litigant claimed he was the ball’s true owner.\textsuperscript{76} In \textit{Popov v. Hayashi}, the court decided that both parties had “an equal and undivided interest in the ball.”\textsuperscript{77} By recognizing each litigant’s concurrent property interest, the court, in effect, deemed the two litigants co-owners\textsuperscript{78} of the ball, before ordering the ball sold and the proceeds divided evenly between them.\textsuperscript{79}

II. NORMATIVE CRITERIA FOR AN ECONOMIC ANALYSIS OF PARTITION\textsuperscript{80}

This section establishes the tools necessary for an economic analysis of partition. However, it is difficult—if not impossible—for any popular or judicial scheme to conform to all
of these principles, and sometimes the presence of one may preclude the presence of another. When scrutinizing a particular method of partition, it is not enough that the scheme sometimes results in a fair or efficient allocation. Rather, the application of these criteria questions whether the mechanism guarantees a particular result. Moreover, the qualities outlined in this checklist may carry different persuasive weight among concurrent owners, courts, economists, and society at large. As a whole, the focus should be both on the fairness of the outcome as well as the method itself.

A. Proportionality

Proportionality embodies the essence of fair division. An allocation is proportional for \( n \) parties with equal interests if each party feels it received at least \( 1/n \) of the resource.\(^{81}\) Therefore, dividing a pie between two children is proportional if each child feels she received one-half of the pie. However, if Alice contributes $10 and Ben contributes $5 towards the purchase of a $15 pie, fairness dictates that Alice receive at least twice as much pie as Ben.\(^{82}\) This conforms to the Aristotelian doctrine that "[e]quals should be treated equally, and unequals unequally, in proportion to relevant similarities and differences."\(^{83}\) This seems intuitive, but when faced with difficult cases, resorting to arbitrary factors—such as age or gender—may be easier, but less fair, than proportional division.

B. Envy-Freeness

Envy-freeness conceptualizes the emotional underpinnings of unfairness into a rational\(^{84}\) and compelling test.\(^{85}\) According to one author, "A distribution is said to be envy-free if no one prefers another's portion to his own."\(^{86}\) After

\(^{81}\) See Brams & Taylor, Fair Division, supra note 17, at 9 & n.8.
\(^{82}\) Presuming that twice as much pie is worth twice as much.
\(^{83}\) Moulin, supra note 13, at 1 (citing Aristotle's Nicomachean Ethics).
\(^{84}\) The underlying assumption is that the parties each rationally want to gain as much value as possible from the allocation. See Lewinsohn-Zamir, supra note 9, at 228 ("The standard economic game-theoretic prediction is that both players will behave rationally, that is to say, strive to maximize their monetary payoffs.").
\(^{85}\) See Moulin, supra note 13, at 235 ("With heterogeneous individual preferences . . . the test of no envy . . . offers an extremely appealing answer.").
\(^{86}\) Young, supra note 24, at 11 (emphasis omitted); accord Moulin, supra note 13, at 236 ("A distribution of resources is nonenvious, or it passes the no-envy test, if every agent prefers (weakly) his or her share to that of any other agent: I cannot complain about my share because no one else has a share that I would exchange for
an envy-free distribution, each party feels he received the most valuable portion,\textsuperscript{87} or a portion at least as valuable as anyone else’s portion. On the other hand, if after partitioning the resource, one party values another’s share more highly than his own, the partition fails the envy-free test.\textsuperscript{88} Thus, the envy-free test requires consideration of each party’s subjective valuation.

C. Efficiency

Concurrent ownership generally encourages inefficient use of property.\textsuperscript{89} If too many people hold the right to use the property, each user may not be able to internalize the consequences (externalities) of his use, which produces inefficient consumption—a situation called the “tragedy of the commons.”\textsuperscript{90} On the other hand, when there are too many owners, each with the ability to block the others’ use, the result is likewise inefficient because by threatening a block, each owner may extract a premium\textsuperscript{91} from the other owners—a
situation known as the “tragedy of the anticommons.”\textsuperscript{92} By ending co-ownership, partition encourages efficiency; however, the methods of partition themselves are often inefficient.

Efficiency has various definitions,\textsuperscript{93} but this discussion will use “efficient” to mean “Pareto-optimal.” According to one leading economist, “Pareto optimality is the single most important tool of normative economic analysis. Its desirability is undisputed. In the endstate version of distributive justice, it is the one requirement that cannot be dispensed with.”\textsuperscript{94} An allocation is Pareto-optimal if there exists no other feasible allocation that can make one party better-off while keeping all other parties at least as well-off.\textsuperscript{95} In other words, a situation is inefficient if it is possible to make a party better without making anyone else worse.\textsuperscript{96} So if Harry has three gloves but Lloyd only has one glove, the distribution is inefficient because taking away Harry’s extra glove and giving it to Lloyd keeps Lloyd warm without making Harry cold. On the other hand, if Harry has two gloves and Lloyd has only one, the distribution is efficient because the only way to make Lloyd better is to make Harry worse.

After an envy-free distribution, simply because each party is unwilling to switch shares with anyone else does not mean that everyone received the largest share possible.\textsuperscript{97} In other words, an envy-free allocation is not necessarily efficient.\textsuperscript{98} However, as one economist wrote, “A nonenvious distribution that is also efficient is one coherent answer to the fair division puzzle.”\textsuperscript{99}

Fitness is an aspect of efficiency that mandates giving “the resources to whomever makes the best use of it: the flute

\textsuperscript{92} See Bell & Parchomovsky, Theory of Property, supra note 2, at 564; see also Miceli & Sirmans, supra note 53, at 783, 787 (but noting, also, that “[c]ommon ownership can promote efficient use of land”).
\textsuperscript{93} See ZAJAC, supra note 88, at 11-12.
\textsuperscript{94} MOULIN, supra note 13, at 8. But see ZAJAC, supra note 88, at 69 (“[E]conomic efficiency is a necessary but far from sufficient condition for an economically just or fair economy.”).
\textsuperscript{95} See B. Lockwood, Pareto Efficiency, in 3 The New Palgrave Dictionary of Economics 811, 811 (Eatwell et al. eds., 1987). Note that in considering a Pareto-superior position we look only at the allocation to the parties involved. So, an allocation that leaves all parties better off, but dispenses with the need for attorneys, is considered more efficient even though the lawyer counting on a big paycheck is now unhappy. See infra Part II.D.
\textsuperscript{96} See BRAMS & TAYLOR, FAIR DIVISION, supra note 17, at 44.
\textsuperscript{97} See id. at 2.
\textsuperscript{98} See id. at 2 n.2.
\textsuperscript{99} MOULIN, supra note 13, at 236.
to the flutist, [and] the books to the avid reader."\textsuperscript{100} The utility a party derives from an allocation is discernable from how much that party values the allocation. Therefore, a child who likes pie will value a slice of pie more than he will value a torts casebook, because he can derive greater utility from the pie than from the casebook. Therefore, efficiency dictates allocating a resource to whichever party values it most.\textsuperscript{101}

\textbf{D. Administrability}

While not conceptually distinct from Pareto-superiority,\textsuperscript{102} administrability emphasizes efficiency in method as well as result. It is incontrovertible that partition schemes should be easy to administer, costing the courts and the parties as little time and effort as possible. Ideally, the parties themselves should be able to achieve the partition cheaply and easily without the need of a third-party mediator (such as the courts).\textsuperscript{103} However, in every jurisdiction,\textsuperscript{104} partition requires peripheral entities—such as auctioneers,\textsuperscript{105} appraisers,\textsuperscript{106} panels of commissioners or referees,\textsuperscript{107}

\textsuperscript{100} Id. at 2. Gauging utility often involves social judgments. Do farmers make "better use" of a field than do football players? Is it better to give the flute to the flutist if no one likes hearing her play? This note will ignore externalities and assume that the party who values the resource the most derives the greatest utility from it. Eschewing social judgment may prove unworkable if society prefers one use over another more profitable use, as in a situation where a young girl’s cherished pet horse may be another’s prizewinner. In this note, utility will be purely subjective.

\textsuperscript{101} See Fennell, supra note 14, at 1403 (“From an efficiency perspective, we would want the entitlement to end up in the hands of the party who values it the most—whether or not the entitlement was originally assigned to that party.”); James E. Krier & Stewart J. Schwab, Property Rules and Liability Rules: The Cathedral in Another Light, 70 N.Y.U. L. REV. 440, 446 (1995) (“From the standpoint of efficiency, a judge should . . . assign the entitlement . . . such that it \textit{ends up} in the hands of that party . . . who values it most (or can do without it at least cost). From the standpoint of justice, the judge should assign the entitlement such that it \textit{starts out} in the hands of the party who is most deserving in light of the applicable justice norm . . . .”).

\textsuperscript{102} See Calabresi & Melamed, supra note 8, at 1093-95 ("[A]dministrative efficiency is just one aspect of the broader concept of economic efficiency."). Pareto-superior means more efficient, or closer to Pareto-optimal.

\textsuperscript{103} This discussion only contemplates the benefits to the parties, so any loss in revenue for auctioneers, appraisers, commissioners, and lawyers is not considered.

\textsuperscript{104} See, e.g., CAL. CIV. PROC. CODE § 874.010 (West 1980) ("The costs of partition include [inter alia]: . . . attorney's fees . . . fee and expenses of the referee . . . compensation . . . for services of a surveyor or other person employed by the referee . . . [and] costs of a title report . . . ").


\textsuperscript{107} See Girard, supra note 65, at 1429-30.
surveyors, and of course, attorneys—whose costs all eat away at the parties’ final allocations. A method of partition that eliminates these costs will result in Pareto-superiority because each party will receive a greater allocation without the deduction of these extra expenses.

E. Equitability

Even in situations where all the parties feel they received more than their fair share, the amount by which they feel they received more than their fair share may differ. Equitability ensures that each party’s surplus (the value exceeding the proportional share) is equal. So if Alice feels she got 51 percent of the pie, but Ben feels he got 91 percent of the pie, the allocation is not equitable. For a more concrete example, imagine that Alice loves boats but hates airplanes, and Ben loves airplanes but hates boats. If Alice is given a toy boat, but Ben is given an actual Harrier jet, the allocation will be envy-free because neither party will want to switch places with the other. However, although Alice is happy with her boat, Ben is much happier with his jet, and therefore the allocation is not equitable. A fair method of partition ensures equitability.

F. Strategy-Proofness

Any method susceptible to gamesmanship does not guarantee fair results. Therefore, a fair partition method must ensure that it is not in any player’s best interest to lie. If an allocation method is strategy-proof then “truthful report is a dominant strategy.” This occurs if the party making the

---

109 See id.
110 See discussion under “Efficiency,” supra Part II.C.
111 Presuming the jet cannot be sold to buy many toy boats.
112 See MOULIN, supra note 13, at 237 (noting the importance of “strategyproofness” in nonenvious assignments); see also Fennell, supra note 14, at 1401-02 (noting in a game-theoretic discussion of Calabresi-Melamed entitlements that the “challenge, then, is to structure legal entitlements in a way that induces people to truthfully reveal their valuations”); Ayres & Talley, supra note 89, at 1030 (noting that “self-interested bargainers have a strong incentive to misrepresent their private valuations so as to capture a larger share of the bargaining ‘pie’” and developing a scheme to facilitate trade by eliminating this incentive).
113 See MOULIN, supra note 13, at 237 (noting that when honesty was the dominant strategy, the assignment was strategy-proof).
114 Vito Fragnelli & Maria Erminia Marina, Strategic Manipulations and Collusions in Knaster Procedure, 3 AUO CZECH ECON. REV. 143, 144 (2009).
evaluation either knows that lying will be disadvantageous\footnote{Consider a rule that allows homeowners to set the value of their property for tax purposes and for eminent domain purposes. Clearly, for property tax purposes, owners would gain by setting a price much lower than their true valuation of the property. Similarly, if homeowners were allowed to set the “just compensation” value of their home after it was taken through eminent domain, owners would quote a price much higher than their true valuation of the property. If strategy-proof, it would be in the owner’s best interest to tell the truth about the property’s value.} or does not know whether fixing a high or low evaluation will be to her advantage.\footnote{See Fennell, supra note 14, at 1411, 1418-19 (citing John Rawls’s “veil of ignorance” as an antidote to self-interest bargaining).} Therefore, a strategy-proof method disincentivizes both high and low valuations.\footnote{See id. at 1411.} Consonant with the criterion of strategy-proofness is the requirement that the method be impervious to collusion. In other words, it should also not be possible for two or more players to better their positions by furtively agreeing to collaborate.\footnote{See Fragnelli & Marina, supra note 114, at 144 (quoting Moulin’s definition of coalition-strategy-proof as “when ‘if a joint misreport by a coalition strictly benefits one member of the coalition, it must strictly hurt at least one (other) member’”).} Much of game theory fair division is devoted to concocting strategy-proofing mechanisms.

III. Testing Partition Schemes for Fairness and Efficiency

This section analyzes several methods of partition using the criteria established in Part II. Although these are not nearly the only modes of partition, they represent an array of schemes most likely to impact the rights of co-owners.\footnote{The several modes of partition discussed primarily by economists, which have gained little if any attention by the courts or legal academics, are best handled in a separate treatment. See Matz, supra note 15.}

A. Partition in Kind

It is axiomatic that one cannot physically divide an indivisible resource. Although most modern partition actions involve indivisible property, statutes in nearly\footnote{See IOWA CODE ANN. § 1.1201(2) (West 2002) (“Property shall be partitioned by sale . . . unless a party prays for partition in kind . . . .”); Fannin v. Fannin, 75 S.W.2d 1042, 1043 (Ky. 1934) (“[T]he court will presume . . . that a town lot is not susceptible of advantageous division.”).} every jurisdiction prefer partition in kind.\footnote{See Loyd, supra note 27, at 188; Delfino v. Vealencis, 436 A.2d 27 (Conn. 1980); Miceli & Sirmans, supra note 53, at 784.} In practice however, courts are likely to allow sale “where the . . . fragmentation [resulting from physical division] would materially reduce the
aggregate value of the [property]." Thus, "indivisible" does not mean "incapable of division," but rather that partition in kind would reduce the total value of the resource. A six-karat diamond that can be cut into two three-karat diamonds is therefore considered indivisible because as the karat of a single diamond increases its price per karat also increases. Moreover, because of removal and disposal costs, physically partitioning a boat, for example, may leave the parties even worse off than if they had received nothing at all. Even if the aggregate value of the divided parcels were not reduced, but remained exactly equal to the value of the undivided property, the high transaction costs associated with partition means that a partition in kind scheme would result in a Pareto-inferior distribution.

Despite the imprudence of physical division, courts presented with an indivisible resource may order partition in kind to encourage an out-of-court settlement. For instance, the court in Turner v. Morgan sanctioned physically partitioning a house hoping the parties would "agree to buy and sell" each other's shares. Yet such a ruling results in an anticommons situation where each party may block the other's use and extract hold-up costs in excess of their fair share. Thus in a world with transaction costs, courts that use a property rule by ordering partition in kind encourage inefficiency.

On the other hand, "where there are no scale economies [meaning that] the aggregate market value of the [property] when subdivided . . . is no less than the value of the undivided parcel . . . partition in kind is always the efficient remedy." So

---

122 Miceli & Sirmans, supra note 53, at 796. Here the authors provide a balancing test for determining when courts should order sale and when they should partition in kind, considering the goals of maximizing the aggregate value of the land and protecting subjective value. Id. at 783; see also Bell & Parchomovsky, Theory of Property, supra note 2, at 601 ("[C]ourts have often favored tilting the balance toward partition by sale . . . by collapsing the two-part test for partition in kind into a one-step inquiry into whether value would be lost by opting for partition in kind.").
123 See Bell & Parchomovsky, Theory of Property, supra note 2, at 601-02.
124 In other words, a six-karat diamond is worth more than twice as much as a three-karat diamond. Similarly, the value of a living baby is worth much more than twice the value of half a baby.
126 See "Administrability," supra Part II.D.
128 See Baucells & Lippman, supra note 52, at 1195.
129 See Miceli & Sirmans, supra note 53, at 787-88.
130 Id. at 788-89. Except when transaction costs are high. See supra Part II.D.
for instance, a giant bluefin tuna (which can top $300,000\textsuperscript{131}) may be physically partitioned. Because such resources are easy to divide, this note will focus exclusively on indivisible property.

B. Chance

Chance has a long history in fair division.\textsuperscript{132} Some courts have accepted randomness in resource allocation,\textsuperscript{133} while others have no tolerance for arbitrariness.\textsuperscript{134} In the context of partition, after real property has been divided into parcels pursuant to a partition in kind, assignment of the parcels by lottery is acceptable\textsuperscript{135} and may even be provided for by statute.\textsuperscript{136}

Although the use of chance has many appealing qualities, it is not a fair method of partition. At first blush, a coin flip seems to provide a simple, impartial,\textsuperscript{137} and strategy-proof\textsuperscript{138} method for allocating disputed resources: it requires no mediator, and the only cost is acquiring the coin. The winner-take-all outcome of coin flipping is also efficient\textsuperscript{139} because making the loser happier requires taking value away from the winner.

Chance appears fair only when the parties view the value of the resource at issue as equivalent to the probability of obtaining the disputed property. If the undivided chance of

\begin{footnotesize}
\begin{enumerate}
\item See Jonah 1:7 (“So they cast lots, and the lot fell upon Jonah.”).
\item See Brown, supra note 86, at 65, 66-74 (discussing the history, legality, and efficacy of distribution and allocation of resources by chance); see also Adam H. Samaha, Randomization in Adjudication, 51 WM. & MARY L. REV. 1 (2009) (finding a proper place for random decision making in American jurisprudence).
\item See Wilk v. Wilk, 795 A.2d 1192, 1195 (Vt. 2002) (in dicta, explaining that where there is no basis for allotting property to one party over another, a court should resort to sale rather than abuse its discretion by making an arbitrary decision); see also Ayres & Talley, supra note 89, at 1078 (explaining that “the legal system’s use of analogy and precedent is inconsistent with a decision-making process that ultimately resembles a flip of a coin”).
\item See Townbridge v. Donner, 40 N.W.2d 655, 657 (Neb. 1950) (referee may allot partitioned portions, or parties may draw lots); Gray v. Von Crotts, 293 S.E.2d 626, 629 (N.C. Ct. App. 1982); 2 WILLIAM BLACKSTONE, COMMENTARIES *189 (coparceners distribute parcels by lots); see also Jay M. Zitter, Judicial Partition of Land by Lot or Chance, 32 A.L.R. 4th 909, 909-15 (West 1984) (collecting cases requiring or allowing parcels to be distributed by chance pursuant to a partition in kind).
\item See, e.g., MISS. CODE ANN. § 11-21-21 (West 2008).
\item See Brown, supra note 86, at 113.
\item That is, presuming the coin is fair, and both parties can see the coin being flipped and landing.
\item See Brown, supra note 86, at 113.
\end{enumerate}
\end{footnotesize}
obtaining the resource is 1, and there are \( n \) parties with equal interests, then a decision that gives each party an equal \( \frac{1}{n} \) chance of winning at first seems proportional. However, the resource at issue is not a probability, but the actual underlying property. Thus, the allocation of the property after the coin-flip is not proportional, envy-free, or equitable, because the winner of the toss gets the entire allocation, whereas the loser receives nothing.

C. Rotation

Rotation describes the concept of taking turns,\(^{140}\) examples of which are ubiquitous.\(^{141}\) Roommates take turns doing the dishes and divorced spouses take turns with custody of children.\(^{142}\) When two motorists converge at an intersection and both have the right to use the road (although they cannot both safely do so at the same time), the traffic light divides the resource temporally by giving one party the use of the road now and the other party the use of the road later.

As a judicial method of partition, rotation has gained intermittent support. Blackstone\(^{143}\) and Coke\(^{144}\) wrote that in circumstances where property, such as a house or mill, cannot be partitioned physically, the co-owners could take turns using and profiting from the resource. A Minnesota statute similarly provides: “When the premises consist of a mill or other tenement which cannot be divided . . . the referees may assign the exclusive occupancy and enjoyment . . . to each of the parties alternately for specified times, in proportion to their respective interests.”\(^{145}\)

---

\(^{140}\) See Young, supra note 24, at 21 (using the term “rotation” to describe alternating time-sharing).

\(^{141}\) For a fictional and humorous example of rotation, see The Simpsons: Three Men and a Comic Book (FOX television broadcast May 9, 1991) (Bart, Millhouse, and Martin contribute to buy an expensive comic book and agree that Bart will have possession Mondays and Thursdays, Millhouse Tuesdays and Fridays, and Martin Wednesdays and Saturdays, while Sunday’s possession will be determined by a random number generator.).

\(^{142}\) See Young, supra note 24, at 21, 34.

\(^{143}\) See 2 William Blackstone, Commentaries *189-90 (“[T]hey shall have the profits of the thing by turns . . . .”).

\(^{144}\) See Loyd, supra note 27, at 167 (alternating use of a mill).

\(^{145}\) Minn. Stat. Ann. § 558.12 (West 2010). Maine has a similar statute. 14 Me. Rev. Stat. Ann. § 6506 (West 2003) (“Tenants in common of a sawmill may have a division of the time during which each may occupy according to his interest . . . .”).
One notable case of partition by rotation is *In re McDowell.* In this probate case, a brother and sister claimed rights to their late grandfather's rocking chair, which had great sentimental value to the parties but "only nominal" value otherwise. While admitting that his ruling "may sound strange," the judge ordered that each sibling have possession of the chair for alternating six-month intervals, with the survivor to take exclusive ownership.

The court's solution in this case is analogous to a partition in kind, except here the court divided the chair temporally instead of physically. Likewise, just as if the chair had been sawed in two, the division of the chair in time substantially decreases the aggregate value of the chair. Just as twice the value of a physical half of a chair is worth less than the value of the whole chair intact, twice the value of each sibling's temporal allotment is worth less than the value of the undivided chair. In other words, the decision to divide the ownership of the chair in time leaves each sibling with much less than half the actual value of the chair. This allocation is not Pareto-optimal because other allocations exist that can leave both parties happier.

In a *McDowell* scenario, rotation adds additional inefficiency because the expense incurred in transporting the chair back and forth between the siblings' houses wastes resources that could otherwise be used more effectively. While in and out of possession, the chair cannot be put to its maximal productive use—which may include something as simple as tying a room together, or as significant as its utility as a cherished heirloom. Furthermore, the party out of possession will have to buy another chair to take the rocking chair's place.

---

147 Id. at 830.
148 Id. One author wrote of the *McDowell* decision: "The court could, I suppose, have adopted the judgment of Solomon and required the chair be cut in two. Instead the judge . . . ordered that the two children should be permitted to take the chair for six month turns . . . ." Robert Pearce, *What Kind of Castle?*, 7 DENNING L.J. 153, 164 (1992).
149 See *In re McDowell*, 345 N.Y.S.2d at 830. According to Abraham Bell and Gideon Parchomovsky, by recognizing that the siblings co-owned the chair and ordering rotation, "the court rejected the traditional owner-oriented [partition in kind] and asset oriented [partition by sale] resolutions of partition problems and instead invented one oriented toward dominion through forced time-sharing." Bell & Parchomovsky, *Reconfiguring Property*, supra note 12, at 1021 (analyzing property in the dimensions of "owner," "asset," and "dominion").
150 See supra notes 123-26 and accompanying text.
151 Consider how much you would pay to rent an apartment for a year, and compare it to how much you would pay to rent the same apartment every other day for two years.
in its absence, and incur the expense of storing the superfluous chair while back in possession. A Pareto-superior distribution would eliminate these added costs, giving each party a greater allocation. Because of these costs, the siblings will have an incentive to resolve their dispute out of court, since the sibling with less resources will either simply give up transporting the chair after a few turns, or will settle for an unfairly small amount.

For resources susceptible to waste, rotation leads to inefficiency by encouraging earlier users to maximize their own utility early lest there be nothing left when their time comes again. If two mining companies take turns with a parcel of mineral-rich land, each company will want to mine as much ore as possible—even more ore than they need—before their tenure ends. Of course, rotation does not apply to resources that are destroyed when used, such as a single piece of candy.

A central problem arises when adjudicators must decide who starts first—or more specifically—which time period to allot to which party. In the McDowell decree, the sister took the chair every July through December. If the brother preferred using the chair during autumn, he feels unhappy that he may only have the chair from January to June. If the sister dreamed of rocking on her porch in spring, she will likewise be upset at her allotment.

Even if the time allotments are substantially similar, the problem becomes who goes first and who goes second. Scholarship suggests that people overwhelmingly value obtaining a resource now more than they value obtaining the same resource at some point in the future. A person’s intertemporal choice may, in part, be psychologically

152 The McDowell Court contemplated the parties ending the arrangement voluntarily. 345 N.Y.S.2d at 830. Although not saying so, the judge may even have intended this outcome.
153 See infra notes 89-90 and accompanying text.
154 If the parties’ preferences are exactly opposite of their allotment then the distribution is clearly inefficient. However, it would be a simple matter of switching the dates to effectuate a Pareto-superior allocation. Ronald Coase argues that regardless of the initial allocation of entitlements, in the absence of transaction costs, parties will themselves bargain for an efficient distribution. See generally Ronald H. Coase, The Problem of Social Cost, 3 J.L. & ECON. 1 (1960). Though, in a family feud, acrimony may prevent frictionless bargaining. See Farnsworth, supra note 91.
motivated,\textsuperscript{156} or it may be a rational economic decision.\textsuperscript{157} Economists agree that temporal discounting plays a fundamental role in evaluating outcomes, even with nonmonetary resources.\textsuperscript{158} Abe will therefore prefer to be the first to try the new game, but will want to hold off cleaning the dishes, and it will be economically rational for him to do so. Thus under either an objective or a subjective test of evaluation, the first taker in rotation gains a larger share of the allocation than do subsequent players. So without adjusting\textsuperscript{159} the shares by discounting the allocation to subsequent parties\textsuperscript{160} the rotation method fails envy-freeness and proportionality.

\textbf{D. Partition by Sale}

Because most partition actions today involve houses and other developed land, courts in practice usually order partition by sale,\textsuperscript{161} despite the stated preference for partition in kind.\textsuperscript{162} In some jurisdictions sale is now the preferred method,\textsuperscript{163} and many scholars have praised the de facto shift to a preference in

\begin{itemize}
\item \textsuperscript{156} See Soman et al., supra note 155, at 351-52 (accounting for factors such as self-control and delayed gratification).
\item \textsuperscript{157} See id. at 356.
\item \textsuperscript{158} That is, even for resources that cannot be invested, and do not succumb to inflation or depreciation (such as eating candy now versus being overweight later, getting a massage now versus getting a massage later). See Shane Frederick, George Loewenstein & Ted O’Donoghue, \textit{Time Discounting and Time Preference: A Critical Review}, 40 J. ECON. LITERATURE 351 (2002), available at http://www.jstor.org/stable/2698382.
\item \textsuperscript{159} See RAMS & TAYLOR, WIN-WIN SOLUTION, supra note 17, at 38-39. The authors contrast strict alternation, where players take turns choosing one of many items to be distributed, to balanced alternation, compensating subsequent choosers for not choosing earlier. Strict and balanced alternation work when many distinct resources must be divided, as when basketball teams take turns drafting players. See id. at 27. If one or few items are at issue, strict and balanced alternation are inapplicable.
\item \textsuperscript{160} See Ayres & Talley, supra note 89, at 1080-81. Here, authors Ayres and Talley temporally divide ownership of real property into an initial term of years to one party, with the remainder to the other party in fee simple. Using a discount rate of 10 percent, they calculate that the parties “should be indifferent between receiving a claim to the first 6.93 years or receiving a claim to all subsequent years (in perpetuity).” Id.
\item \textsuperscript{161} See 2 AMERICAN LAW OF PROPERTY, supra note 43, § 6.26 (writing in 1952 that “[a]s a practical matter the modern practice is to decree a sale . . . . in the great majority of cases, which usually involve single parcels of improved property”).
\item \textsuperscript{162} See Loyd, supra note 27, at 188; Delfino v. Vealencis, 436 A.2d 27, 30 (Conn. 1980); Miceli & Sirmans, supra note 53, at 787.
\item \textsuperscript{163} See, e.g., IOWA CODE ANN. § 1.1201(2) (West 2002) (“Property shall be partitioned by sale . . . . unless a party prays for partition in kind . . . .”); Fannin v. Fannin, 75 S.W.2d 1042, 1043 (Ky. 1934) (“[T]he court will presume . . . . that a town lot is not susceptible of advantageous division . . . .”).
\end{itemize}
favor of sale.\textsuperscript{164} Most of this academic praise, however, is warranted only as a preference for sale as the better alternative to physical division. However, when partition by sale is examined independently, rather than as the better alternative to partition in kind, several problems emerge. First, forced sale does not adequately compensate owners for the subjective value of their property. This is particularly problematic if the property has purely sentimental value. Second, forced sale may work hardship by dispossessing people of their homes or ancestral land. Finally, the many shortcomings of auctions and other methods of judicial sale result in unfair allocations.

1. Protecting Subjective Value

Generally, people cannot be forced to sell their property, but rather may decide whether to sell their property and how much they must get in return for doing so.\textsuperscript{165} Yet in partition by sale, owners are forced to sell their property interests, and do not get to decide their compensation for doing so. Instead, partition by sale protects concurrent owners' interests with a liability rule\textsuperscript{166}: the owner is forced to relinquish the entitlement in exchange for the entitlement's objective value.\textsuperscript{167} In partition actions, an appraiser, panel of commissioners, or the will of the free market by sale at public auction determines the property's objective, or “market,” value. However, owners are often unwilling\textsuperscript{168} to part with their property at market value, and instead will subjectively evaluate their property at a price

\textsuperscript{164} See, e.g., Miceli & Sirmans, \textit{supra} note 53 (indicating the courts' preference for sale conforms to their standard); Bell & Parchomovsky, \textit{Theory of Property, supra} note 2, at 601 (noting that in consonance with the article's value theory of property, by favoring partition by sale “courts have reached precisely the right result”); see also Reid, \textit{supra} note 57, at 856 (suggesting that “courts promote economic efficiency when they favor judicial sales”).

\textsuperscript{165} See generally Calabresi & Melamed, \textit{supra} note 8 (discussing the protection of such entitlements with a property rule).

\textsuperscript{166} See Miceli & Sirmans, \textit{supra} note 53, at 784, 789-90.

\textsuperscript{167} See Calabresi & Melamed, \textit{supra} note 8. In eminent domain cases, it is generally held that “the condemnee is entitled to market value based on what the appraiser determines to be the highest or best use of the property . . . not merely the market value based on the condition of the property at that time.” Asper, \textit{supra} note 11, at 498.

\textsuperscript{168} That is, if their entitlement is protected with a property rule, giving them the option of retaining the entitlement or voluntarily relinquishing the entitlement for whatever compensation they feel adequate. If protected by a liability rule, the unwilling owner may be forced to part with his property for an objective price. See Calabresi & Melamed, \textit{supra} note 8, at 1108 (“Taney may be sentimentally attached to his land. As a result, eminent domain [using a liability rule] may grossly undervalue what Taney would actually sell for.”).
much higher than market value. In fact, a substantial gap often exists between a unique object’s market and reserve prices. Thus, a liability rule undercompensates owners by the amount that the owner’s subjective valuation of the entitlement exceeds its objective value.

The McDowell rocking chair scenario confounded the court because the court believed it could not sell a chair with no market value. The true problem, however, was not the chair’s lack of market value, but rather that the chair’s sentimental value outweighed its market value. Because an owner’s subjective valuation will often exceed the property’s fair market value, courts should analyze most partition actions—even of a multimillion-dollar home—the same way as that of an old rocking chair.

There are several reasons why subjective value will be greater than objective value. First, sentimental value creates “emotional utility” for the owner that is not accounted for in the market price. Whereas an old clock only holds intrinsic utility to the market (as just an old clock), to an owner with sentimental attachment to the property the clock has both the utility of an old clock plus its additional value as a family heirloom. Second, a quality known as delight in ownership describes the utility derived from the ownership rather than from the value of the item. In other words, this additional value results not from the worth of an object, but simply from the satisfaction in owning it.

Delight in ownership is related to, but distinct from, another concept called the endowment effect, which describes an irrational notion people have that items they own are more

169 See Bell & Parchomovsky, Theory of Property, supra note 2, at 567-68 (elaborating on the personhood perspective of property established in Margaret Jane Radin, Property and Personhood, 34 STAN. L. REV. 957 (1982)).
170 See Miceli & Sirmans, supra note 53, at 783-84.
172 See Bell & Parchomovsky, Theory of Property, supra note 2, at 567-68.
173 See id. at 558 (referencing Meir Dan-Cohen, The Value of Ownership, GLOBAL JURIST FRONTIERS, at 1 (2001)). Consider the additional value in being a homeowner, rather than a renter with otherwise comparable rights to the property. See Asper, supra note 11, at 491 (“Subjective value in the home results from the personal dignity and social status that accompany homeownership, as well as the sentimental value an individual places on the home and surrounding land.”).
valuable than identical items they do not own. Essentially, whereas a person would not be willing to spend more than $100 for a clock at a store, if a houseguest offered to buy the homeowner’s clock, the lowest price the homeowner would be willing to sell it for would be much higher than $100.

A fourth reason an owner’s value of property would be higher than market price is that the owner may have a unique skill or ability that allows him to derive utility from the property others cannot. If Ben is one of the only people who knows how to play the paixiao (a type of Chinese panpipe last popular during the Song Dynasty), he can gain usefulness from it that others cannot. He will therefore value the paixiao more highly than the market would, since it cannot extract similar utility. For the same reason, a company holding the only license to mine uranium will be, by virtue of the state-sanctioned monopoly, the only entity allowed to generate the greatest use out of uranium-rich land, and will therefore value the land more highly than would the market.

Finally, a person may have invested a great deal of time and effort learning how to use a particular irreplaceable object to its greatest effectiveness; if the object were lost or destroyed, he would have to reinvest the same amount of time and effort to learn how to use a replacement object just as well. For instance, Adam may have spent years getting used to the nuances of his baseball glove, which now conforms and responds perfectly to his hand. If he had to use a new glove, he would have to spend just as long readjusting both the glove and

---


176 See Levinsohn-Zamir, *supra* note 9, at 250-51.

177 See Bell & Parchomovsky, *Theory of Property*, supra note 2, at 568.

178 Abraham Bell and Gideon Parchomovsky use the example of the harpsichord. See id. at 568. Though the skills necessary to extract utility from the harpsichord seem too transferable to be rhetorically effective here.


181 See id. at 569 (citing RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* 95-96 (5th ed. 1998)).
himself to use it as well as he had used his old glove.\textsuperscript{182} Thus, Adam factors in the time and effort invested in learning how to use his glove into his valuation of it\textsuperscript{183} and will discount that investment in the cost of continuing to use the glove.\textsuperscript{184} The market has not made the same investment, and will therefore value his glove at a lower price.\textsuperscript{185}

By compensating owners with mere market value, forced sale harms those who have subjective ties to the property.\textsuperscript{186} Some courts in partition actions have taken sentimental attachment into consideration,\textsuperscript{187} although often only as one factor among many rather than as dispositive.\textsuperscript{188} Because partition is an equitable remedy,\textsuperscript{189} courts should consider these types of equitable circumstances, though they are curiously unwilling to do so.\textsuperscript{190} Still, fair division allocation must consider subjective value,\textsuperscript{191} at least because it is a necessary component of the no-envy test.\textsuperscript{192}

2. Dispossession Through Forced Sale

Although courts are generally unwilling to entertain subjective concerns in partition actions, where forced sale would work hardship, some courts have been willing to reject sale in favor of partition in kind.\textsuperscript{193} Courts appear most willing to consider the particular hardships resulting from

\begin{enumerate}
\item See id. at 568. For a more substantial example, consider a company that after decades of production has learned to operate a nuanced factory to utmost efficiency.\textsuperscript{182}
\item See id.\textsuperscript{186}
\item See id. at 569 n.205 (“Where there are only close, but not identical substitutes, substantial costs may be involved in learning how to enjoy a substitute item’s full value, whereas all such costs in the currently possessed item have already been sunk. Consequently, the marginal cost of continuing to use the possessed item will no longer include the cost of learning how to use it, while such costs will continue to be reflected in market prices.”).\textsuperscript{184}
\item See id. at 568.\textsuperscript{185}
\item See Ark Land Co. v. Harper, 599 S.E.2d 754, 761 (W. Va. 2004).\textsuperscript{186}
\item See id. at 760-61 (citing Schnell v. Schnell, 364 N.W.2d 713, 721 (N.D. 1984); Fike v. Sharer, 571 P.2d 1252, 1254 (Or. 1977)).\textsuperscript{187}
\item See Zimmerman v. Marsh, 618 S.E.2d 898, 901 (S.C. 2005); see also Campbell v. Jordan, 675 S.E.2d 801, 804 (S.C. 2009) (citing Zimmerman, 618 S.E.2d 898).\textsuperscript{188}
\item See Reitmeier v. Kalinoski, 631 F. Supp. 565, 575 (D.N.J. 1986).\textsuperscript{189}
\item See 4 THOMPSON ON REAL PROPERTY, supra note 54, § 38.03(a)(2)(iii); Girard, supra note 65, at 1433.\textsuperscript{190}
\item See Asper, supra note 11, at 501 (“Subjective values that can be monetized must be part of any compensation that is just.”).\textsuperscript{191}
\item See supra Part II.B.\textsuperscript{192}
\item See Wilk v. Wilk, 795 A.2d 1191, 1192 (Vt. 2002) (“Forced sale is disfavored because the Legislature, and the common law . . . sought to minimize the forced divestiture of family property where avoidable.”).\textsuperscript{193}
\end{enumerate}
dispossession of a family in residence,\textsuperscript{194} and so, where a family will be dispossessed, a court may deny partition altogether.\textsuperscript{195} Furthermore, courts and scholars have criticized partition by sale for exacerbating the loss of ancestral lands (generally family farms) to opportunistic developers.\textsuperscript{196} In many particularly egregious cases, properties that had been in black families for generations were taken over by developers through forced partition, an act which scholarship suggests has been a major reason for the precipitous decline in black landownership over the last century.\textsuperscript{197} The recently published UPHPA attempts, in part, to prevent “unscrupulous real estate speculators [from] purchas[ing] a very small interest in family-owned tenancy-in-common property with the sole purpose of seeking a court-ordered partition by sale,” thereby winning the property at auction for a price below market-value.\textsuperscript{198} At least one jurisdiction has enacted corrective legislation to keep property within families.\textsuperscript{199}

3. Problems with Auctions

Concededly, where the resource to be partitioned is fungible and extremely liquid, no real problems exist with selling the property and dividing the proceeds among the co-owners. However, when faced with the facts of \textit{McDowell}, partition by sale does not result in a fair solution. Because the rocking chair has purely sentimental value, it will fetch a paltry sum at auction; each litigant will get half of this small sum, and so both will leave as losers. If the siblings could bid on the chair at this auction, one of them would end up as the

\textsuperscript{194} See \textit{id.} at 1195 (“Forcing partition by sale when more than one co-tenant is willing to take assignment of the property could result in the unnecessary forced divestment of numerous family farms.”).  

\textsuperscript{195} See \textit{Reitmeier v. Kalinoski,} 631 F. Supp. 565, 575 (1986) (“Where a family will be put out of its home by the sale in partition, a bill for partition will be denied.”). \textit{But see Heldt v. Heldt,} 193 N.E.2d 7, 64-65 (Ill. 1963); \textit{see also 4 THOMPSON ON REAL PROPERTY, supra} note 54, § 38.03(a)(1) (citing \textit{Heldt,} 193 N.E.2d 7).  

\textsuperscript{196} See generally Anna Stolley Persky, \textit{In the Cross-Heirs: A Loophole in Real Estate Law Pits Families Against Developers and Each Other. Some Say There’s More Than Money at Stake,} 95 A.B.A. J. 44 (2009) (discussing how “land owned for generations is suddenly lost” to developers and land speculators through partition sales).  

\textsuperscript{197} See generally \textit{Casagrande, supra} note 56; \textit{see also UPHPA, supra} note 57, preface & cmts.; Persky, \textit{supra} note 196 (discussing how particularly African-Americans are displaced from “heir’s property” and how the UPHPA serves to abate the problem).  

\textsuperscript{198} UPHPA, \textit{supra} note 57, prefatory notes.  

\textsuperscript{199} See ARK. CODE ANN. § 18-60-404 (2010) (providing that a minority shareholder party not related to other co-owners by four degrees of consanguinity cannot petition for partition within three years of the purchase of that interest).
higher bidder (because no member of the public would be willing to pay extra for a rocking chair that only has sentimental value). If \( x \) is the price the winning sibling, \( A \), pays at auction, \( x \) is divided between the two, so at the end of the day, \( A \) will leave with the rocking chair minus \( x/2 \), and the other sibling, \( B \), will leave with \( x/2 \). Here both parties leave with ostensibly the same net value: \( B \) with \( x/2 \), and \( A \) with the chair (valued at \( x \)) minus \( x/2 \). However, because of the nature of bidding at an auction, \( x \) does not accurately reflect winning bidder \( A \)'s subjective valuation of the chair; rather it simply reflects the price higher than which \( B \) was unwilling to spend.

Suppose \( A \) valued the chair at $10,000 (meaning he subjectively valued his interest at $5000) and \( B \) valued the chair at $6000 (meaning she subjectively valued her interest at $3000). If \( A \) places a bid for $6000, \( B \) will be unwilling to go higher and will not bid; therefore \( A \) will win the chair and pay half this amount ($3000) to \( B \). Because \( A \) values the chair at $10,000, he values his final allocation at $7000 (the value of the chair minus the $3000 he paid to \( B \)). \( A \) therefore gains a $2000 surplus (the $7000 he received over the $5000 he thought his interest was worth) while \( B \) gains no surplus at all (since the $3000 she received was equal to her initial valuation). Thus, although the allocation satisfies the envy-free criterion, it violates equitability, because the surpluses are unequal.

In a sealed-bid auction, parties submit one bid each, with the chair going to whomever submitted the highest amount (here, \( A \)'s $10,000 bid). The result in a sealed-bid auction is the reverse of the result in the normal auction: \( A \) leaves with the $5000 he expected and thus with no surplus, but \( B \) gains a $2000 surplus over the $3000 she expected.

The transaction costs associated with holding an auction also make this solution inefficient. When it orders a judicial sale, the court will normally appoint a panel of commissioners or a referee to conduct a public or private sale of the property. The proceeds of the sale go toward paying attorneys’ fees, fees for the commissioner’s appraisal, expenses of the auction, and other costs, with the remainder paid to the

---


201 See Girard, supra note 65, at 1429-30.

202 See CAL. CIV. PROC. CODE § 874.010 (Deering 2010) (“The costs of partition include [inter alia]: . . . attorney’s fees . . . fee and expenses of the
parties based on their respective shares.293 With these costs eliminated, both parties would leave with larger shares, resulting in a Pareto-superior allocation. Finally, judicial sales notoriously yield submarket prices. For instance, following the Popov v. Hayashi decision, the Barry Bonds ball did not fetch nearly as much at auction as had been originally estimated (an amount, as it turns out, which was several thousand dollars less than Popov’s legal bill).294

E. Partition by Allotment

The two recognized judicial methods of partition are partition in kind and partition by sale. However, in 30 percent of the states, courts have occasionally engaged in a third,295 generally unacknowledged296 method of partition known as partition by allotment.297 The Supreme Court of South Carolina defines “partition by allotment” as a mode of partition “whereby one joint owner is allotted the entire property” and pays the others for their respective interests.298 Many

referee . . . compensation . . . for services of a surveyor or other person employed by the referee . . . costs of a title report . . . ?); see also Riley v. Turpin, 349 P.2d 63 (Cal. 1960).


294 See DUKEMINIER ET AL., supra note 5, at 111.

295 Variations on partition by allotment may avoid its characterization as something separate from partition by sale or in kind. For instance, in an in kind division with an owelty payment, it is conceivable that the property be physically divided into one part containing the entire property and the other containing nothing. Furthermore, the winning bidder at a partition sale may be one of the co-owners.

296 The term “partition by allotment” has no entry in American Jurisprudence, American Law Reports, Corpus Juris Secundum, Black’s Law Dictionary, Ballentine’s Law Dictionary, or Words and Phrases, and is not mentioned in the Restatement (First) of Property. Confusingly, the term is sometimes used to refer to partition in kind. See, e.g., Wilk v. Wilk, 795 A.2d 1191, 1193 (Vt. 2002).

297 In Maine it is referred to as “partition by buy-out,” see Libby v. Lorrain, 430 A.2d 37, 39-40 (Me. 1981), and in Vermont it is referred to as “assignment,” see Wilk, 795 A.2d at 1193 (“When . . . the real estate . . . cannot be divided without great inconvenience . . . the court may order it assigned to one of the parties, provided he pays to the other . . . such sum . . . as the commissioners judge equitable.” (quoting VT. STAT. ANN. tit. 12, § 5174)).

298 Zimmerman v. Marsh, 618 S.E.2d 898, 901 (S.C. 2005); see also id. at 902 (Peleicones, J., dissenting); Pruitt v. Pruitt, 380 S.E.2d 862, 864 (S.C. Ct. App. 1989) (“Partition by allotment to one of the parties”); see also Austin v. Dobbins, 252 S.E.2d 588, 590-91 (Va. 1979) (Virginia’s statute “expressly authorizes partition by allotment of the whole property to one or more coparceners . . . or to a tenant in common”); Faith Rivers, Inequity in Equity: The Tragedy of Tenancy in Common for Heirs’ Property Owners Facing Partition in Equity, 17 TEMP. POL. & CIV. RTS. L. REV. 1, 59, 72-73 (2007) (“Where one cotenant seeks to maintain the real property, a court may utilize partition by allotment to allow one cotenant to buy out other cotenants through payment of the appraised value.” (citing Zimmerman, 618 S.E.2d at 900)).
jurisdictions do not permit this mode of partition. Thirteen states, however, recognize partition by allotment under statute. Courts in two states allow partition by allotment.

29 See, e.g., Thompson v. Celestain, 936 So.2d 219, 222 (La. Ct. App. 2006) (holding the property must be sold publicly, and that assigning the property to one of the parties was in error); Onderdonk v. Onderdonk, 307 A.2d 710, 712 (Md. 1973) (holding partition with owelty does not allow dispossessing but reimbursing some co-owners); see also Zimmerman, 618 S.E.2d at 902 (Pelicones, J., dissenting); see also Soriano v. Soriano, 643 P.2d 450, 452 (Wash. Ct. App. 1982) (in a divorce proceeding, court violated statutory duty by providing for a private auction where "the husband and wife [were] the only persons permitted to bid and that each item subject to auction be awarded to the person submitting the highest bid on that item"). But see Dougherty v. Dougherty, 210 N.W.2d 151, 153 (Mich. Ct. App. 1973) (divorce court's use of sealed bids in lieu of partition was not abuse of discretion).

29 See Alabama: ALA. CODE § 35-6-100 (LexisNexis 2010); Prince v. Hunter, 388 So. 2d 546, 547 (Ala. 1980) (sale reversed because statute mandates joint owners be given opportunity to purchase other owners' shares). But see Jolly v. Knopf, 465 So. 2d 150 (Ala. 1986) (challenging the statute's constitutionality). Georgia: GA. CODE ANN. § 44-6-166.1 (2010) (allowing parties to buy out others' shares at appraised price); Lassiter Props. v. Gresham, 371 S.E.2d 650, 651 (Ga. 1988) (holding that statute allows party to purchase others' shares to avoid sale); Clements v. Seaboard Air Line Ry. Co., 124 S.E. 516, 517 (Ga. 1924) (where three railroads owned depot as tenants in common, court ordered value of the property be set and defendants be given opportunity to buy plaintiff's share). Maine: ME. REV. STAT. ANN. tit. 14, § 6515 (2010) ("When any parcel . . . is of greater value than either party's share and cannot be divided without great inconvenience, it may be assigned to one party by his paying the sum of the money awarded to the parties who have less than their shares."); Libby v. Lorrain, 430 A.2d 37, 39-40 (Me. 1981) (holding "partition by buy-out" is permitted, but inappropriate where party requesting the allotment failed to show she had the financial ability to pay); Dyer v. Lowell, 30 Me. 217, 219 (1849) ("[I]f the estate was incapable of division, [the court] should have set off the whole to one of the co-tenants" under the statute.). Maryland: Catlin v. Catlin, 60 Md. 573 (1883) (under statute, eldest heir of intestate descendent may elect to take entire estate and pay to the other heirs their proportionate shares). But see Onderdonk, 307 A.2d at 712 (concept of owelty does not allow partition dispossessing but reimbursing co-owners). Minnesota: MINN. STAT. ANN. § 558.12 (West 2010) (With a "tenement which cannot be divided . . . the whole premises . . . may be set off to any party who will accept it . . . paying to . . . the others such sums . . . as the referees award."); Cozzi v. Cozzi, 391 N.W.2d 25, 28 (Minn. Ct. App. 1986). Ohio: OHIO REV. CODE ANN. § 5307.09 (LexisNexis 2004) ("If one or more of the parties elects to take the estate at the [commissioners'] appraised value, it shall be adjudged to them, upon their paying to the other parties their proportion of its appraised value."); Sword v. Sword, 620 N.E.2d 199, 204 (Ohio Ct. App. 1993) (a party may elect to take estate at appraised value). Oklahoma: OKLA. STAT. ANN. tit. 12, § 1512 (West 2010) ("If partition cannot be made, and the property shall have been . . . appraised, any one . . . of the parties may elect to take . . . at the appraisement . . . on payment . . . of their proportion of the appraised value."); Sun Inv. & Loan Corp. v. McIntyre, 537 P.2d 341, 344 (Okla. 1975); Herron Trust v. Swarts, 361 P.2d 280 (Okla. 1961). Pennsylvania: 42 PA. CONS. STAT. ANN. § 1563 (West 2002) (party objecting to sale may be awarded the property and pay parties requesting partition and sale the amounts of their interests based upon the court's valuation); Beall v. Hare, 174 A.2d 847, 849 (Pa. 1961); cf. Harbin v. Harde, 14 A.2d 866, 867 (Pa. 1940) (deciding under an older statute). South Carolina: S.C. CODE ANN. § 15-61-25 (2009); Zimmerman, 618 S.E.2d at 901; Cox v. Frierson, 451 S.E.2d 392, 393 (S.C. 1994) (per curium) (reversing sale because referee ordered sale without first considering partition by allotment). Vermont: VT. STAT. ANN. tit. 12, § 5174 (2002) ("When the real estate . . . cannot be divided . . . the court may order it assigned to one of the parties, provided he pays to the other party such sum . . . as the
even in the absence of statutory authority. Five states explicitly prefer partition by allotment to partition by sale, and thus courts may hold sale improper if allotment is not first considered. The Uniform Probate Code section 3-911, adopted by seventeen states, allows a court to sell estate “property which cannot be partitioned without prejudice to the commissioners judge equitable.”; Wilk, 795 A.2d at 1194 (holding under statute, partition in kind is preferred over partition by allotment, and allotment preferred over partition by sale). Virginia: VA. CODE ANN. § 8.01-83 (2010) (“When partition cannot be conveniently made, the entire subject may be allotted to any one or more of the parties.”); Quillen v. Tull, 312 S.E.2d 278, 281 (Va. 1984); Austin, 252 S.E.2d at 590-91 (statute “expressly authorizes partition by allotment of the whole property to one or more coparceners . . . or to a tenant in common”); Price v. Simpson, 29 S.E.2d 394, 396 (Va. 1944); Roberts v. Hagan, 93 S.E. 619, 621 (Va. 1917) (under statute “it would have been entirely within the power of the court . . . to have assigned to [Plaintiff] the entire estate to be partitioned, upon his payment to [Defendant] of the amount to which she would be entitled for her interest”). West Virginia: W. VA. CODE ANN. § 37-4-3 (LexisNexis 2005) (“When partition cannot be conveniently made, the entire subject may be allotted to any party.”); State ex rel. Bowser v. Hill, 550 S.E.2d 62, 65-66 (W. Va. 2001) (discussing Corrothers v. Jolliffe, 9 S.E. 889 (W. Va. 1899)); Smith v. Smith, 376 S.E.2d 97, 102 (W. Va. 1988) (allowing the allotment of the entire property, with payment based on appraised value of a panel of “three disinterested and qualified persons”); Corrothers, 9 S.E. at 890 (by statute, “when partition in kind cannot be conveniently made, the court may . . . allot the entire subject to a party . . . who offers the largest proportional price for it”). Wyoming: WYO. STAT. ANN. § 1-32-109 (2011) (“When . . . the estate cannot be divided . . . without manifest injury . . . and one . . . of the parties elects to take the estate at [the commissioners’] appraised value, it shall be adjudged to him upon his paying to the other parties their proportion of the appraised value.”); Hutchins v. Payless Auto Sales, Inc., 85 P.3d 1010, 1013 (Wyo. 2004); In re Estate of Sorenson, 9 P.3d 259, 262 (Wyo. 2000).

211 See Morris v. Tracy, 48 P. 571, 572-73 (Kan. 1897) (where commissioners report that partition in kind cannot be made without manifest injury, it was error to direct a sale of the property without first allowing one or more of the parties to take the land at the appraised value); Reitmeier v. Kaliolesi, 631 F. Supp. 565, 578 (D.N.J. 1986) (predicting that New Jersey “would permit a partition in which one party took the entire property and compensated the other with an owelty”); Baker v. Drabik, 541 A.2d 229, 232-33 (N.J. Super. Ct. App. Div. 1988) (analyzing Reitmeier, 631 F. Supp. 565, and holding party may purchase the other’s interest at its fair market value).

212 See Prince, 388 So. 2d at 547 (order of sale reversed because Alabama statute mandates joint owners be given opportunity to purchase other owners’ shares); Dyer, 30 Me. at 219 (“If the estate was incapable of division, they should have set off the whole to one of the co-tenants” according to Maine statute); Morris, 48 P. at 573 (holding it was error to direct a sale of the property without first allowing one or more of the parties to take the land at the appraised value); Cox, 451 S.E.2d at 393 (per curium) (reversing order for judicial sale because referee ordered the sale without first considering partition by allotment); Wilk, 795 A.2d at 1194 (under statute, partition in kind is preferred over partition by allotment, and partition by allotment is preferred over partition by sale).

owners and which cannot conveniently be *allotted to any one party.* Additionally, the UPHPA requires that courts allow the co-owners who did not request partition by sale to buy the interests of those cotenants who did request sale at a price based on the initial appraisal valuation of the property.\(^{216}\)

Under most, if not all, such allotment schemes, the price for which a party may purchase the other owners' interests is set objectively *ex ante*, usually by a panel of three commissioners charged with appraising the property.\(^{216}\) Because of commissioners' and surveyors' appraisal fees, the resulting allocation cannot be Pareto-optimal. Also, if the appraised value is too low (that is, lower than the parties' subjective value), then more than one party will be willing to take the property.\(^{218}\) A Pennsylvania court ruled that where both parties were willing to take the property at valuation, it was proper to have the parties submit sealed bids and allot the property to the highest bidder.\(^{219}\) Vermont "allows for assignment even when more than one co-tenant is willing to accept it, and gives the trial court discretion over whether to order an assignment and the choice of assignee."\(^{220}\)


\(^{216}\) See UPHPA, supra note 57, § 7 & accompanying cmts.


\(^{218}\) Clearly valuation was set too low in these instances.

\(^{219}\) See Harbin v. Harde, 14 A.2d 866, 867 (Pa. Super. Ct. 1940) (so holding, even though the party with the losing bid inherited his share and the winner bought her interest from other heirs).

\(^{220}\) See Wilk v. Wilk, 795 A.2d 1191, 1196 (Vt. 2002) (holding no abuse of discretion where court ordered 1/8 owning brother with adjacent junkyard to transfer his share to 7/8 owning brother operating a paving business on the property, where both brothers were willing to take the property at valuation).
Table 1: Partition by Allotment

<table>
<thead>
<tr>
<th></th>
<th>Subjective value</th>
<th>Interest subjective value</th>
<th>Objective value</th>
<th>Interest objective value (owelty)</th>
<th>Allocation</th>
<th>Allocation subjective value</th>
<th>Envy value</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer</td>
<td>$v_A$</td>
<td>$v_A / 2$</td>
<td>$M$</td>
<td>$M - M / 2$</td>
<td>$v_A / 2$</td>
<td>$M / 2$</td>
<td>$v_A / 2 - M / 2$</td>
<td></td>
</tr>
<tr>
<td>Seller</td>
<td>$v_B$</td>
<td>$v_B / 2$</td>
<td>$M$</td>
<td>$M / 2$</td>
<td>$v_B / 2$</td>
<td>$M / 2$</td>
<td>$v_B / 2 - M / 2$</td>
<td></td>
</tr>
</tbody>
</table>

The allocation resulting from partition by allotment is not necessarily proportional, equitable, or envy-free. Say Amy and Bruce are partitioning a painting, $p$, created by Amy’s grandfather, a famous artist. The market value ($M$) of the painting is $70,000. Because of its sentimental value, Amy’s valuation ($v_A$) of the painting is $100,000. Bruce, on the other hand, because of the delight in owning a genuine masterpiece and the endowment effect, values the painting at $v_B$, $80,000. Thus, Amy values her half interest at $50,000 and Bruce values his half interest at $40,000. However, because the market value is $70,000, the objective value of each of their interests is only $35,000, and so the owelty is set at that amount. Under their state’s allotment statute, Amy buys Bruce’s interest in the painting for $35,000 and gets full ownership of it.

Table 2: Amy and Bruce

<table>
<thead>
<tr>
<th></th>
<th>Subjective value of $p$</th>
<th>Interest subjective value of $p$</th>
<th>Objective value of $p$</th>
<th>Interest objective value of $p$</th>
<th>Allocation</th>
<th>Allocation subjective value</th>
<th>Envy value</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amy</td>
<td>$100,000</td>
<td>$50,000</td>
<td>$70,000</td>
<td>$35,000</td>
<td>$p - $35,000</td>
<td>$v_A / 2 - $35,000</td>
<td>$15,000</td>
<td></td>
</tr>
<tr>
<td>Bruce</td>
<td>$80,000</td>
<td>$40,000</td>
<td>$70,000</td>
<td>$35,000</td>
<td>$v_B / 2 - $35,000</td>
<td>$45,000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The allocation fails the no-envy test because Bruce sees Amy’s allocation as more valuable than his own. Since he values the painting at $80,000, to him Amy’s allocation is worth $45,000, (which of course is greater than the $35,000 he received in cash), and so Bruce would rather have her allocation. Additionally, the allocation is not proportional because Bruce expected to get at least $40,000 (the subjective value of his interest) but left with $35,000. Finally, the allocation is not equitable because Amy’s surplus is not equal to Bruce’s surplus. Partition by allotment may sometimes meet
many of the normative criteria of fair division, but it cannot guarantee fulfilling most or all of them.

A slightly different scenario demonstrates that in partition by allotment the surplus values will always be different unless the owelty is equal to the average of the initial subjective evaluations. Say Cody and Daron buy a house, Blackacre, expecting to move in together. Things fall apart; Daron moves to the next town, but Cody continues to occupy the house, both living and operating a profitable orthodontics practice there. A year later, Daron is in need of cash and brings an action for partition of Blackacre, now valued at $90,000. Daron is only interested in the money and would be happy to get half of whatever Blackacre is worth on the open market. Cody, on the other hand, having an established reputation in the community, does not want to move and so values retaining Blackacre at $120,000.

Table 3: Cody and Daron

<table>
<thead>
<tr>
<th></th>
<th>Subjective value of BA</th>
<th>Interest subjective value</th>
<th>Objective value of BA</th>
<th>Interest objective value</th>
<th>Allocation</th>
<th>Allocation subjective value</th>
<th>Envy value</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cody</td>
<td>$120,000</td>
<td>$60,000</td>
<td>$90,000</td>
<td>$45,000</td>
<td>BA - $45,000</td>
<td>$75,000</td>
<td>$45,000</td>
<td>+$15,000</td>
</tr>
<tr>
<td>Daron</td>
<td>$90,000</td>
<td>$45,000</td>
<td>$45,000</td>
<td>$45,000</td>
<td>$45,000</td>
<td>$45,000</td>
<td>$0</td>
<td></td>
</tr>
</tbody>
</table>

Here the allocation is nonenvious, because neither party prefers the other’s allocation, and it is proportional, because each party receives at least half of their value of the property. However, it is not equitable, because Cody’s surplus is larger than Daron’s. Also, because of court costs, both parties actually leave with a little less, thereby making the distribution inefficient.

IV. EQUITABLE ALLOTMENT

The main problem with partition by allotment is that it uses market value to set the payment price. If market value were eliminated from the equation and replaced with an owelty determined from the parties’ subjective valuations of the resource, the method can be reformulated\(^{221}\) to assure a proportional, envy-

---

\(^{221}\) As a variation of partition by allotment, equitable allotment may be allowable under a state’s partition statutes or rules without necessity of further legislation. See supra notes 205-11 and accompanying text.
free, efficient, and equitable allocation. This reformulated method will be referred to here as equitable allotment.

A. The Method: Overview

Equitable allotment is a fair and efficient partition method for two-party disputes over a single indivisible resource \((r)\). Under this method, the parties submit sealed bids with their subjective valuations of the resource. The buyout price, \((w)\), is half the average of these two amounts. The party submitting the highest bid will be allotted the entire resource and will pay to the other the predetermined owelty (the buyout price).  

Equitable allotment is a simplified version of Knaster’s procedure of sealed bids, which, when dividing multiple indivisible goods among several parties, guarantees proportionality but not envy-freeness. However, with only two parties, the result is always envy-free. When Knaster’s procedure is performed for two parties and one item, it is exactly the same as equitable allotment.

Table 4: Equitable Allotment

<table>
<thead>
<tr>
<th>Subjective value of (r)</th>
<th>Interest subjective value</th>
<th>Owelty ((w))</th>
<th>Allocation subjective value</th>
<th>Envy value</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer (V_a)</td>
<td>(\frac{V_a}{2})</td>
<td>(\frac{V_a + V_b}{4})</td>
<td>(r - w)</td>
<td>(V_a - w)</td>
<td>(w)</td>
</tr>
<tr>
<td>Seller (V_b)</td>
<td>(\frac{V_b}{2})</td>
<td>(\frac{V_a + V_b}{4})</td>
<td>(w)</td>
<td>(w)</td>
<td>(V_b - w)</td>
</tr>
</tbody>
</table>

To demonstrate that the allocation is truly envy-free, the court may give one of the parties the choice between either taking the resource or taking the payment. But regardless of who chooses, the party submitting the higher bid will end up with \(r\), paying \(w\) to the party submitting the lower bid. This is because, to the party submitting the lower bid, the cash is the more valuable of the two choices. On the other hand, to the party submitting the higher bid, \(r-w\) is more valuable than \(w\), and so he will gladly pay the buyout price to have the resource. The prospect of not knowing who the chooser will be may, however, give even greater incentive for honest initial evaluation. This, of course, assumes that both parties will act rationally, and strive to maximize their own wealth. See Lewinsohn-Zamir, supra note 9, at 228 (“The standard economic game-theoretic prediction is that both players will behave rationally, that is to say, strive to maximize their monetary payoffs.”).

See Brans & Taylor, Fair Division, supra note 17, at 52-56; see also Steinhaus, supra note 14 (first describing Knaster’s procedure).

See Brans & Taylor, Fair Division, supra note 17, at 55.

See id.
Table 5: Equitable Allotment Showing Inherent Values for \( r \) and \( w \)

<table>
<thead>
<tr>
<th></th>
<th>Subjective value of ( r )</th>
<th>Interest subjective value</th>
<th>Owelty (( w ))</th>
<th>Allocation</th>
<th>Allocation subjective value</th>
<th>Envy value</th>
<th>Surplus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer</td>
<td>( V_b )</td>
<td>( \frac{V_b}{2} )</td>
<td>( V_b + V_s )</td>
<td>( V_b )</td>
<td>( \frac{V_b + V_s}{4} )</td>
<td>( \frac{3V_b - V_s}{4} )</td>
<td>( \frac{-V_s}{2} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( V_b + V_s )</td>
<td>( V_b )</td>
<td>( \frac{V_b + V_s}{4} )</td>
<td>( \frac{3V_b - V_s}{4} )</td>
<td>( \frac{V_b - V_s}{4} )</td>
</tr>
<tr>
<td>Seller</td>
<td>( V_s )</td>
<td>( \frac{V_s}{2} )</td>
<td>( V_b + V_s )</td>
<td>( V_s )</td>
<td>( \frac{V_b + V_s}{4} )</td>
<td>( \frac{3V_b - V_s}{4} )</td>
<td>( \frac{-V_b}{2} )</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( V_b + V_s )</td>
<td>( V_s )</td>
<td>( \frac{V_b + V_s}{4} )</td>
<td>( \frac{3V_b - V_s}{4} )</td>
<td>( \frac{V_s - V_b}{4} )</td>
</tr>
</tbody>
</table>

Table 5 shows that the allocations and surplus values are identical for both the buyer and seller.

B. Setting the Owelty

Equitable allotment finds the proper value for \( r \) by averaging the two parties’ subjective values (\( V_b \) and \( V_s \)). If the parties have equal half-interests then the value of each interest (and hence the owelty value) is half of this average. If the parties have unequal interests in the property, the owelty can simply be adjusted for each party to reflect the ratio of their interests.

Whether the parties value \( r \) equally or differently, the owelty value arrived at will reflect the proper value of \( r \). Unless the litigants are not reasonable people, their bid amounts will very closely approximate the proper value they attach to the resource. Neither will want to pay more than he has to, and neither will want to give up the item for a lower sum than necessary. Because neither knows the other’s bid, both will have great incentive to arrive at a figure they perceive as the proper value of the property. In that case, one party leaves with the property, and the other leaves with the fair value of his interest.

Equitable allotment is “equitable” because it assures that each party’s surplus is the same. Unlike in an auction, where only one party (at most) determines the price, or in
partition by allotment, where neither party determines the price, in equitable allotment, the subjective values of both parties determine the buy-out price. By basing the owelty payment on the average of the parties’ subjective values, equitable allotment guarantees that the difference between each party’s final allotment and their proportional share is equal, even if the parties value $r$ differently. Thus, in equitable allotment, the amount by which each party feels her allocation exceeds the value of her interest will always be the same.

C. Efficiency

By keeping the resource intact, both physically and temporally, equitable allotment maximizes the overall allocation. However, it need not be used solely for indivisible resources. The equitable allotment method works just as well for property that could also have been divided physically and for resources that are liquid or fungible. More significantly, without the need for an objective appraisal of the property, both parties save on court costs, yielding a Pareto-superior outcome. In fact, the parties can bypass the judicial procedure altogether, because all equitable allotment requires is both parties together in a room with two pieces of paper and a pen. Furthermore, because this method always gives the resource to the party that values it the most, it satisfies the requirement of fitness.

D. Application

Consider how equitable allotment would be applied in a McDowell-like situation. In this scenario, Ethel and Frank both want their grandfather’s old rocking chair ($c$), which would sell on eBay for at most $150. Ethel has very fond memories of her grandfather reading to her in the chair and would pay up to $8000 to have it and be able to pass it on to her children. Frank also has sentimental attachment to the chair, but finds it somewhat uncomfortable to sit in, so he would pay up to $6000 to keep it.
If Ethel submits a bid of $8000 and Frank submits a bid of $6000, the average of the bids is $7000, so the owelty will be set at $3500. Because Ethel submitted the highest bid, she will be allotted the chair and be ordered to pay Frank $3500. Ethel views the total value of her allocation at $4500 and so is not envious of Frank’s $3500 award. And from Frank’s perspective Ethel’s allocation is only worth $2500, so he is happier with his payment and is not envious of her allocation.

The facts of Pugh v. NPC Services, Inc. present a unique circumstance where equitable allotment would have helped the court reach a more efficient and fair outcome. In that case, the court was confounded when, because of hazardous waste contamination, the indivisible subject property had either “no value or a negative value.” What the court did not realize was that the value to the parties—and the apparent reason for the partition action—was to divest themselves of the contaminated land. Therefore, equitable allotment should apportion this divestiture.

Using equitable allotment in this situation, the question each party should be asked is, “What is it worth to you to step away from this mess?” The answer to this will be based on what each party estimates it will cost them to clean up the waste, as well as what the underlying property would be worth to them if the waste were cleaned up. The party submitting the higher bid would be allotted the right to step away, while the party submitting the lower bid would be awarded the owelty but would remain with the property.

E. Bidding Strategy

There are several limitations to equitable allotment, one of which is its susceptibility to gamesmanship. However, in the
overwhelming majority of situations, this flaw is not fatal since
gamesmanship is prohibitively risky. Recall Ethel and Frank’s
partition of their rocking chair. Ethel knows that in equitable
allotment, the chair goes to the person who submits the highest
bid, so her initial strategy is to bid high. But then she
remembers that the higher she bids, the more she will have to
pay. Frank reflects on his reserve price and suspects his sister
will probably be willing to pay somewhat more than he would,
considering he knows how fond she is of the chair. His initial
strategy then is to raise his bid just enough that Ethel will get
the chair and pay him more than she would have otherwise for
his interest. However, this strategy is very risky; he does not
know exactly what Ethel’s subjective valuation is, so if he
raises his bid too much, he will get the chair and overpay for it.
The opposite is going through Ethel’s head: she suspects her
bid is greater than her brother’s, so if she lowers her valuation
just enough, she can still end up with the chair but pay less for
it than she would otherwise. Again, this too is risky because if
she lowers her bid too much, she will lose the chair and get a
paltry sum in compensation.

This shows how equitable allotment is not strategy-proof,
but if the parties’ valuations are greater than the market price,
gamesmanship is very risky. If the players reveal their
strategy, the entire method falls apart. This will most likely
happen if one of the parties makes known he is only interested
in the market value.

Another problem with equitable allotment is that it is
unavailable if at least one of the parties cannot pay the owelty.
Unlike with chance, rotation, partition in kind and by sale, the
parties must have enough money to cover the owelty in order to
participate in equitable allotment (or partition by allotment). If a party’s only asset is the resource being partitioned, equitable allotment is not possible.

---

229 See BRAMS & TAYLOR, FAIR DIVISION, supra note 17, at 56. Discussing
strategic misrepresentation in Knaster’s procedure, the authors conclude that while
the procedure is not strategy-proof, because of the high risk of misrepresentation
“honest evaluations, in many situations . . . may be the pragmatic thing to do.” Id.
(quoting HOWARD RAIFFA, THE ART AND SCIENCE OF NEGOTIATION 296 (1982)). Cf.
Fragnelli & Marina, supra note 114.

230 See BRAMS & TAYLOR, FAIR DIVISION, supra note 17, at 56.
CONCLUSION

Although centuries old, American partition law has not developed the subtleties of contract or tort law. Instead, the laws of partition have developed haphazardly and bluntly, and so co-owners are left with only a few narrow options. The problem with the available methods is that the courts are so focused on the fairest solution that they fail to use a solution that is both fair and efficient.231 On the other hand, economists often focus solely on efficiency, and leave fairness considerations to other disciplines.232 Instead of being beholden to the old rules of partition, courts should use economic theories to fairly and efficiently allocate resources. Of course, reality is more complex than theory, and in the real world, there are often factors in family disputes or divorces that defy simplistic mathematics. There are no easy solutions.233 The goal of this note, therefore, is to spark debate about judicial partition schemes, and encourage an economic approach to partition that takes subjective value into consideration.

Zachary D. Kuperman†

231 See Blonquist v. Frandsen, 694 P.2d 595, 596 (Utah 1984) (noting “the fundamental objective in a partition action is to divide the property so as to be fair and equitable and confer no unfair advantage on any of the cotenants”); see also ZAJAC, supra note 88, at 69 (explaining that “at best, economic efficiency is a necessary but far from sufficient condition for an economically just or fair economy”).

232 See ZAJAC, supra note 88, at 76.

233 See Meurer, supra note 13, at 940 (writing that economists insist “there is no single fair method of division”).

Corporate Corruption & the New Gold Mine

HOW THE DODD-FRANK ACT OVERINCENTIVIZES WHISTLEBLOWING

INTRODUCTION

On September 15, 2008, Lehman Brothers Holdings Inc. filed for relief under Chapter 11 of the Bankruptcy Code. It had debts of $613 billion against total assets of $639 billion, and its bankruptcy filing stands as the largest in the United States. On the following day, the U.S. government seized control of American International Group (AIG), one of the world’s largest insurers, in an $85 billion deal that “signaled the intensity of its concerns about the danger a collapse could pose to the financial system.” Federal Reserve Vice Chairman Donald Kohn stated that the failure of AIG posed “unacceptably large” risks to “consumers, municipalities, small business . . . as well as the risks to the wider economy,” and in essence, provided a “too big to fail” rationale to support the federal government’s $85 billion bailout. To some, this troubled financial state represented the “worst economy since the Great Depression.”

---

This economic crisis led to widespread support for changes in the financial regulatory system. President Barack Obama, when commenting on twenty-first-century financial regulatory reform, stated, “It is indisputable that one of the most significant contributors to our economic downturn was . . . the lack of adequate regulatory structures to prevent abuse and excess.” In response to the lack of such regulatory structures, Democratic Representative Barney Frank and Senate Banking Committee Chairman Chris Dodd proposed a financial regulatory overhaul. President Obama signed the resulting Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank Act) on July 21, 2010.

The Dodd-Frank Act faced fierce resistance and nearly unanimous Republican opposition. Critics of the Dodd-Frank Act claim it is a “radical expansion of the federal government that will hurt small businesses, community banks, and everyday taxpayers” and that the “new and expanded regulations . . . will limit the ability of banks . . . to extend credit.” Others argue that the Dodd-Frank Act is not strong

---


7 Barack Obama, President, United States, Remarks by the President on Regulatory Reform (June 17, 2009) (transcript available at http://www.whitehouse.gov/the_press_office/Remarks-of-the-President-on-Regulatory-Reform/) [hereinafter Obama, Remarks].


enough to revive the nation’s economic health. Proponents, however, praise the Dodd-Frank Act as legislation that “marks the end of more than a generation in which the prevailing posture of Washington toward the financial industry was largely one of hands-off cheering, evidenced by steady deregulation,” and hail the Act as a “clear turning point, highlighting ... [a] renewed reliance on government to protect the little guy.”

Despite the cloud of political controversy surrounding the Dodd-Frank Act, President Obama has firmly held that this overhaul of the financial regulatory system “place[s] rules that will allow our markets to promote innovation while discouraging abuse.” In an effort to help discourage such abuse, the Dodd-Frank Act includes a new whistleblower protection provision and also amends preexisting provisions to provide significant monetary incentives to potential whistleblowers. The new whistleblower provision drastically expands preexisting whistleblower legislation, primarily by expanding the scope of persons prohibited from taking retaliatory action and the scope of persons protected from retaliatory action. This note will examine two aspects of the Dodd-Frank Act’s whistleblower provision: its whistleblower bounty program and its antiretaliation statute. It argues that the Dodd-Frank Act’s whistleblower provisions overincentivize whistleblowing by providing excessive and unnecessary bounties, and by granting expansive whistleblower protection to too large a scope of individuals. In effect, the Dodd-Frank Act’s whistleblower laws transform corporate corruption into a “gold mine” by giving individuals the opportunity to reap enormous benefits from reporting alleged violations.

13 See Gretchen Morgenson, Strong Enough for Tough Stains?, N.Y. TIMES, June 26, 2010, at BU1, available at http://www.nytimes.com/2010/06/27/business/27gret.html (stating that the Dodd-Frank Act “fails completely” in curbing “dangerous risk taking by institutions and cutting big and interconnected financial entities down to size,” and that “the nation’s financial industry will still be dominated by a handful of institutions that are too large, too interconnected and too politically powerful to be allowed to go bankrupt if they make unwise decisions or make huge wrong-way bets”).
14 Ewing, supra note 6.
15 Obama, Remarks, supra note 7.
17 The “gold mine” metaphor incorporated into this note’s title was inspired by Letter from the Association of Corporate Counsel, to Elizabeth M. Murphy, Sec’y, Sec. & Exch. Comm’n (Dec. 15, 2010) (on file with author) (stating that as a result of the Dodd-Frank Act’s whistleblower provisions, “[f]raudulent misconduct, the bane of good compliance systems, then becomes the gold mine”).
Part I will discuss the preexisting whistleblower protection and bounty provisions. Part II will examine the Dodd-Frank Act’s amendments to the preexisting whistleblower laws, as well as its own antiretaliation statute. Part III will discuss the Dodd-Frank Act’s whistleblower bounty program and argue that it is likely to be ineffective in furthering the Act’s goals of encouraging individuals to report tips of a “higher quality.” Part IV will analyze the Dodd-Frank Act’s antiretaliation provision in light of judicial interpretation of similar provisions contained in the Sarbanes-Oxley Act, and argue that Dodd-Frank’s dramatic expansion of the scope of antiretaliation laws has damaging effects. Finally, Part V will discuss suggestions and recommendations for the future.

I. PREEXISTING WHISTLEBLOWER PROTECTION AND BOUNTY PROVISIONS

Before the Dodd-Frank Act was enacted, several statutes already protected whistleblowers “who report[] illegal or wrongful activities of his employer or fellow employees” and provided for rewarding such informants with bounties. Such provisions include the Sarbanes-Oxley Act of 2002, the False Claims Act, the Internal Revenue Code’s whistleblower provision, and the Insider Trading and Securities Enforcement Act of 1988. These statutes, however, only protect whistleblowers in narrow circumstances and provide weaker bounty incentives, both of which were imprudently expanded by Dodd-Frank.

A. The Sarbanes-Oxley Act of 2002

Under the Sarbanes-Oxley Act of 2002 (SOX), employees who provide information or assist in an investigation regarding conduct that they reasonably believe is a violation of any rule of the Securities Exchange Commission (SEC) or “any provision of federal law relating to fraud against shareholders” are

\[18\] Jessica Holzer & Fawn Johnson, Larger Bounties Spur Surge in Fraud Tips, WALL ST. J., Sept. 7, 2010, at C3 (quoting SEC official Stephen Cohen when writing that “[t]he goal is not just to get more tips, we want to get more high-quality tips”).  
\[23\] Id.  
protected from discharge, demotion, suspension, harassment, or any other form of discrimination by their employers.\textsuperscript{25} A whistleblower alleging an injury from a discriminatory or retaliatory act must make a prima facie showing that his whistleblowing behavior was a “contributing factor” to such act.\textsuperscript{26} If the whistleblower-employee prevails in the action, he is entitled to compensatory relief, including back pay with interest, reinstatement to the same seniority status the employee would have had but for the employee's act of discrimination, and compensation for any special damages incurred, including reasonable attorneys’ fees.\textsuperscript{27}

Although SOX protects whistleblowers from retaliation, it applies only when the whistleblower provides information to, or the investigation is conducted by, “a federal regulatory enforcement agency or law,” a congressional committee or member, or an individual “with supervisory authority over the employee.”\textsuperscript{28} SOX is further limited to apply only to publicly traded companies.\textsuperscript{29}

**B. The False Claims Act**

The False Claims Act (FCA) similarly protects whistleblowers from retaliatory action. It prohibits an employer from discharging, demoting, suspending, harassing, or otherwise discriminating against the whistleblowing

\textsuperscript{25} Id. § 1514A(a).
\textsuperscript{27} See, e.g., Ray v. Henderson, 217 F.3d 1234, 1242-43 (9th Cir. 2000) (noting the persuasiveness of the EEOC’s guidance that “any adverse treatment that is based on a retaliatory motive and is reasonably likely to deter the charging party or others from engaging in protected activity” is prohibited (quoting EEOC Compliance Manual Section 8, “Retaliation,” ¶ 8008 (1998))); Collins v. Beazer Homes USA, Inc., 334 F. Supp. 2d 1365, 1379 (N.D. Ga. 2004) (ruling that a two-week span between the whistleblower’s complaints of alleged violations and her termination was sufficient to establish circumstances suggesting that the protected activity “was a contributing factor to the unfavorable personnel action” for summary judgment purposes).
\textsuperscript{28} 18 U.S.C. § 1514A(c)(2)(A)-(C). Once a party has established its entitlement to an award of attorneys’ fees, the court must determine what is reasonable. Van Asdale v. Int’l Game, Tech., No. 3:04-CV-703-RAM, 2010 U.S. Dist. LEXIS 46725, at *21 (D. Nev. Apr. 13, 2010). Under federal law, reasonable attorneys’ fees are generally calculated using the “lodestar” method based on the number of hours the attorney worked. Id.
\textsuperscript{29} 18 U.S.C. § 1514A(a)(1)(A)-(C).
\textsuperscript{29} The provision prohibits any company with a class of securities “registered under section 12 of the Securities Exchange Act of 1934, or that is required to file reports under section 15(d) of the Securities Exchange Act of 1934 including any subsidiary or affiliate whose financial information is included in the consolidated financial statements of such company” from retaliating against an employee. Id. § 1514A(a).
employee. Relief for the injured whistleblower includes reinstatement to the same seniority status that the employee would have had but for the discrimination, double the amount of back pay plus interest, and compensation for any special damages sustained, which also includes reasonable attorneys’ fees.

In addition to its antiretaliation provision, the FCA offers a monetary reward to whistleblowers. The whistleblower, however, must initiate the action himself. If the government (through the Department of Justice) elects to proceed with the complaint, the whistleblower is entitled to 15 to 25 percent of the action’s proceeds or settlement. The ultimate amount of the reward depends on the extent to which the whistleblower “substantially contributed to the prosecution of the action.” If the government declines to pursue an action initiated by the whistleblower, then the whistleblower will have the right to conduct it himself and be eligible for a reward the court deems reasonable (between 25 and 30 percent of the proceeds of the action). In making this determination, the court must consider the whistleblower’s role in advancing the case to litigation, as well as the significance of the information he furnishes. Under the FCA, the whistleblower

---

30 An employee, contractor, or agent is entitled to “all relief to make [him] whole[,]” if he “is discharged, demoted, suspended . . . or in any other manner discriminated against in the terms and conditions of employment . . . because of [his] lawful acts . . . in furtherance of an action under this section . . . .” 31 U.S.C. § 3730(h) (2006).

31 Id.

32 Id. § 3730(d).

33 Id. § 3730(b)(1).

34 Id. § 3730(d)(1).

35 Id.

36 Within 60 days after the whistleblower files his report, the Government will either proceed with the action, in which case the action is to be conducted by the Government, or notify the court that it declines to take over the action, in which case the person bringing the action shall have the right to conduct the action. Id. § 3730(b)(4).

37 Id. § 3730(d)(2). Some courts have interpreted the False Claims Act’s whistleblower bounty provision to provide the minimum amount of recovery for individuals who “substantially and independently contribute to the government’s recovery.” See e.g., United States ex rel. Burr v. Blue Cross & Blue Shield, 882 F. Supp. 166, 168 (M.D. Fla. 1995). If, however, a whistleblower suffers “considerable personal and professional expense,” he is entitled to recover the full thirty percent. Id. The courts have further held that the “maximum recovery is reserved for situations where the [whistleblower] actively and uniquely aids the government in the prosecution of the case.” Id. Other courts believe that an “important factor in determining whether a [whistleblower] has actively and uniquely aided the government is if the matter proceeds all the way through trial.” United States ex rel. Fox v. Nw. Nephrology Assocs., 87 F. Supp. 2d 1103, 1112 (E.D. Wash. 2000). Such courts have held that “[i]t should be a rare occurrence that the maximum percentage is awarded in a case that has settled short of trial.” Id.

may also be rewarded up to 10 percent of the action’s proceeds even if the court determines that the action was based primarily on information that has already been made public.\footnote{Id.}

C. Internal Revenue Code’s Whistleblower Provision

The Internal Revenue Code’s whistleblower provision also provides bounties to whistleblowers reporting tax violations. The provision states that a whistleblower who furnishes information regarding an employer’s underpayment of taxes or violations of tax laws\footnote{If the Secretary of the Treasury proceeds with an administrative or judicial action for the underpayment of taxes or a violation of the internal revenue laws based on information brought by a whistleblower, the whistleblower is entitled to receive a monetary reward. 26 U.S.C. § 7623(b)(1) (2006).} may receive an award of 15 to 30 percent of the collected proceeds.\footnote{The whistleblower shall “receive as an award at least 15 percent but not more than 30 percent of the collected proceeds . . . resulting from the action (including any related actions) or from any settlement in response to such action.” Id. § 7623(b)(1).} The amount of the whistleblower’s reward will be determined in light of the extent to which he “substantially contributed to such action.”\footnote{Id.} If the information provided by the whistleblower is based principally on public information,\footnote{Id. § 7623(b)(2)(A).} the whistleblower may receive an award of up to 10 percent of the collected proceeds or settlement.\footnote{Id.} This determination must take into account the significance of the whistleblower’s role in contributing to the action.\footnote{Id.} In this respect, the Internal Revenue Code’s whistleblower incentive system is similar to that of the False Claims Act.\footnote{Under the False Claims Act, the degree to which the whistleblower contributed to the prosecution of the action is a factor that determines the amount of the reward. 31 U.S.C. § 3730(d)(1) (2006).} The Internal Revenue Code’s whistleblower provisions, however, are limited in scope and do not apply to illegal actions by individual taxpayers whose gross annual income is not more than $200,000 or where the amounts in dispute do not exceed $2 million.\footnote{26 U.S.C. § 7623(b)(5).} In addition, unlike under the
False Claims Act, the payment of the whistleblower reward here is discretionary, not mandatory.\footnote{Courts have held that the whistleblower reward provision in 26 U.S.C § 7623(b)(1) is not “money-mandating.” Wilson v. United States, No. 07-191T, 2007 U.S. Claims LEXIS 268, at *5-6 (Fed. Cl. July 13, 2007). Instead, the provision gives the Internal Revenue Service the “broad discretion to decide whether to make an award or how much to grant.” Merrick v. United States, 846 F.2d 725, 726 (Fed. Cir. 1988); see also Schmidt v. IRS, No. 08-10037, 2008 U.S. Dist. LEXIS 39376, at *3-4 (E.D. Mich. May 15, 2008) (concluding that an award under the Internal Revenue Code’s whistleblower provision is discretionary unless there have been negotiations with the whistleblower, and the whistleblower entered into an explicit agreement on the amount of the award with the Internal Revenue Service); Conner v. United States, No. 06-655C, 2007 U.S. Claims LEXIS 104, at *2 (Fed. Cl. Mar. 28, 2007) (holding that the statute only gives the Internal Revenue Status “broad discretion” to determine whether to make an award); Destefano v. United States, 52 Fed. Cl. 291, 293 (2002) (stating that the Internal Revenue Code’s whistleblower reward provision is a “discretionary statute” that does “not mandate monetary rewards and consequently do[es] not create a substantive right to money damages”); Krug v. United States, 41 Fed. Cl. 96, 97 (1998) (ruling that the Internal Revenue Service is not obligated to reward informants).}

D. Insider Trading and Securities Fraud Enforcement Act of 1988

The Insider Trading and Securities Fraud Enforcement Act of 1988 (Insider Trading Act) added a whistleblower bounty provision to the Securities Exchange Act of 1934\footnote{Insider Trading Act and Securities Fraud Enforcement Act of 1988, Pub. L. No. 100-704, 102 Stat. 4677.} that applies only to insider trading.\footnote{Id. § 21A, 102 Stat. at 4677-78.} The Insider Trading Act prohibits any person from purchasing or selling a security while in possession of material and nonpublic information\footnote{Id.} and mandates an award to whistleblowers of up to 10 percent of the proceeds of an action brought for any such violation they helped expose.\footnote{“[T]here shall be paid from amounts imposed as a penalty…such sums, not to exceed 10 percent of such amounts, as the [SEC] deems appropriate, to the person or persons who provide information leading to the imposition of such penalty.” Id. § 21A(e), 102 Stat. at 4679.}

These preexisting whistleblower statutes protect whistleblowers from employer retaliation, and even provide monetary rewards to incent whistleblowing. These provisions, however, protect and reward whistleblowers only in limited circumstances. For instance, SOX’s whistleblower protection provision applies only to publicly traded companies,\footnote{The provision prohibits any company with a class of securities “registered under section 12 of the Securities Exchange Act of 1934, or that is required to file reports under section 15(d) of the Securities Exchange Act of 1934” from retaliating against an employee. 18 U.S.C. § 1514A(a) (2006).} and the
Internal Revenue Code’s whistleblower bounty program applies to an individual only if the disputed amount exceeds $2 million.\(^{54}\) Further, such rewards are not mandatory.\(^{55}\) Although bounties under the False Claims Act are mandated, the whistleblower himself must initiate the action,\(^{56}\) thereby saving the government from administrative expenses. The Dodd-Frank Act’s whistleblower provisions, however, largely eliminate these limitations and drastically expand the applicability of whistleblower laws.

II. WHISTLEBLOWER INCENTIVES AND PROTECTION UNDER THE DODD-FRANK ACT

The Dodd-Frank Act greatly expands preexisting whistleblower bounty and antiretaliation provisions. It amends the SEC’s whistleblower provision by expanding its scope to cases other than those involving insider trading.\(^{57}\) The newly amended whistleblower provision now applies to any violation of securities laws and prohibits employers from discharging, demoting, suspending, threatening, or otherwise discriminating against a whistleblower.\(^{58}\) The provision also expands the definition of a whistleblower, which is now defined as an individual, or two or more individuals acting jointly, who provide information “relating to a violation of the securities laws.”\(^{59}\)

The Dodd-Frank Act also incentivizes whistleblowing by mandating a monetary reward to informants who provide “original information” regarding illegal activity.\(^{60}\) The provision

\(^{55}\) See supra note 48.
\(^{58}\) Id. § 78u-6(h)(1)(A). The provision reads:

No employer may discharge, demote, suspend, threaten, harass, directly or indirectly, or in any other manner discriminate against, a whistleblower in the terms and conditions of employment because of any lawful act done by the whistleblower (i) in providing information to the [SEC] in accordance with this section; (ii) in initiating, testifying in, or assisting in any investigation or judicial or administrative action of the [SEC] based upon or related to such information; or (iii) in making disclosures that are required or protected under the Sarbanes-Oxley Act of 2002 (15 U.S.C. 7201 et seq.), the Securities Exchange Act of 1934 (15 U.S.C. 78a et seq.), including section 10A(m) of such Act (15 U.S.C. 78f(m)), section 1513(e) of title 18, United States Code, and any other law, rule, or regulation subject to the jurisdiction of the [SEC].

Id.

\(^{59}\) Id. § 78u-6(a)(6).
\(^{60}\) Id. § 78u-6(b)(1).
defines "original information" as information that is derived from the independent knowledge or analysis of a whistleblower; is not known to the Commission from any other source, unless the whistleblower is the original source of the information; and is not exclusively derived from an allegation made in a judicial or administrative hearing, in a governmental report, hearing, audit or investigation, or from the news media, unless the whistleblower is a source of the information.

Further, the Act expands the type of person that can be a whistleblower. For instance, a whistleblower is not required to be an employee to receive the reward. In order to be eligible for a reward, however, the information furnished by the whistleblower must result in sanctions exceeding $1 million. Thus, if the whistleblower qualifies for the reward, he is guaranteed to receive a payment of at least 10 percent of $1 million, or $100,000. Additionally, the Dodd-Frank Act inserts a nearly identical whistleblower bounty provision into the Commodity Exchange Act.

Section 922 of the Dodd-Frank Act also expands the types of companies these provisions apply to by amending SOX's whistleblower provision. Unamended, SOX's antiretaliation statute applied only to publicly traded companies. Under the Dodd-Frank Act amendments, however, now any "subsidiary or affiliate whose financial information is included in the consolidated financial statements of such company" is within the purview of SOX's whistleblower protection provision.

Although the Dodd-Frank Act makes significant changes in preexisting whistleblower statutes, section 1057 of the Act most notably creates a broad private right of action for employees in the financial services industry who are retaliated
against for disclosing information regarding a violation of the Dodd-Frank Act, or any other provision that is subject to the Bureau of Consumer Financial Protection (Bureau).  

Under section 1057, “covered employees” are protected from retaliation if they engage in certain protected activities. A covered employee is any individual who performs “tasks related to the offering or provision of a consumer financial product or service,” or an authorized representative of such an individual. An employer (or any affiliate of such employer if the affiliate is a service provider) engaged in the offering or provision of a “consumer financial product or service” cannot discriminate against the whistleblower-employee, or cause the whistleblower-employee to be discriminated against. The complainant has the burden of making a prima facie showing that his disclosure of information was a contributing factor to the alleged retaliatory action. But if the employer demonstrates by clear and convincing evidence that he would have taken the same unfavorable personnel action even in the absence of that behavior, then the complainant is not entitled to any relief. If it is ultimately determined that the whistleblower was the victim of discrimination in violation of section 1057, he is entitled to reinstatement to his former position, as well as compensatory damages.

---

68 Id. § 1057, 124 Stat. at 2031. Section 1011 of the Dodd-Frank Act establishes an independent executive agency called the “Bureau of Consumer Financial Protection.” The Bureau is to “regulate the offering and provision of consumer financial products or services under the Federal consumer financial laws.” Id. § 1011(a), 124 Stat. at 1964.

69 The Dodd-Frank Act prohibits an employer’s retaliatory action against “any covered employee or any authorized representative of covered employees.” Id. § 1057(a), 124 Stat. at 2031. A whistleblower’s protected activity includes providing information to the employer or any government authority relating to any violation of any law that is subject to the Bureau’s jurisdiction, as well as “testify[ing] . . . in any proceeding resulting from the . . . enforcement of any . . . law that is subject to the [Bureau’s jurisdiction] . . . or object[ing] . . . to . . . any activity . . . that the employee . . . reasonably believed to be a violation of any law subject to the [Bureau’s jurisdiction].” Id. § 1057(a)(1)-(4), 124 Stat. at 2031-32.

70 Id. § 1057(b), 124 Stat. at 2032.

71 The Dodd-Frank Act prohibits a covered person or service provider from discriminating against “any covered employee or any authorized representative of covered employees” for reporting a violation. Id. § 1057(a), 124 Stat. at 2031.

72 Id. §§ 1002(6), 1057(a), 124 Stat. at 1961, 2031.

73 Id. § 1057(c)(3)(A), 124 Stat. at 2033.

74 Id. § 1057(c)(3)(C), 124 Stat. at 2033. Note that these standards of proof are essentially identical to those governing whistleblower retaliation claims brought under the Sarbanes-Oxley Act. See supra Part I.A.

75 Dodd-Frank Act § 1057(c)(4)(B), 124 Stat. at 2033.
Section 1057 of the Dodd-Frank Act as well as the Act’s amendments to preexisting whistleblower laws dramatically expands the scope of liability under antiretaliation provisions. The SEC’s whistleblower program now extends to cases other than insider trading, and whistleblowers engaged in the consumer financial product or service industries are offered expansive protection from retaliation or discrimination. The Dodd-Frank Act also creates generous bounty programs offering large monetary rewards to qualifying whistleblowers in efforts to encourage individuals to report violations. However, the effectiveness of the Dodd-Frank Act’s whistleblower provisions is doubtful, and likely overincentivizes whistleblowing.

III. The Ineffectiveness of the Dodd-Frank Act’s Whistleblower Bounty Program

Despite legislators’ good intentions, the Dodd-Frank Act’s bounty program overincentivizes whistleblowing and will waste administrative resources because it provides what studies show are unnecessarily excessive awards. Although the bounty program was enacted to encourage whistleblowing, the monetary rewards are likely unnecessary in advancing the provision’s purported goals. Further, the bounty provisions lack significant threshold considerations and fail to discourage the submission of frivolous claims. As a result, whistleblowers are overincentivized.

A. Goals of the Dodd-Frank Act’s Whistleblower Bounty Program

The legislative history of the Dodd-Frank Act reveals Congress’s intent “to motivate those with inside knowledge to come forward and assist the Government to identify and prosecute persons who have violated securities laws and recover money for victims of financial fraud.” Congress emphasized the importance of bounty programs because whistleblowers “often face the difficult choice between telling the truth and the risk of committing ‘career suicide.’”

---

77 See infra Part IV.
78 The program “[r]ecognizes that whistleblowers often face the difficult choice between telling the truth and the risk of committing ‘career suicide.’” S. REP. NO. 111-176, at 110 (2010).
79 Id.
80 Id.
Markopolos, Certified Fraud Examiner and Madoff whistleblower, attested to the efficiency and effectiveness of whistleblower bounty programs. He cited statistics holding that whistleblower tips uncovered 54 percent of fraud schemes in public companies, as opposed to the 4 percent that external auditors—including the SEC—exposed.\(^8\) SEC Chairman Mary L. Schapiro further noted that “[w]histleblowers can be a source of valuable firsthand information that may otherwise not come to light”\(^8\) and that “[t]hese high-quality leads can be crucial to protecting investors and recovering ill-gotten gains from wrongdoers.”\(^8\) Congress has made it abundantly clear that one of the primary purposes of the Dodd-Frank Act’s whistleblower bounty program is to encourage the reporting of “high-quality” tips,\(^8\) and the $1 million requirement that whistleblowers must meet in order to qualify for a reward evidences legislators’ efforts to encourage the disclosure of major violations. Although it may seem logical that providing monetary rewards incentivizes whistleblowers to report illegal activity, studies show that this assumption is not entirely true. In fact, in light of recent research, the Dodd-Frank Act’s whistleblower bounty program is counterproductive.

B. Research Indicating the Ineffectiveness of the Dodd-Frank Act’s Whistleblower Bounty Program

A recent study demonstrates the ineffectiveness of the Dodd-Frank Act’s bounty program. The study examined the role incentives play in whistleblowers’ decisions to report illegal activity.\(^8\) It concluded that in cases where the whistleblower has a “greater ethical stake in the outcome” monetary incentives might be unnecessary and counterproductive because they may offset the whistleblower’s internal ethical motivation.\(^8\) Conversely, when the perceived severity of the misconduct is

\(^8\) Id.
\(^8\) Holzer & Johnson, supra note 18.
\(^8\) Id. at 1207.
low, “external incentives,” such as monetary rewards, “matter much more” to the whistleblower’s decision to report the illegality. In other words, in cases where the activity at issue has significant ethical and moral implications, research suggests that the whistleblower does not need monetary incentives to compel him to report the violation. On the other hand, situations involving less severe conduct may require financial rewards to encourage the whistleblower to report the misconduct. These results are “contrary to the basic intuition of the legal policy maker to give higher rewards as the misconduct is more severe.” Given its legislative intent to encourage the reporting of major violations, the Dodd-Frank Act’s whistleblower bounty program is likely to be counterproductive, as it offers large monetary rewards, which studies indicate are unnecessary for cases involving significant moral implications. Other studies show that “extrinsic motivators do not alter the attitudes that underlie our behaviors. They do not create an enduring commitment to any value or action.” Thus, it is unlikely that the Dodd-Frank Act’s whistleblower bounty program will be successful in furthering legislators’ goals. The Dodd-Frank Act also lacks significant thresholds that the whistleblower must overcome, which is likely to result in a surge of reports of trivial claims.

1. The Ineffectiveness of the Dodd-Frank Act’s Large Monetary Rewards

The general perception regarding the relationship between the amount of sanctions imposed and the severity of misconduct is that the greater the amount of sanctions, the more severe the misconduct. Thus, through the Dodd-Frank Act’s imposition of the $1 million minimum, the legislature appears to intend to reward whistleblowers only in cases where the severity of misconduct and moral offensiveness is high. The study generally found that “[i]n areas where the misconduct is expected to trigger high internal motivation, there is less need

---

87 Id. at 1194.
88 Id. at 1204.
89 See id.
91 Feldman & Lobel, supra note 85, at 1204 (stating that “legal policy maker[s] give higher rewards as the misconduct is more severe (given its likely correlation with greater harm to society)”).
The Dodd-Frank Act’s bounty program appears to contradict this research, as it guarantees a qualifying whistleblower a minimum reward of $100,000 for a tip regarding a violation worthy of sanctions of at least $1 million. Consequently, the Dodd-Frank Act’s bounty program is excessive, as it offers monetary incentives that studies show are not necessary. The large rewards offered to whistleblowers overincentivize whistleblowing, and may ultimately prove to be a waste of limited agency resources. This is not to suggest that all whistleblower bounty programs should be eliminated. Due to its inherent risks, whistleblowing, to some extent, should be incentivized through regulatory policies that “encourage individuals to break the code of silence in corrupt organizations.” Eliminating all whistleblower bounty provisions and instead implementing a legal duty to report, along with a fine for a failure to report, is unlikely to be effective in encouraging whistleblowing, as there is a “growing body of studies both in social psychology and in behavioral economics indicating that people respond more strongly to incentives than penalties.” However, monetary incentives may not be necessary or effective in all situations, and thus it is imperative that legislators accurately assess and determine the optimal level of rewards to ensure the effective application of limited agency resources. Although the Dodd-Frank Act’s whistleblower bounty program attempts to achieve this ideal level of whistleblower incentives, it fails to do so. By mandating large rewards to whistleblowers, the Dodd-Frank Act’s whistleblower provision invites a flood of whistleblower reports, which may not necessarily be of the “high-quality”

---

92 Id.
93 Under section 21F of the Securities Exchange Act, a whistleblower is to receive a minimum of ten percent of the sanctions recovered, which must be at least $1,000,000, resulting in a minimum recovery of $100,000. 15 U.S.C.A. § 78u-6(a)-(b)(1) (West 2009 & Supp. 2011). The law further provides that a “covered judicial or administrative action” is an action that “results in monetary sanctions exceeding $1,000,000.” Id. § 78u-6(a)(1).
94 See Feldman & Lobel, supra note 85, at 1157-58 (“The decision of whether to blow the whistle is a complex one and inevitably involves certain risks... In addition to direct employment retaliation, reporting often entails psychological and societal costs, including fear, guilt and mistreatment by peers and community... One commentator has described whistle-blowing as ‘professional suicide.’” (quoting James Gobert & Maurice Punch, Whistleblowers, the Public Interest, and the Public Interest Disclosure Act 1998, 63 Mod. L. Rev. 25, 35 (2000))).
95 Id. at 1159.
96 Id. at 1181.
97 See id. at 1182.
legislators intended to encourage.\textsuperscript{98} Although some whistleblowers may be discouraged from bringing such claims forward because of fear of retaliation, in light of the expansive protection granted by the Dodd-Frank Act's whistleblower protection provisions, this ultimately may not prove to be a significant disincentive.\textsuperscript{99}

2. The Dodd-Frank Act's Lack of Significant Threshold Considerations

The excessive rewards provided by the Dodd-Frank Act could be justifiable if there were meaningful thresholds the whistleblower had to overcome in order to receive his reward. Threshold considerations involve what the whistleblower must first do and what the result of the disclosures must be in order to qualify for a reward.\textsuperscript{100} For example, the FCA contains significant threshold considerations\textsuperscript{101}—namely, the whistleblower must initiate the litigation himself.\textsuperscript{102} As a result of this prerequisite, whistleblowers may be discouraged from reporting frivolous claims.\textsuperscript{103} Such meaningful threshold considerations are absent in the Dodd-Frank Act's whistleblower bounty provisions.

For instance, the Dodd-Frank Act does not contain significant threshold considerations with respect to who can benefit from its provisions. In fact, the whistleblower does not even need to be an employee of the entity allegedly engaging in the illegal activity.\textsuperscript{104} Although the whistleblower must provide information that is not known to the agency from another

\textsuperscript{98} Holzer & Johnson, supra note 18.
\textsuperscript{99} See discussion supra Part II; see also Marsha J. Ferziger & Daniel G. Currell, Snitching for Dollars: The Economics and Public Policy of Federal Civil Bounty Programs, 1999 U. ILL. L. REV. 1141, 1173-74 (“Although an informant’s discounted losses can arise in many ways, the ‘big ticket’ potential harms are harm to her livelihood . . . . In this vein, discounted retaliation costs arise from a host of actions a defendant could take against the informant . . . .”). However, federal and state antiretaliation statutes protect the whistleblower, and “[a]lthough an informant may not find his former workplace a pleasant post-informing environment, these statutes can ensure that he remains gainfully employed or receives compensation for any harm suffered in the workplace, thus mitigating his discounted retaliation losses.” Id. at 1174. Such discounted losses may also include “reputational harms and the mental and emotional costs of testifying in litigation.” Id.
\textsuperscript{100} Id. at 1150.
\textsuperscript{101} Id.
\textsuperscript{102} Id.
\textsuperscript{103} Id. at 1159.
source, thus overcoming a threshold requirement concerning the type of information covered under the Act, this may not ultimately prove to be a difficult standard to overcome as whistleblowers can race to furnish the agencies with their inside information. Another arguable “threshold” contained in the Dodd-Frank Act is the $1 million requirement whistleblowers must satisfy in order to be eligible for a reward. Although this million-dollar minimum does provide some limitation on the whistleblower’s eligibility for the bounty, there is no provision preventing the whistleblower from reporting frivolous claims in hopes that the sanctions exceed $1 million. Additionally, whistleblowers do not need to demonstrate the veracity of their allegations. Under the finalized rules, the whistleblower submits his or her information regarding potential securities law violations under the penalty of perjury. However, “there is little to deter individuals from making unjustified accusations of wrongdoing,” and the possibility that a whistleblower will provide a knowingly false statement that could be the basis for prosecution of perjury is “remote at best.”

Thus, the Dodd-Frank Act’s whistleblower provisions lack significant threshold considerations and fail to discourage whistleblowers from reporting trivial claims. The Act also offers excessive monetary rewards, and as a result, overincentivizes whistleblowing.


106 A whistleblower is rewarded only in a “covered judicial or administrative action,” 15 U.S.C.A § 78u-6(b)(1), which is defined as an action that results in monetary sanctions exceeding $1 million. Dodd-Frank Act §§ 748(a)(1), 922(a)(1), 124 Stat. at 1841, 1842.

107 Bruce Carton, Pitfalls Emerge in Dodd-Frank Whistleblower Bounty Provision, SEC. DOCKET (Sept. 9, 2010), http://www.securitiesdocket.com/2010/09/09/pitfalls-emerge-in-dodd-frank-whistleblower-bounty-provision/ (noting that “[t]he millions that whistleblowers might potentially reap could also encourage a lottery mentality, where people file complaints on weak or wholly illegitimate claims ‘just in case.’”). However, there is a provision that penalizes the whistleblower for willfully furnishing false information. Dodd-Frank Act §§ 748(m), 922(i), 124 Stat. at 1746, 1847.


109 Letter from Jones Day, to Elizabeth M. Murphy, Sec’y, Sec. & Exch. Comm’n (Dec. 17, 2010) (on file with author) [hereinafter J ones Day Letter].

110 Under 18 U.S.C. § 1001(a) (2006), any person who “knowingly and willfully (1) falsifies, conceals, or covers up...a material fact; (2) makes any materially...fraudulent statement or representation; or (3) makes or uses any false writing or document knowing the same to contain any materially...fraudulent statement...will be fined and imprisoned for not more than 5 years.” Id.

111 Jones Day Letter, supra note 109.
C. Future Implications of the Dodd-Frank Act’s Whistleblower Bounty Program

The excessive rewards provided by the Dodd-Frank Act’s whistleblower bounty program, as well as its lack of significant threshold considerations, is likely to have several implications, including an increase in the number of whistleblower claims, a burden on administrative costs, and employees’ circumvention of companies’ internal compliance systems.

1. Substantial Increase in the Number of Whistleblower Claims

Expected bounty payments are one of the most important factors that influence whistleblowers’ decisions to disclose a violation.112 In fact, “[m]uch of the [whistleblower’s] uncertainty [is attributed to] the agency’s discretion to award a reduced bounty or no bounty at all.”113 Under the Dodd-Frank Act, a whistleblower’s uncertainty regarding payment is largely eliminated. Provided that his information results in sanctions exceeding $1 million, the whistleblower is certain that he will receive a reward of at least $100,000. But if monetary rewards are high, “every potential informant with a crumb of information might crawl out of the woodwork hoping to hit the bounty jackpot.”114 In fact, claims began “trickling in” shortly after the enactment of the Dodd-Frank Act.115 Although “no flood of tips” has occurred yet since the SEC finalized the whistleblower provisions,116 Sean McKessy, head of the Whistleblower Office, is expanding his staff117 and the SEC predicts it will receive approximately thirty-thousand tips per year.118 Further, it is likely that a SEC announcement of a large

112 Ferziger & Currell, supra note 99, at 1152.
113 Id.
114 Id.
117 See id.
118 Eaglesham & Jones, supra note 115.
reward arising from a Dodd-Frank whistleblower initiative will trigger an increased number of claims.\textsuperscript{119} Due to the Act’s lack of significant threshold considerations,\textsuperscript{120} whistleblowers may rush to report alleged violations without confirming that the allegations are valid in order to ensure that they are the original source of information. Ultimately, the unnecessary payments the whistleblower receives is likely to be ineffective in furthering the Dodd-Frank Act’s goal of soliciting “high quality” tips. Instead, the Dodd-Frank Act’s whistleblower bounty program overincentivizes whistleblowing, and is likely to lead to a surge of reports and a burden on administrative costs.

2. Burden on Administrative Costs

Whistleblower bounty programs may ostensibly lower the cost of obtaining vital information, but the Dodd-Frank Act will not ultimately save administrative resources.\textsuperscript{121} A reduction in administrative costs may result if the statute contained meaningful threshold considerations, such as in the False Claims Act, which permits private individuals to bring the suit on behalf of the government.\textsuperscript{122} However, the administration under the Dodd-Frank Act will have to “review—and occasionally litigate—a substantial number of claims that turn out to be grounded on poor information or information the [agency] or the public already possessed.”\textsuperscript{123} Unlike the False Claims Act, where the government does not conduct most of the litigation,\textsuperscript{124} the administration under the Dodd-Frank Act will have to devote further resources to investigate the alleged misconduct. Additionally, given the likelihood that the Dodd-Frank Act will result in a drastically increased amount of whistleblower reports, agencies will expend additional “time and resources when sorting good tips from the bad.”\textsuperscript{125} The

\textsuperscript{120} See supra Part III.B.2.
\textsuperscript{121} Ferziger & Currell, supra note 99, at 1158-59.
\textsuperscript{123} Ferziger & Currell, supra note 99, at 1159.
\textsuperscript{124} Id. “Of the first four hundred [False Claims Act] cases filed, the Department of Justice…joined as a litigant in only seventy; the others proceeded privately.” Id.
\textsuperscript{125} Id. at 1171. Aside from its practical consequences, implementing whistleblower incentive programs also has several ethical implications, and the morality of the practice of rewarding informants has been disputed. Id. at 1191. The
administrative cost of processing these voluminous reports may exceed the benefit gained from enticing a few whistleblowers holding excellent information on high-level crimes. The rules provide that whistleblowers who wish to participate in the whistleblower program must “declare, under penalty of perjury, that their submission is truthful to the best of their knowledge.” Thus, the SEC argues, the whistleblower rules sufficiently discourage frivolous claims. It reasons that “[t]his should reduce the costs incurred by the [SEC] from devoting resources to review and evaluate frivolous submissions, and also create efficiency gains by permitting the [SEC] to place greater reliance on the accuracy of information that is received.” But the rules lack threshold considerations that prevent whistleblowers from reporting claims in hopes that they will result in sanctions exceeding $1 million. As a result, while their submissions may not be perjurious, they may not be of the high quality sought by legislators when drafting the Dodd-Frank Act.

Although the Internal Revenue Code’s whistleblower bounty program was profitable before its 2006 amendment, it is important to note that the Internal Revenue Service (IRS) did not pay whistleblowers the maximum reward of 15 percent. In fact, in 1993, the IRS paid an average reward of 3 percent of resulting sanctions. Data further indicate that the IRS “not only pa[id] small percentages and small rewards but also pa[id] them to a small number of claimants.” Thus, it is general opinion is that informants should voluntarily come forward with information, rather than being “brib[ed]” to disclose any illegal activity. One U.S. District Court Judge noted, “I don’t think that turkeys like that ought to receive a dime of my money.” (quoting Judge Alcee Hastings). Others, however, do not find such whistleblower incentives as immoral. Some view such programs as “right and honorable,” and hold the viewpoint that whistleblowers are protecting the public. Although such moral issues present an interesting debate, it is beyond the scope of this note.
not unreasonable to posit that the Internal Revenue Code's previous profits are attributable, at least in part, to its small and infrequent payouts. Under the Dodd-Frank Act, however, agencies must reward qualifying whistleblowers at least 10 percent of the resulting sanctions. As a result, it is questionable whether Dodd-Frank's whistleblower program will turn a similar profit.

Interestingly, the IRS paid its first whistleblower reward in 2011, four years after the enactment of its whistleblower bounty program. Further, the IRS recently reported that it would delay payments under its pre-amendment program for up to two years. According to the Treasury Department's Inspector General for Tax Administration, the IRS's whistleblower program suffered from defects in the "control and timely resolution of whistleblower claims." In light of the shortcomings of preexisting whistleblower bounty programs, it is likely that the Dodd-Frank Act's bounty program will also be ineffective in furthering the Act's purported goals.

3. Eradication of Companies' Internal Compliance Systems

The Dodd-Frank Act's whistleblower program is also likely to undermine companies' established internal compliance systems. Rather than reporting alleged illegal activity to the company itself, potential whistleblowers may opt to forgo internal compliance methods and report the misconduct directly to the SEC. The National Association of Corporate Directors (NACD) recently argued to the SEC that the Dodd-Frank Act's whistleblower provision “encourage[s] employees to bypass their own [company's] compliance departments in their

---

136 In fact, the SEC has delayed the opening of its whistleblower office due to budgeting concerns. Eaglesham & Jones, supra note 115.
138 Donmoyer, supra note 137.
139 Eaglesham & Jones, supra note 115.
eagerness to inform the SEC of suspected foul play.\textsuperscript{140} It is inappropriate for such SEC investigations to be conducted at taxpayer expense,\textsuperscript{141} and instead, the company should be afforded an opportunity to remedy the alleged violation. Bypassing internal compliance essentially denies management the opportunity to take remedial action because of the whistleblower’s pursuit of a profit. As a result, the Dodd-Frank Act vitiates companies’ responsible efforts to create and implement effective compliance systems and reporting schemes.\textsuperscript{142}

In an attempt to mitigate the subversion of internal compliance programs, the SEC now offers a potentially larger reward if a whistleblower first utilizes the company’s compliance program before reporting the alleged violation to the SEC.\textsuperscript{143} The SEC does this in several ways. For instance, if a whistleblower first reports the alleged misconduct to the company’s internal compliance program and the company later investigates and reports the results of its investigation to the SEC, all the information provided to the SEC by the company will be attributed to the whistleblower,\textsuperscript{144} which can lead to an increased reward. Additionally, the rules expressly state that a whistleblower’s participation in his or her company’s internal compliance program is a factor that can increase his or her reward.\textsuperscript{145} Thus, this arguably remedies the issue of the eradication of companies’ internal compliance programs.\textsuperscript{146}

\textsuperscript{140} Crossman, supra note 17. The NACD also argued that the whistleblower program “provide[s] an incentive for persons having “independent knowledge” of possible corporate wrongdoing to report directly to the SEC,” and that “the legislators who enacted the original provisions in the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010 did not weigh the consequences the provisions could have on the ethical and compliance-based cultures of corporations.” Id. (quoting the NACD).

\textsuperscript{141} Id.

\textsuperscript{142} Eaglesham & Jones, supra note 115 (quoting Susan Hackett, Senior Vice President of the Association of Corporate Counsel, who stated that “[t]he proposals cut to the very core of what it is that every responsible U.S. company has been trying to do for the last couple of decades, which is to create effective, robust compliance reporting systems”).

\textsuperscript{143} Memorandum from Cleary Gottlieb Steen & Hamilton LLP 3 (May 26, 2011), available at http://www.cgsr.com/files/News/af83ed9a-fbdf-442e-8e7b-10f5c1bc0a54/ Presentation/NewsAttachment/b281f060-f24c-4224-b24b-23990d2e9a1CGSH%20Alert%20-%20SEC%20Approves%20New%20Whistleblower%20Program.pdf (“Despite controversy and numerous comments from the business community, the final rules do not require whistleblowers to avail themselves of internal compliance programs before reporting to the SEC. Instead, the [SEC] made several changes to its proposal to add incentives for whistleblowers to use internal compliance programs before or when going to the [SEC].”).

\textsuperscript{144} 17 C.F.R. § 240.21F-4(c) (2010).

\textsuperscript{145} Id. § 240.21F-6(a)(4).

\textsuperscript{146} The SEC states that although it did not require whistleblowers to report violations internally, it has “made additional changes to the rules to further incentivize
However, it is unclear whether increased monetary rewards will be effective in encouraging whistleblowers to first report alleged violations internally. First, as noted above,\textsuperscript{147} studies indicate that large monetary rewards offset internal motivations, and may prove to be unnecessary and counterproductive.\textsuperscript{148} Further, in light of the court’s interpretation in Egan v. Tradingscreen, Inc.,\textsuperscript{149} whistleblowers are likely to bypass the internal compliance system despite the increased reward.

Egan is the first reported decision under the Dodd-Frank Act’s whistleblower protection provision,\textsuperscript{150} and concerns a “major issue addressed as part of the Dodd-Frank whistleblower rulemaking proceedings: the integrity of corporate internal compliance and reporting programs.”\textsuperscript{151} In Egan, the whistleblower-employee reported suspected fraudulent activity to the company’s internal compliance system.\textsuperscript{152} Despite assurances that he would not be fired, he was later terminated and thus filed suit, alleging a violation of the Dodd-Frank Act’s antiretaliation provision.\textsuperscript{153} The employer argued that the antiretaliation provisions did not apply because the whistleblower-employee did not directly report the alleged violation to the SEC.\textsuperscript{154} The court interpreted the whistleblower protection provisions as follows:

Plaintiff must either allege that his information was reported to the [SEC], or that his disclosures fell under the four categories of disclosure delineated by 15 U.S.C. § 78u-6(h)(1)(A)(iii) that do not require such reporting: those under the Sarbanes-Oxley Act, the Securities Exchange Act, 18 U.S.C. § 1513(e), or other laws and regulations subject to the jurisdiction of the [SEC].\textsuperscript{155}

The Egan Court reasoned that the legislature could have easily provided broader protection for whistleblowers to utilize their companies’ internal compliance and reporting systems when appropriate.” SEC Final Rules, supra note 108, at 5.

\textsuperscript{147} See supra Part III.B.
\textsuperscript{148} See Feldman & Label, supra note 85, at 1154-55.
\textsuperscript{149} No. 10 Civ. 8202 (LBS), 2011 U.S. Dist. LEXIS 47713, at *1-2 (S.D.N.Y. May 4, 2011).
\textsuperscript{150} Letter from Stephen M. Kohn, Exec. Dir., Nat'l Whistleblowers Ctr., to Elizabeth Murphy, Sec'y, Sec. & Exch. Comm'n, and David A. Stawick, Sec'y, Commodity Futures Trading Comm'n, at 1 (May 17, 2011) (on file with author) [hereinafter National Whistleblowers Center Letter].
\textsuperscript{151} Id.
\textsuperscript{153} Id. at *5-6.
\textsuperscript{154} Id. at *9.
\textsuperscript{155} Id. at *13-14.
alleging securities laws and thus the “absence of similarly broad protections...indicates that Congress intended to encourage whistleblowers reporting such violations to report to the SEC.” 156 As such, the court’s interpretation can be read to mean that “employees have no choice but to bypass internal reporting systems and directly raise concerns regarding violations of securities laws with federal regulatory agencies and the Justice Department.” 157 In light of the court’s interpretation, the possibility of an increased monetary reward for using internal compliance systems does not outweigh whistleblowers’ concern of retaliation, and thus they are likely to report directly to the SEC. In fact, some organizations assert that they “will do everything in their power to ensure that employees bypass such channels....[as it] would be the height of irresponsibility for whistleblower advocates to urge employees to use internal reporting programs.” 158 As a result, the important policy objectives identified by the SEC would be “seriously undermined.” 159 The Egan Court further noted that “[o]bviously, a whistleblower must directly report to the SEC to receive a bounty award from the SEC.” 160 Consequently, the requirement that a whistleblower must report the alleged violation to the SEC in order to receive the reward, along with the confusion about whether whistleblowers must report to the SEC in order to receive protection under the Dodd-Frank Act’s antiretaliation provisions, is likely to lead to the eradication of internal compliance programs. Thus, the SEC’s attempt to encourage whistleblowers to first use companies’ internal compliance systems by offering larger rewards is ineffective.

The Dodd-Frank Act’s excessive bounties and lack of significant threshold considerations invite a flood of whistleblower reports. Supporters may argue that such monetary incentives are necessary, as whistleblowers are unwilling to report violations for fear of retaliation or discrimination by their employer. But the Dodd-Frank Act also offers expansive protection under its antiretaliation statute, largely reducing such fears. As a result, the Dodd-Frank Act’s unnecessary and excessive bounty program, coupled with its expansive antiretaliation protection,
overincentivizes whistleblowing and is likely to lead to the waste of administrative resources.

IV. THE DODD-FRANK ACT’S EXPANSION OF PREEXISTING WHISTLEBLOWER PROTECTION PROVISIONS

The Dodd-Frank Act’s antiretaliation provisions drastically expand preexisting laws in two respects: first, by expanding the scope of persons prohibited from taking retaliatory action, and second, by enlarging the scope of persons protected from retaliatory action.

A. The Expansion of the Class of Persons Prohibited from Taking Retaliatory Action

Unamended, SOX’s whistleblower protection provision applies only to certain publicly traded companies. In applying the statute’s plain meaning, courts have held that a narrow reading—applying the provisions only to public companies—is “necessary” to limit the scope of SOX’s antiretaliation protection. Here courts have argued that a contrary holding might have the effect of extending the statute “far beyond” what Congress envisioned. Additionally, courts have held that to subject nonpublic subsidiaries of publicly traded parent companies to SOX’s whistleblower statute would “widen the scope of the whistleblower protection provisions beyond what Congress appears to have intended.” Congress, however, has effectuated such intent through the Dodd-Frank Act. The Dodd-Frank Act amended SOX’s antiretaliation provision, which now provides that “any subsidiary or affiliate whose financial information is included in the consolidated financial

---

161 18 U.S.C. § 1514A(a) (2006); see also Brady v. Calyon Sec. (USA), 406 F. Supp. 2d 307, 317 (S.D.N.Y. 2005) (holding that an employee of nonpublicly traded companies was not covered by SOX and that a “specific requirement . . . is that defendant be a publicly traded company”); Flake v. New World Pasta Co., 03-126, Final Decision and Order (Dep’t of Labor Feb. 25, 2004) (concluding that pursuant to the plain language of the Sarbanes-Oxley Act, even when an employer is a publicly traded company, it is not covered if it is not registered under section 12 or required to file reports under section 15(d) of the Securities and Exchange Act of 1934).


163 Id. (quoting Goodman v. Decisive Analytics Corp., 2006 SOX 11 (Dep’t of Labor Jan. 10, 2006)).

statements of a publicly traded company is prohibited from taking discriminatory or retaliatory action against a whistleblower-employee. Under this amendment, the scope of SOX’s antiretaliation provision has been significantly expanded, and now prohibits non-publicly traded subsidiaries or affiliates of publicly traded companies from taking discriminatory action against whistleblowers. Arguably, the scope of persons prohibited from taking retaliatory action under SOX is still restricted, as the statute applies only to allegations of violations of securities regulations. In this respect, the scope of SOX’s whistleblower provision may remain limited. It is unlikely, however, that this “limitation” will ease employers’ concerns, since individuals hoping to blow the whistle on violations of regulations other than securities laws can claim protection under a new antiretaliation statute provided in section 1057 of the Dodd-Frank Act, which implements a very broad antiretaliation provision.

The amended antiretaliation provision protects whistleblowers from retaliation against “any person that engages in offering or providing a consumer financial product or service,” rather than to a narrow band of certain publicly traded companies. The provision also broadly defines “financial products or services” and includes appraisers, check cashers, and lenders. Affiliates of financial service providers are also subject
to SOX’s expanded laws, provided that the affiliate acts as a “service provider,” or someone who furnishes a “material service” to the financial service provider in connection with the offering of a financial product or service. As a result, under the Dodd-Frank Act’s amendment to SOX’s preexisting whistleblower provisions, persons who provide or offer financial products or services, and even those who may indirectly offer or provide financial services, are subject to antiretaliation provisions.

This expansion of liability under the Dodd-Frank Act is startling. Application of the preexisting statutes was limited to cases of insider trading, tax evasion, fraudulent transactions with the government, and those involving publicly traded companies. Under section 1057, however, persons offering or providing financial products or services to consumers, and even those who provide a “material” service to such providers of financial products or services, are subject to antiretaliation provisions.

The Bureau’s ability to insert additional examples and definitions of “financial product or service” further evidences the broad scope of the individuals and entities subject to section 1057. If the Bureau concludes that a financial product or service is executed with the purpose of evading any federal consumer financial law, or is one that a bank or financial holding company is permitted to offer and is likely to have a material impact on consumers, the Bureau is explicitly authorized to insert that product or service into the Dodd-Frank Act’s definition of “financial product or service.” The scope of persons subject to section 1057’s prohibition against retaliatory action is noninclusive and can be amended to include other persons in the future, further expanding its scope. The expanded protection that is now provided to whistleblowers leads to increased employer liability, which can have numerous negative effects like increased costs.

---

170 Id. An affiliate is a person that “controls, is controlled by, or is under common control with another person” of a person engaged in the offering or provision of a consumer financial product or service. Id. § 5481(1), (6)(A).

171 Id. § 5481(26). The Dodd-Frank Act explicitly provides that the term “service provider” is not to include a person who offers or provides to a covered person a support service that is generally provided to businesses or a “similar ministerial service,” id. § 5481(26)(B)(i), nor a person who provides “time or space for an advertisement for a consumer financial product or service through print, newspaper, or electronic media.” Id. § 5481(26)(B)(ii). Note that if the service provider offers or profits from its own consumer financial product or service, it is deemed to be a covered person. Id. § 5481(26)(C).

172 Id. § 5481(15)(A)(xi).

173 Id.

174 See infra Part IV.C.
B. The Expansion of the Class of Persons Protected from Retaliatory Action

The Dodd-Frank Act imprudently expands the class of whistleblowers from the previously narrowly defined constraints of the preexisting antiretaliation provisions to the current definition protecting any whistleblower who performs tasks “related” to “the offering or provision of a consumer financial product or service.”\textsuperscript{175} The use of the broad, undefined term “related” greatly increases the scope of antiretaliation protection beyond what is appropriate to achieve congressional goals.

In addition, unlike SOX’s antiretaliation provision, section 1057 is not limited to whistleblowers alleging certain types of violations.\textsuperscript{176} Instead, legislators set section 1057’s parameters broadly, and the statute explicitly applies to whistleblowers who allege violations of any law subject to the jurisdiction of the Bureau.\textsuperscript{177} The Bureau’s jurisdiction extends to the regulation of the offering and provision of consumer financial products or services under the federal consumer financial laws, as well as all federal laws concerning public or federal contracts, property, works, employees, budgets, or funds.\textsuperscript{178} Protection from retaliation is no longer limited to whistleblowers claiming violations of securities laws, tax codes, or cases of insider trading. Rather, section 1057 significantly expands the class of individuals protected under antiretaliation laws.\textsuperscript{179} Due to the greater protection afforded to them, whistleblowers now have additional incentives to blow the whistle. The Dodd-Frank Act also increases employers’ liability

\textsuperscript{175} 12 U.S.C.A. § 5567(b).
\textsuperscript{176} Section 1057 also protects authorized representatives of covered employees. Id. § 5567(a). This is similar to SOX’s whistleblower provision, which protects “employees” from retaliation. 18 U.S.C.A. § 1514A(a) (2006). Accordingly, an “employee” includes a “company representative,” which is defined as “any officer, employee, contractor, subcontractor, or agent of a company.” 29 C.F.R. § 1980.101 (2010).
\textsuperscript{178} Id. § 1011(a), 124 Stat. at 1964.
\textsuperscript{179} Sections 748 and 922 of the Dodd-Frank Act amend the Commodity Exchange Act and Securities and Exchange Act of 1934, respectively. The provisions provide that no employer may discriminate or retaliate against an employee because of the employee’s furnishing of information to either the Commodity Futures Trading Commission or the SEC regarding the employer’s misconduct. Id. §§ 748(h)(1)(A), 922(h)(1)(A), 124 Stat. at 1744, 1845. These provisions include additional classes of individuals who are protected from retaliatory and discriminatory action, further expanding the scope of whistleblower protection laws. See 15 U.S.C.A. § 78u-6(h) (West 2009 & Supp. 2011).
to such provisions. This expansion of liability, however, is likely to have damaging effects.

C. The Negative Effects of Employers' Expanded Liability Under Dodd-Frank's Whistleblower Protection Statutes

It is undeniable that antiretaliation laws justly protect whistleblowers who report legitimate claims of employers' violations. The positive effects of whistleblower protection laws, however, do not justify the Dodd-Frank Act's dramatic expansion of liability under such provisions. Employers' increased liability under the Dodd-Frank Act is likely to lead to several damaging consequences. For instance, given the direct costs associated with expanded corporate liability, such as indemnity and defense costs, an increase in liability may "affect the economy by influencing the behavior of individual corporations." It is likely that such expenses will increase as a result of employers' greater exposure to liability, thereby impacting management decisions, including the cost-benefit analyses associated with such decisions.

Another cost that may result from the Dodd-Frank Act's whistleblower provisions is the "negative effect on organizational culture." Management shapes the ethical nature of an organization, and by undermining management's internal compliance efforts, the whistleblower provisions are "harming the organizational culture." Further, "as organizational culture affects organizational performance, Dodd-Frank is harming the bottom line." Other costs include the damage to organizational reputation and a decrease in shareholder wealth caused by SEC actions. Additionally, employers may implement more stringent employment policies in response to their increased liability. Although employers may refrain from unjustly terminating competent employees as

---

181 Id. at vi.
182 Id. at v.
184 Id.
185 Id.
186 Id. at 139-40.
187 Reuter, supra note 180, at vi.
a result of the employment policies,\textsuperscript{188} they may be hesitant to terminate incompetent employees for fear of accusations of violating antiretaliation laws.\textsuperscript{189} Consequently, there may be an increase in the number of employees retained who in fact should be terminated,\textsuperscript{190} and employers are likely to become “hampered in their ability to adjust their employment levels expeditiously.”\textsuperscript{191} Moreover, each new employee represents a potential liability in the event of their termination,\textsuperscript{192} and thus employers may incur greater expenditures for recruitment efforts.\textsuperscript{193} The fear of wrongful termination actions can prevent managers from “being as flexible in their response to changing market conditions, risky investment opportunities, or technology advances.”\textsuperscript{194} In effect, expanded exposure to liability may prevent firms from making short-run adjustments in light of business fluctuations.\textsuperscript{195} As a result, companies may increasingly rely on overtime contractors or temporary agencies during transition periods rather than permanent employees.\textsuperscript{196}

The expanded protection granted to whistleblowers in itself is arguably sufficient to encourage employees to report violations, as it eliminates potential whistleblowers’ fears of retaliation. But coupled with the excessive bounties offered to whistleblowers under the Dodd-Frank Act, whistleblowers are overincentivized, leading to more whistleblower reports\textsuperscript{197} and burdening both employers and administrative agencies.

V. SUGGESTED APPROACH FOR THE FUTURE

The ideal whistleblower provision is one that “maximize[s] a potential informant’s discounted reward and minimize[s] his discounted losses without making the mix so attractive as to induce the disclosure of large amounts of bad information.”\textsuperscript{198} The Dodd-Frank Act’s whistleblower provisions,
however, fail to achieve this ideal balance. These provisions overincentivize whistleblowing through their exorbitant bounty payouts and expansive protection against retaliation. Whistleblowers need not overcome significant threshold considerations to be eligible for the rewards, and these employees are protected by the liberal interpretation of antiretaliation provisions. Thus, risk is comparably low for whistleblowers, but the potential reward is high. In such circumstances, “most informants will come forward with even a low level of certainty” and inundate government agencies with trivial reports. In order to provide a more efficient and effective means of policing employers through the use of whistleblowers, Congress should ensure that the Dodd-Frank Act’s whistleblower provision includes meaningful thresholds that the whistleblower must overcome, lower rewards for the whistleblower to recover, and appropriate penalties if the whistleblower intentionally delays reporting a violation in order to increase the amount of his reward.

A. Implement Significant Threshold Considerations

In efforts to discourage whistleblowers from reporting trivial claims, the Dodd-Frank Act should encourage whistleblowers to report only those violations involving a substantial amount of money. Although the Dodd-Frank Act attempts to achieve this by requiring that the whistleblower’s information lead to sanctions resulting in at least $1 million, this requirement alone is unlikely to prevent whistleblowers from bringing forth insignificant claims. First, the rules now provide that, “for the purposes of making an award, [the SEC] will aggregate two or more smaller actions that arise from the same nucleus of operative facts.” This is likely to lead to a “just in case” mentality because a whistleblower may think that the amount will be reduced or not awarded at all.”

199 See supra Part III.
200 See supra Part IV.
201 Ferziger & Currell, supra note 99, at 1180.
202 “[I]f a bounty program is to avoid” incentivizing citizens to “snitch on their neighbors for insignificant transgressions and . . . the administrative costs attendant to sifting through such bogus claims, it should encourage . . . informants to come forward with information only when (1) the violation is factually and legally clear, and (2) it involves a substantial quantity of money.”
203 See supra note 63.
204 SEC Final Rules, supra note 108, at 6-7.
that another person is also reporting a violation that arises out
of the "same nucleus of operative facts." Instead, the Dodd-
Frank Act should impose additional threshold requirements.

One meaningful threshold is a requirement that
whistleblowers demonstrate the reasonable likelihood that the
alleged violation will result in sanctions of at least $1 million.
This requirement is not likely to be particularly cumbersome to
the potential whistleblower, as an informant is

able to cheaply discover the maximum and minimum awards for
which he is potentially eligible, and, unlike an agency, he may know
approximately how much is at stake in the potential litigation.

Based on this nonpublic information, an informant should be able to
estimate with some accuracy whether the government is likely to
recover a penalty from the defendant.\footnote{Ferziger & Currell, supra note 99, at 1183-84.}

This requirement would discourage whistleblowers from
reporting trivial claims "just in case" they lead to sanctions of
over $1 million, thereby reducing administrative costs.\footnote{Commentators advocated for including a standard of reasonableness in the
definition of a "whistleblower." Some recommended that an individual have a
"reasonable" or "good faith belief" that the information he or she possesses relates to a
securities law violation. See Jones Day Letter, supra note 109, at 2-3; Letter from
Morgan, Lewis & Bockius LLP, to Elizabeth M. Murphy, Sec'y, Sec. & Exch. Comm'n,
at 3 (Dec. 17, 2010) (on file with author) (suggesting a requirement for both a
subjective and objectively reasonable belief of a violation); Letter from Ronald C. Long,
Dir. Regulatory Affairs, Wells Fargo, to Elizabeth M. Murphy, Sec'y, Sec. & Exch.
Comm'n, at 3 (Dec. 17, 2010) (noting that defining whistleblowers as individuals who
provide information regarding "potential violations' without any threshold defining
criteria or good faith standard would only serve to frustrate the effective and efficient
administration of genuine whistleblower claims"). The SEC, however, commented that
"a higher standard requiring a 'probable' or 'likely' violation is unnecessary, and would
make it difficult for the staff to promptly assess whether to accord whistleblower status
to a submission." SEC Final Rules, supra note 108, at 13.}

\footnote{17 C.F.R. § 240.21F-2 (2010).}
violation. For instance, Jones Day presents a scenario where a company's Chief Financial Officer makes optimistic statements about the company's prospects. At the end of the financial quarter, however, the company reports disappointing results and the company's stock falls. An employee of the company then blows the whistle and alleges that prior to making the optimistic statements, the Chief Financial Officer made statements during an internal meeting suggesting the company faced "significant challenges." Jones Day notes that the contrast between the negative internal statements and the more optimistic statements may suggest a potential violation of section 10(b) of the Securities Exchange Act of 1934. The whistleblower-employee, however, knows the full context of the internal statements and in what respects the company faced such challenges. Thus, the full context of the internal statements makes it apparent that the Chief Financial Officer has not engaged in any misrepresentation. In this scenario, the whistleblower-employee's allegation is "literally true, and it relates to a possible violation of the securities laws; but even if [the whistleblower-employee] acted in bad faith, she is absolutely protected in her employment due to the [SEC's] expansion of the definition of a 'whistleblower' to include information relating to 'potential' violations." In this respect, the fact that the SEC requires whistleblowers to bring claims regarding "possible" violations will not likely serve as a meaningful threshold consideration.

An additional threshold that legislators should implement is a requirement that whistleblowers first report the alleged discrepancy to employers' internal compliance systems. Only if the issue is not resolved may the whistleblower resort to government assistance. For instance, a

208 Jones Day Letter, supra note 109, at 3.
209 Id.
210 Id. Section 10(b) of the Securities Exchange Act of 1934 provides that it is unlawful for any person, directly or indirectly, by the use of any means or instrumentality of interstate commerce or of the mails, or of any facility of any national securities exchange [t]o effect a short sale . . . [and] [t]o use or employ, in connection with the purchase or sale of any security registered on a national securities exchange or any security not so registered . . . any manipulative or deceptive device or contrivance in contravention of such rules and regulations as the [SEC] may prescribe as necessary or appropriate in the public interest or for the protection of investors.

211 Jones Day Letter, supra note 109, at 3.
provision similar to section 10A of the Securities Exchange Act would provide a more “logical whistleblowing framework.”

Under section 10A, auditors who believe they have discovered an illegal act at a company are required to report it first to company management and the audit committee. If the company fails to take remedial action, only then is the auditor required to report the violation to the SEC. As it stands now, the Dodd-Frank Act’s whistleblower provision is likely to “reverse a decade of effort promoting integrity, self-remediation, and corporate self-reporting.” Requiring whistleblowers to report first through companies’ internal compliance systems would provide senior management with the opportunity to remedy the alleged violation, and avoid employees’ circumvention of internal compliance programs. The finalized rules provide that a whistleblower’s reward may be increased if he or she utilizes the internal compliance system. However, as displayed in Egan v. Tradingscreen, Inc., this might not always be the most effective means of encouraging whistleblowers to use internal compliance programs. Under the Egan Court’s interpretation, the Dodd-Frank Act’s whistleblower provision requires whistleblowers to report to the SEC in order to receive both protection against retaliation and their monetary reward. As a result, it is

---

212 Carton, supra note 107 (quoting former SEC enforcement attorney, Jacob Frenkel, and providing that “absent egregious misconduct condoned (or even conducted) by senior management, employees have a responsibility to attempt to correct errors and misconduct through existing corporate compliance systems”).

213 15 U.S.C. § 78j-1(b)(1) (“If, in the course of conducting an audit . . . the registered public accounting firm detects . . . an illegal act . . . the firm shall . . . inform the appropriate level of the management of the issuer and assure that the audit committee of the issuer . . . is adequately informed with respect to illegal acts that have been detected.”).

214 Id. § 78j-1(b)(2). “If . . . the registered public accounting firm concludes that . . . the senior management has not taken . . . timely and appropriate remedial actions with respect to the illegal act . . . the registered public accounting firm shall, as soon as practicable, directly report its conclusions to the board of directors.” Id. Once the issuer’s board of directors receives the report, it must “inform the [SEC] by notice not later than 1 business day after the receipt of such report and shall furnish the registered public accounting firm making such report with a copy of the notice furnished to the [SEC].” Id. § 78j-1(b)(3).

215 Carton, supra note 107 (quoting former SEC enforcement attorney, Jacob Frenkel).

216 See supra Part III.C.3.

217 See 17 C.F.R. §§ 240.21F-4(c), 240.21F-6(a)(4) (2010).

218 See supra Part III.C.


220 Id. at *24 (“Obviously, a whistleblower must directly report to the SEC to receive a bounty award from the SEC . . . .”).
likely that whistleblowers will bypass internal compliance programs. The SEC states that “internal investigations can be an important component of corporate compliance...[but] providing information to persons conducting an internal investigation, or simply being contacted by them, may not, without more, achieve the statutory purpose of getting high-quality, original information about securities violations directly into the hands of [SEC] staff.” 221 However, Congress should implement a program that replicates the auditing system where the whistleblower first reports the violation internally, and receives protection from retaliation, and then reports it to the SEC if a certain period of time passes and no remedial action has been taken. Such a reporting system would avoid the eradication of companies’ internal compliance programs and mitigate the issue of threshold considerations,222 thereby encouraging only those who have nontrivial claims to come forward.

B. Reduce the Amount of the Mandated Bounties Awarded to Eligible Whistleblowers

Legislators should also reduce the amount of the mandated reward given to eligible whistleblowers. The imposition of high monetary rewards for violations involving high levels of moral outrage is unnecessary and overincentivizes whistleblowing.223 Instead, a low, fixed-percentage award provides the best route.224 Although a low bounty may affect potential whistleblowers’ decisions to report alleged violations, a lower reward is not likely to affect whistleblowers’ decisions to report cases of serious infractions.225 Instead, a low bounty’s most profound effect will likely be limited to whistleblowers’ decisions to expose cases involving “insignificant” fraud.226 Even if the bounties offered under the Dodd-Frank Act were limited to 5

221 SEC Final Rules, supra note 108, at 34.
223 See supra Part III.B.
224 Ferziger & Currell, supra note 99, at 1197. The authors also propose that the ideal whistleblower bounty provision “guarantee[s] that (1) the maximum allowable bounty will always be paid where an agency recovers a penalty based on an informant’s tip, and (2) the agency will make all possible efforts to maintain an informant’s anonymity within the constraints of the litigation process.” Id.
225 Id. at 1198.
226 Id.
percent of sanctions exceeding $1 million, this would guarantee a minimum payment of $50,000—hardly an insignificant amount.

C. Penalize Whistleblowers Who Intentionally Delay Reporting the Alleged Violation

Another concern implicated by the Dodd-Frank Act's whistleblower provision is the fear that whistleblowers may intentionally delay reporting violations to increase the amount of the resulting sanctions, thereby increasing the amount of their reward. To discourage such behavior, lawmakers should penalize whistleblowers who unreasonably delay reporting the illegality. Research has indicated that penalties are not always effective in inducing action.\(^{227}\) Thus, if such a penalty were imposed, it should be implemented only in limited circumstances where the employer demonstrates that the whistleblower unreasonably delayed reporting the violation with the clear intent to increase the amount of his reward or to increase his chances of satisfying the Dodd-Frank Act's $1 million minimum. The use of a penalty in limited circumstances may prevent whistleblowers' fraudulent behavior while minimizing the negative sociological effects of penalties.

Arguably, whistleblowers are penalized under Dodd-Frank if they attempt to undermine the integrity of internal compliance systems. The SEC provides that a whistleblower's attempts to undermine the company's internal compliance program can decrease the amount of an award.\(^{228}\) However, the mere diminution of an award will not be sufficient to deter people from acting in this manner. The Dodd-Frank Act whistleblower provisions were implemented in hopes to "promote effective enforcement of federal securities laws by providing incentives for persons with knowledge of misconduct to come forward and share their information with the [SEC]."\(^{229}\) If, however, a whistleblower takes "any steps to undermine the integrity of...[employers' internal compliance] systems or processes,"\(^{230}\) his or her reward is merely reduced rather than eliminated, and this contradicts the goals of the provisions. Instead, elimination of the reward is a better option.

\(^{227}\) Feldman & Lobel, supra note 85, at 1181.

\(^{228}\) SEC Final Rules, supra note 108, at 5; 17 C.F.R. § 240.21F-6(b)(3) (2010).

\(^{229}\) SEC Final Rules, supra note 108, at 34.

\(^{230}\) Id.
CONCLUSION

Drafters of the Dodd-Frank Act attempted to cure the defects that many alleged were the contributing factors to the 2008 financial crisis, including regulatory failure. Commentators, however, have noted that “reform is premature when the exact nature and causes of the financial crisis are yet to be determined.” Legislators, without fully knowing the extent and precise causes of the financial crisis, implemented provisions calling for comprehensive regulatory reform and created a broad whistleblower provision within the Dodd-Frank Act. As a result, the Dodd-Frank Act fails to provide an effective whistleblower program, and instead overincentivizes whistleblowing through its expansive whistleblower protection and excessive bounties. Instead, legislators should carefully examine the incentives and consequences of whistleblowing and draft a provision that reflects those findings accordingly.

Jenny Lee†

† B.A., English, Villanova University, 2009; J.D. Candidate, Brooklyn Law School, 2012. I wish to thank my classmates and colleagues on the Brooklyn Law Review for their incredible dedication and effort. I would also like to extend a special thank you to my family and friends for their unconditional love and support, without which none of my accomplishments would have been possible.
An Unconstitutional Response to Citizens United

MANDATING SHAREHOLDER APPROVAL OF CORPORATE POLITICAL EXPENDITURES

In accepting corporate money, I promise to respect federal election laws the same way I respect the must-shower-before-swimming law at the Y. As a candidate, I am under no obligation to promote the zesty, robust taste of Doritos brand tortilla chips regardless of how great a snack they may be for lunchtime, munch time, anytime.

—Stephen Colbert

INTRODUCTION

Mr. Colbert, the host of a satirical political news show, The Colbert Report, took some significant steps toward running for President of the United States in the 2008 election. Although Mr. Colbert’s candidacy was sardonic in nature, the real steps he took toward procuring a spot on the ballot in South Carolina brought the potential of sanction by the Federal Election Commission (FEC). Doritos, manufactured by PepsiCo, Inc. (PepsiCo), sponsored Mr. Colbert on Comedy Central at the time he launched his purported candidacy in 2007. A former FEC general counsel stated that PepsiCo’s sponsorship of Mr. Colbert’s show, and ostensibly of his

4 Klein, supra note 1.
5 Jones, supra note 2, at 299. Professor Jones notes that “[t]here was apparently no actual sponsorship and Pepsi[Co], maker of Nacho Cheese Doritos, paid no consideration for being featured in the campaign.” Id. at 299 n.17 (citing Richard L. Hasen, Stephen Colbert’s “Hail to the Cheese” Presidential Candidacy: Why the Comedian’s Campaign Raises Serious Questions About the Role of Corporate Money in Elections, FINDLAW.COM (Nov. 9, 2007), http://writ.lp.findlaw.com/commentary/20071109_hasen.html).
candidacy, raised “serious questions.” One question was whether PepsiCo’s sponsorship violated federal campaign-finance laws that—at the time—barred corporations from directly contributing to candidates or making independent political expenditures from their corporate treasuries. Just a few years later, the legal landscape regulating corporate political activity has changed drastically.

When the Supreme Court decided Citizens United v. FEC on January 21, 2010, it caused a major shift in campaign-finance law. In that seminal decision, the Court held a ban on independent corporate political expenditures to be unconstitutional under the First Amendment. Specifically, the Court held that Congress may not ban political speech on the basis of “the corporate identity of the speaker and the content of the political speech.” As a result of the five-to-four decision, corporations may now legally use their general treasuries to fund political advertising or make other independent political expenditures.

Within this new paradigm, PepsiCo could seemingly pay for an advertisement that featured Mr. Colbert’s candidacy.

---

6 Klein, supra note 1 (quoting Lawrence M. Noble, counsel at Skadden, Arps, Slate, Meagher & Flom LLP); see also] ones, supra note 2, at 307-09.
8 130 S. Ct. 876 (2010). This note takes no position as to the wisdom of the majority opinion in Citizens United. Rather, the focus of this note is to evaluate legislative proposals in light of the decision.
9 See Michael S. Kang, After Citizens United, 44 IND. L. REV. 243, 243 (2010) (“Citizens United . . . represents the Roberts Court’s clear reversal of [the trend of expansion of campaign finance regulation] and a narrow focus on quid pro quo corruption as the exclusive grounds for government regulation.”); Molly J. Walker Wilson, Too Much of a Good Thing: Campaign Speech After Citizens United, 31 CARDOZO L. REV. 2365, 2392 (2010) (“Unfortunately, the Citizens United decision does more than to give corporate interests a place at the table. It gives them a place at the head of the table and a bullhorn.”); Adam Liptak, Justices, 5-4, Reject Corporate Spending Limit, N.Y. TIMES, Jan. 22, 2010, at A1 (Citizens United was a “sharp doctrinal shift”). Although Citizens United essentially extended the same theoretical framework announced in FEC v. Wisconsin Right to Life, Inc., 551 U.S. 449 (2007), the practical implications of the Citizens United decision were extremely broad.
10 2 U.S.C. § 441b; Citizens United, 130 S. Ct. at 897 (“Section 441b makes it a felony for all corporations—including nonprofit advocacy corporations—either to expressly advocate the election or defeat of candidates or to broadcast electioneering communications within [thirty] days of a primary election and [sixty] days of a general election.”).
11 Citizens United, 130 S. Ct. at 917.
(should he run in 2012) and perhaps also feature Doritos-brand tortilla chips, so long as PepsiCo made the expenditure independently and not in coordination with Mr. Colbert. If PepsiCo were to fund such an advertisement, its shareholders could take issue with corporate spending on advertising that featured Mr. Colbert’s candidacy. Surely, some shareholders would not want PepsiCo to wade into politics and others might oppose Mr. Colbert’s candidacy, whether satirical or not. In Citizens United, Justice Kennedy, writing for the majority, addressed the interests of shareholders by stating that “the procedures of corporate democracy” could deal with such concerns.

14 Dean Erwin Chemerinsky has suggested that the Supreme Court will go even further in extending political speech rights to corporations in the near future and “hold that corporations have the right to contribute money to candidates for elective office...” Erwin Chemerinsky, The Future of the First Amendment, 46 WILLAMETTE L. REV. 623, 638 (2010). If Dean Chemerinsky’s prediction comes true, then the direct corporate sponsorship of a political campaign would have constitutional protection. In a recent decision, a district court held the ban on companies’ direct contributions to federal candidates to be unconstitutional. United States v. Danielczyk, No. 1:11cr85 (JCC), 2011 WL 2161794, at *19 (E.D. Va. May 26, 2011) (invalidating 2 U.S.C. § 441b(a)); see also Nathan Koppel, Judge Rules Ban on Corporate Campaign Contributions Unconstitutional, WALL ST. J. L. BLOG (May 27, 2011, 2:44 PM), http://blogs.wsj.com/law/2011/05/27/judge-rules-ban-on-corporate-campaign-contributions-unconstitutional/. However, Professor Hasen has written that he expects the Danielczyk decision to be reversed on appeal. Rick Hasen, Federal District Court, in Criminal Case, Holds that Ban on Direct Corporate Contributions to Candidates Is Unconstitutional Under Citizens United, ELECTION L. BLOG (May 26, 2011, 9:58 PM), http://electionlawblog.org/?p=18342.

15 See Justin J. Wert, Ronald Kether Gaddie & Charles S. Bullock, III, Of Benedick and Beatrice: Citizens United and the Reign of the Laggard Court, 20 CORNELL J. L. & PUB. POL’Y 719, 726-27 (2011) (“Taking controversial and highly visible political stands can potentially cost clients and therefore lead to financial costs. Corporate stocks and corporate products have been punished by consumers for overt political activity... A rise in overt, direct political action by most corporations carries with it risks far exceeding the political gains that might be achieved by acting through other agents.” (footnotes omitted)); see also Matthew A. Melone, Citizens United and Corporate Political Speech: Did the Supreme Court Enhance Political Discourse or Invite Corruption?, 60 DEPAUL L. REV. 29, 95-96 (2010) (“Corporations are constrained by the potential reaction of customers, employees, shareholders, public interest groups, and non-governmental organizations to open advocacy. Unseemly corporate campaigning may result in loss of customers, employee dissatisfaction, or shareholder agitation in the form of proxy fights.” (footnotes omitted)).

16 Citizens United, 130 S. Ct. at 911 (quoting First Nat’l Bank of Bos. v. Belotti, 435 U.S. 765, 794 (1978)); but see Ciara Torres-Spelliscy, Corporate Campaigning, FORBES.COM (Jan. 29, 2010, 10:00 AM), http://www.forbes.com/2010/01/28/corporate-campaign-politics-legal-opinions-contributors-ciara-torres-spelliscy.html. This note does not challenge the constitutionality of shareholders su sponte restricting their corporation’s spending on political endeavors. However, one commentator has pointed out that while “a shareholder could simply dissociate himself from a corporation should its expressed political ideals conflict with his own,” the shareholder might already have been harmed “by the time a shareholder first learns his political views conflict with those disseminated by the corporation.” Alex Osterlind, Note, Giving a Voice to the Inanimate The Right of a Corporation to Political Free Speech, 76 Mo. L. REV. 259, 281 (2011); see also Elizabeth Pollman, Citizens Not United: The Lack of Stockholder
In the wake of the Citizens United decision, academics,\(^{17}\) think tanks,\(^{18}\) and legislators\(^{19}\) have put forth many proposals in an attempt to counteract the increased potential for corporate political involvement that has resulted from the Supreme Court’s “robust conception of corporate political speech.”\(^{20}\) Although shareholder-driven governance will certainly be permitted to direct corporations in this arena as the shareholders see fit, some members of Congress and think tanks have proposed to make this voluntary procedure mandatory.\(^{21}\) Three prominent proposals would make significant changes to the federal securities laws and would mandate shareholder approval of a corporation’s political expenditures,\(^{22}\) similar to the current campaign-finance law in

---


Some commentators have argued that the fears that these proposals seek to address are “overblown, or at the very least misstated” because, in part, “the corporate … moneys that are opened up for use by the Citizens United decision do not exist in a vacuum, sitting in a massive Scrooge McDuck money vault. These moneys, whether from profits or member dues, also have other obligations … .”

See infra Part III.
the United Kingdom.\textsuperscript{23} For example, the Brennan Center for Justice urges Congress to amend the federal securities laws by “mandating that corporations obtain the consent of shareholders before making political expenditures.”\textsuperscript{24}

Lively and divisive political debate has long centered on the interpretation of the First Amendment in the context of political campaign finance. Because the outcome of this debate will have an unquantifiable impact on the future elections of our national leaders—and will dictate upon whom those leaders may rely for financial support—many factions pull in different directions in an attempt to shape the ongoing interpretation of the First Amendment in the corporate-political-speech context.\textsuperscript{25} The proponents of the proposals that mandate shareholder approval of political expenditures urge an unrealistically narrow interpretation of Citizens United. Although their authors have noble goals, the proposals are unconstitutional under Citizens United and other First Amendment jurisprudence.\textsuperscript{26}


\textsuperscript{24} Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 21; see also Torres-Spelliscy, Corporate Political Spending, supra note 23, (manuscript at 59). For a similar proposal, see Ronald Gilson & Michael Klausner, Corporations Can Now Fund Politicians. What Should Investors Do?, FORBES, Mar. 29, 2010, available at 2010 WLNR 5423725 (“The answer is to mandate that corporations let stockholders vote annually on whether they want the company to exercise the rights that Citizens United gave them to get into political races. Managers who seek stockholder approval of political activity would explain the actions they intend to take, how those actions would be in stockholders’ interests and what the cost will be.”).

\textsuperscript{25} Scholars and experts agree that the Citizens United decision has significantly impacted the legal and political landscape, but its effects have yet to be fully felt or understood. E.g., Christopher Beane, Ad Hominem: How Much Has the Citizens United Case Changed Campaign Finance in 2010?, SLATE (Oct. 6, 2010, 6:09 PM), http://www.slate.com/id/2270036/ (“No doubt Citizens United set back the cause of campaign finance reform. But the jury is still out on its practical effects.”).

\textsuperscript{26} See WHITAKER, supra note 13, at 8 (“The practicalities of when to require [shareholder] votes . . . may need to be carefully considered in order not to run afoul of corporate freedom of speech rights defined by the Supreme Court in Citizens United.”). The proposals could only be found constitutional if the Supreme Court overruled the very case and doctrine that the proposals seek to “remedy.” The proposals, while well-intentioned and logical, also raise several policy and practical issues that would make their implementation difficult or counterproductive. This note will not address these issues due to space limitations.
Part I of this note provides a brief historical overview of campaign-finance law as applied to corporations before _Citizens United_. Part II discusses the _Citizens United_ decision and its effects on the legal and political landscape regarding corporate contributions and expenditures. Part III provides an outline of three proposals that would require mandatory shareholder approval of independent political expenditures by corporations. Part IV offers a critique of the proposals through an analysis under _Citizens United_ and other First Amendment jurisprudence. Finally, this note concludes that mandatory shareholder approval of political expenditures by corporations is not a constitutional response to the potential problems created by _Citizens United_.

I. **PRE-CITIZENS UNITED LANDSCAPE**

Although _Citizens United_ is undoubtedly the new leading case in the area of campaign-finance law,\(^27\) several earlier cases remain good law and will likely bear upon the Court’s analysis of the mandatory shareholder-approval proposals. Since _Citizens United_ overruled two cases, proper consideration of the current doctrine requires an understanding of the historical doctrine.

The Supreme Court first recognized the political speech rights of corporations in _First National Bank of Boston v. Bellotti_,\(^28\) where it struck down a state statute that banned corporate political expenditures on referenda unrelated to that corporation’s proprietary interests.\(^29\) The Court found that, for the purposes of independent expenditures on referenda, corporations had the same free speech rights as individuals.\(^30\)

With regard to analysis in future cases, the Court stated that “extending constitutional guarantees to a corporation depends upon the nature, history, and purpose of the particular constitutional provision.”\(^31\) The Court also noted that free

---

\(^27\) See Breanne Gilpatrick, Removing Corporate Campaign Finance Restrictions in _Citizens United v. Federal Election Commission_, 130 S. Ct. 876 (2010), 34 HAV. J.L. & PUB. POL’Y 405, 419 (2011) (“The effect of _Citizens United_ already is being felt in federal courtrooms around the country as judges and lawyers come to see the decision as a dramatic shift in how the Court views campaign finance laws.”).

\(^28\) _First Nat’l Bank of Bos. v. Bellotti_, 435 U.S. 765, 778 n.14 (1978) (“In cases where corporate speech has been denied the shelter of the First Amendment, there is no suggestion that the reason was because a corporation rather than an individual or association was involved.” (citations omitted)).

\(^29\) Id. at 779, 791-92; see also Matthew Lambert, Beyond Corporate Speech: Corporate Powers in a Federalist System, 37 RUTGERS L. REC. 20, 21 (2010).

\(^30\) _Bellotti_, 435 U.S. at 780 & n.15.

\(^31\) Lambert, supra note 29, at 22 (quoting Bellotti, 435 U.S. at 778 n.14). Importantly, _Citizens United_ viewed _Bellotti_ as one of the bases for its jurisprudential
speech is “indispensable to decisionmaking in a democracy, and this is no less true because the speech comes from a corporation.”

Eight years later, the Supreme Court took a more equivocal and complicated approach to corporate political expenditures in FEC v. Massachusetts Citizens for Life, Inc. (MCFL). Unlike Bellotti, which involved a for-profit corporation that sought to influence votes for a referendum, MCFL resolved a First Amendment challenge from a nonprofit corporation that sought to influence an election for political office. In MCFL, the Court held that the government could not ban ideological nonprofits, which do not take corporate or union contributions, from spending their treasury funds on explicit political advocacy because no risk of dissenting shareholders existed.

The next major case to bear upon the issue of the political rights of corporations was Austin v. Michigan Chamber of Commerce. In Austin, the plaintiffs challenged a
state law that banned corporations from using funds from their general treasuries to make independent expenditures on behalf of, or in opposition to, state electoral candidates. The Austin Court upheld the state law on the basis of a fear that corporations could distort the political process because “the unique state-conferred corporate structure . . . facilitates the amassing of large treasuries . . . [that] can unfairly influence elections when . . . deployed in the form of independent expenditures.” After Austin, corporations had no constitutional right to make independent campaign expenditures out of their general treasuries, but they could still expend treasury funds on issue advocacy that mentioned specific candidates. For example, corporations could not fund advertisements that urged citizens to vote for or against a candidate in a federal election, but corporations could fund advertisements that urged constituents to contact a specific legislator regarding a particular cause to express support or disapproval.

36 Austin, 494 U.S. at 654.
37 Id. at 660. Professor Hasen notes that although the Court explains this opinion with “an anti-corruption rationale, Austin’s emphasis on preventing distortion of the electoral process through large corporate spending suggested the Court in fact was espousing an equality rationale . . . .” Hasen, supra note 32, at 588 (internal quotation marks and citations omitted). It has also been noted that with its decision in Austin, “the Court moved beyond [a] focus on quid pro quo corruption to embrace broader theories” such as “a different type of corruption: the corrosive and distorting effects of immense aggregations of wealth that are accumulated with the help of the corporate form and that have little or no correlation to the public’s support for the corporation’s political ideas.” Abraham, supra note 34, at 1086 (quoting Austin, 494 U.S. at 659-60); see also Wilson, supra note 9, at 2371-72.
39 Excellent examples of these so-called “issue advocacy” advertisements from the 1996 presidential campaign abound. Although it is not a corporation, the AFL-CIO ran one advertisement titled “No Way” that included this voice-over:

[Woman]: My husband and I both work. And next year, we’ll have two children in college. And it will be very hard to put them through, even with the two incomes.

Announcer: Working families are struggling. But Congressman [X] voted with Newt Gingrich to cut college loans, while giving tax breaks to the wealthy. He even wants to eliminate the Department of Education. Congress will vote again on the budget. Tell Congressman [X], don’t write off our children’s future.

[Woman]: Tell him, his priorities are all wrong.

In 2002, Congress reacted to this line of cases in an attempt to increase restrictions on both corporate and private funding of campaigns and independent political expenditures. Congress enacted the Bipartisan Campaign Reform Act (BCRA) and altered the campaign-finance regulatory landscape.\(^4\) The BCRA, in relevant part, banned the use of corporate treasury funds to pay for “electioneering communications” during certain preelection time periods, such as thirty days before a primary election or sixty days before a general election.\(^4\) Professor Hasen has explained that “[e]lectioneering communications are television or radio (not print or Internet) advertisements that feature a candidate for a federal election; they are capable of reaching 50,000 people in the relevant electorate . . . .”\(^4\) In addition, the BCRA preserved the ability of corporations to establish political action committees (PACs) that could solicit and accept contributions from a limited class of individuals, including employees and shareholders.\(^4\) McConnell v. FEC\(^4\) consolidated the multiple lawsuits that challenged this section of the BCRA on First Amendment grounds. In McConnell, the Court held the ban constitutional, because corporations and unions would “remain free to organize and administer segregated funds, or PACs, for [the] purpose” of funding electioneering communications.\(^4\) Importantly, the Court also noted that “Congress’s power to prohibit corporations and unions from using funds in their treasuries to finance advertisements expressly advocating the election or defeat of candidates in federal elections has been firmly embedded in our law.”\(^4\)

Prior to Citizens United, the Supreme Court most recently grappled with the First Amendment implications of

\(^4\) Hasen, supra note 32, at 588-89; see also 2 U.S.C. § 441b(b)(2).
\(^4\) See 11 C.F.R. § 100.6 (2010).
\(^4\) McConnell, 540 U.S. at 204; see also Citizens United, 130 S. Ct. at 894 (“McConnell permitted federal felony punishment for speech by all corporations, including nonprofit ones, that speak on prohibited subjects shortly before federal elections.”). McConnell relied heavily on Austin in its analysis and reasoning. Id. at 913 (“The McConnell Court relied on the antidistortion interest in Austin . . . .”); see also Bopp & Coleson, supra note 38, at 33 (“[McConnell] simply relied on Austin . . . .”); Wilson, supra note 9, at 2372.
\(^4\) McConnell, 540 U.S. at 203.
limiting corporate political expenditures in FEC v. Wisconsin Right to Life (WRTL). WRTL involved a nonprofit organization’s challenge to the BCRA’s ban on the use of corporate funds for “electioneering communications” during certain preelection time periods—the very provision upheld in McConnell.

WRTL planned to broadcast radio advertisements within thirty days of a primary election that would call on Wisconsin’s U.S. Senators to stop delaying a vote on President George W. Bush’s judicial nominees. Despite the Court’s purported affirmation of McConnell, the Court construed “express advocacy” very narrowly and held that “BCRA § 203 [was] unconstitutional as applied to WRTL’s [advertisements],” because the advertisements exemplified issue advocacy, not express advocacy. This decision acted as a harbinger of the hard-line, antiregulation approach to corporate political expenditures that would become the new paradigm.

II. CITIZENS UNITED

On January 21, 2010, the Supreme Court decided Citizens United v. FEC, in which it struck down the ban on independent corporate political expenditures on First Amendment grounds. Specifically, the Court held that the laws banning corporations from using treasury funds for “electioneering communications” unconstitutionally silenced political speech. The decision extended constitutional

---

48 Id. at 455-56; see also 2 U.S.C. § 441b(b)(2); McConnell, 540 U.S. at 207. Because the challenge only regarded the BCRA’s application to WRTL’s specific situation, it was an “as-applied” challenge. Wis. Right to Life, 551 U.S. at 456.
49 Wis. Right to Life, 551 U.S. at 458-59.
50 Id. at 481. Professor Hasen notes that although it was not facially invalidated by the Court, “[t]he new functional equivalency test appeared likely to eviscerate [BCRA] § 203.” Hasen, supra note 32, at 590.
51 See James A. Gardner, Anti-Regulatory Absolutism in the Campaign Arena: Citizens United and the Implied Slippery Slope, 20 CORNELL J.L. & PUB. POL’Y 673, 679-81 (2011) (“Soon after its ruling in McConnell, the Court watered down its holding on corporate and union spending by approving an as-applied challenge to the very provision of BCRA it had recently upheld.” (citing Wis. Right to Life, 551 U.S. at 456)); Zipkin, supra note 34, at 556.
52 2 U.S.C. § 441b (2006); Citizens United, 130 S. Ct. at 897 (“Section 441b makes it a felony for all corporations—including nonprofit advocacy corporations—either to expressly advocate the election or defeat of candidates or to broadcast electioneering communications within 30 days of a primary election and 60 days of a general election.”).
53 Citizens United, 130 S. Ct. at 917.
54 Id. at 913.
protection to corporations’ ability to use general-treasury funds to make independent political expenditures that are not coordinated with candidates for office.\textsuperscript{56}

A. The Facts

Citizens United, a nonprofit corporation, defines itself as “dedicated to restoring our government to citizens’ control...[t]hrough a combination of education, advocacy, and grass roots organization.”\textsuperscript{57} It accepts donations from individuals and from for-profit corporations.\textsuperscript{58} The suit, which would later send shockwaves through the legal and political communities, resulted from a film that Citizens United produced and released in January 2008 entitled Hillary: The Movie (Hillary),\textsuperscript{59} a ninety-minute documentary about then-Senator Hillary Clinton’s candidacy for President.\textsuperscript{60} As the Court noted, “Hillary mention[ed] Senator Clinton by name and depict[ed] interviews with political commentators and other persons, most of them quite critical of Senator Clinton.”\textsuperscript{61} Citizens United released the film to theaters and put it out on DVD, but the organization also wanted to arrange for distribution via video-on-demand and run advertisements on broadcast and cable television that included a statement about Clinton, the name of the film, and the movie’s web address.\textsuperscript{62} Citizens United planned to have at least three of the advertisements broadcasted within thirty days before the 2008 Democratic National Convention and within sixty days before the 2008 general election, ostensibly in violation of the BCRA’s

\textsuperscript{56} WHITAKER, supra note 13, at 1-2; GARRETT, supra note 13, at 1-2.
\textsuperscript{58} Citizens United, 130 S. Ct. at 887. Professor Hasen has noted that “[b]y taking some for-profit corporate money, the corporation appeared ineligible for the MCFL exemption.” Hasen, supra note 32, at 591 n.59; see also supra notes 31-32 and accompanying text.
\textsuperscript{59} H ILLARY: THE MOVIE (Citizens United Productions 2008).
\textsuperscript{60} Citizens United, 130 S. Ct. at 887.
\textsuperscript{61} Id. at 887. Professor Hasen has noted that the documentary “contain[ed] a great many negative statements” about then-Senator Clinton. Hasen, supra note 32, at 591-92.
\textsuperscript{62} Citizens United, 130 S. Ct. at 887 (“Video-on-demand allows digital cable subscribers to select programming from various menus, including movies, television shows, sports, news, and music. The viewer can watch the program at any time and can elect to rewind or pause the program.”). The Court noted that they found the statements about then-Senator Clinton to be “pejorative.” Id. at 887.
ban on electioneering communications within those time periods. Before Citizens United, federal campaign-finance law prohibited corporations and unions from funding independent expenditures that expressly advocated for the election or defeat of a candidate (express advocacy) as well as "any broadcast, cable, or satellite communication that ‘refer[red] to a clearly identified candidate for Federal office’ and [was] made within [thirty] days of a primary or [sixty] days of a general election” (electioneering communication). As an alternative, federal campaign-finance law allowed corporations to establish PACs, which could in turn receive donations from shareholders or employees of the corporation and make independent political expenditures with those funds. If the FEC considered Hillary and the advertisements to be either express advocacy, electioneering communications, or both, then it could have subjected Citizens United to civil and criminal penalties. Therefore, Citizens United brought suit to seek declaratory and injunctive relief against the FEC regarding those provisions. The complaint argued that the prohibition on express advocacy and electioneering communications was unconstitutional as applied to the film and its advertisements.

B. The Analysis and Holding

The Citizens United decision steeply veered away from critical precedents. From the very beginning of the decision, the Court justified its enormous step in overturning two major decisions, Austin and McConnell, by framing the outcome of the case as nearly unavoidable. In its rationalization, the Court

---

63 Citizens United v. FEC, 530 F. Supp. 2d 274, 276 (D.D.C. 2008); see also Citizens United, 130 S. Ct. at 888. "Both periods are within BCRA’s definition of an electioneering communication." Citizens United, 530 F. Supp. 2d at 276. Citizens United’s complaint “contain[ed] two major claims: (1) that [the] prohibition of corporate disbursements for electioneering communications violates the First Amendment on its face and as applied [in this case]; and (2) that [the statute] requiring disclosure and . . . disclaimers are unconstitutional as applied” to Citizen United’s movie and advertisements. Id. at 277 (footnotes omitted).


65 2 U.S.C. § 441b(b)(2); Citizens United, 130 S. Ct. at 887-88; see also supra Part I.

66 Citizens United, 130 S. Ct. at 888; see also 2 U.S.C. § 437g.

explained it could not decide the case on a narrower ground—either through statutory interpretation or a limited constitutional holding. In particular, the Court took great pains to explain the lack of opportunity for a ruling based on the more limited as-applied challenge brought by Citizens United. For example, the Court pointed out that because the District Court addressed the facial validity of the statute . . . it was necessary for the Court to consider . . . facial validity as well. The Court "declined to adopt an interpretation that require[d] intricate case-by-case determinations to verify whether political speech is banned," because such a "course of decision would prolong the substantial, nation-wide chilling effect caused by [section] 441b's prohibitions on corporate expenditures." The Court's conclusion that it could not pursue narrower grounds and its use of statements such as, "First Amendment freedoms need breathing space to survive," indicated what would later become the sweeping holding of the Court.

---

68 Citizens United, 130 S. Ct. at 888 ("[W]e first address whether Citizens United's claim that § 441b cannot be applied to Hillary may be resolved on another, narrower grounds."); see also Richard Briffault, Corporations, Corruption, and Complexity: Campaign Finance After Citizens United, 20 CORNELL J.L. & PUB. POL'y 643, 664 (2011) ("The Court [was] determined not to resolve Citizens United's complaint on any . . . narrower grounds. . . ."); Richard H. Fallon, Jr., Fact and Fiction About Facial Challenges, 99 CALIF. L. REV. 915, 946 (2011) ("[T]he Court could have held a challenged provision of the Bipartisan Campaign Finance Reform Act invalid as applied to the party before it, but instead opted to pronounce it invalid on its face."); Harold Anthony Lloyd, A Right but Wrong Place: Righting and Rewriting Citizens United, 56 S.D. L. REV. 219, 220 (2011) (The Court "[d]eclined to decide the case on narrower grounds and resurrected a facial challenge that Citizens United had waived . . . .").

69 Citizens United, 130 S. Ct. at 890 ("[I]nere is no reasonable interpretation of Hillary other than as an appeal to vote against Senator Clinton. Under the standard stated in McConnell and further elaborated in WRTL, the film qualifies as the functional equivalent of express advocacy."). The Court also noted that although "the Court should construe statutes as necessary to avoid constitutional questions, the series of steps suggested would be difficult to take in view of the language of the statute." Id. at 892.

70 Id. at 891-92. An as-applied challenge would only raise the question of whether the relevant statute was constitutionally invalid in a specific situation. However, a facial challenge is a general challenge to the statute itself. Commentators have noted that this case ended the Roberts Court's trend of a "resistance to constitutional challenges seeking the facial invalidation of laws . . . ." Patricia Millett et al., Mixed Signals: The Roberts Court and Free Speech in the 2009 Term, 5 CHARLESTON L. REV. 1, 14 (2010).

71 Citizens United, 130 S. Ct. at 893.

72 Id. at 892, 894.

73 Id. at 892 (quoting FEC v. Wis. Right to Life, Inc., 551 U.S. 449, 468-69 (2007)).

74 In his dissent, Justice Stevens argued that the Court could and should have decided the case on the narrow grounds and that the majority overreached in its decision. Id. at 931-32 (Stevens, J., dissenting).
Next, the Court found it “necessary to consider Citizens United's challenge to Austin and the facial validity of [section] 441b’s expenditure ban.” As discussed above, section 441b criminalizes express advocacy or electioneering communications by a corporation. Although the Court acknowledged that PACs provide some alternative for corporations seeking to take part in public political discourse, the Court held that the existence of this alternative was “burdensome,” “expensive to administer,” “subject to extensive regulations,” and thus “[d]id not alleviate the First Amendment problems with [section] 441b.” In addition, the Court determined that these practical concerns about PACs rendered “[s]ection 441b’s prohibition on corporate independent expenditures [as]... a ban on speech” whose “purpose and effect [were] to silence entities whose voices the Government deem[ed] to be suspect.”

The Court analyzed section 441b as if it applied to individuals, based on the premise that “political speech of corporations or other associations should [not] be treated differently under the First Amendment simply because such associations are not ‘natural persons.’” Therefore, the Court considered whether a regulation like section 441b could be applied to an individual and concluded that if Congress enacted such a regulation, “no one would believe that it [was] merely a time, place, or manner restriction on speech.” This vast

---

75 Id. at 893 (majority opinion). The Court noted that “[a]ny other course of decision would prolong the substantial, nationwide chilling effect caused by [section] 441b’s prohibitions on corporate expenditures.” Id. at 894.
76 Id. at 897 (“Section 441b makes it a felony for all corporations—including nonprofit advocacy corporations—either to expressly advocate the election or defeat of candidates or to broadcast electioneering communications within [thirty] days of a primary election and [sixty] days of a general election.” (emphasis added)).
77 Id. As part of their exploration of the practical concerns about the PAC alternative to regular corporate political speech, the Court also found that “[g]iven the onerous restrictions, a corporation may not be able to establish a PAC in time to make its views known regarding candidates and issues in a current campaign.” Id. at 898.
78 Id.
79 Id. at 900 (citing First Nat’l Bank of Bos. v. Bellotti, 435 U.S. 765, 776 (1978)).
80 Id. at 898. Generally, time, place, or manner restrictions on speech are constitutional. The Court also noted that
disparity between speakers in the political realm doomed section 441b’s chance of being upheld as constitutional. The Court found that laws that restrict the speech of some speakers, but not others, constitute an attempt to control the content of the restricted speakers.81

In addition to analyzing the law’s actual and potential effects, the Court assessed the interests the law served in order to determine whether its goals justified the means it employed. For these types of assessments, the Court employs various levels of “scrutiny.” The level of scrutiny applied often plays a crucial role in dictating the outcome of a First Amendment challenge.82 In Citizens United, the Court stated that “[l]aws that burden political speech are ‘subject to strict scrutiny,’ which requires the Government to prove that the restriction ‘furthers a compelling interest and is narrowly tailored to achieve that interest.’”83

Under the strict scrutiny analysis, only an interest that the Supreme Court recognizes as compelling will justify some of the most dubious legislation.84 A narrowly tailored law is one that is not over- or underinclusive in serving the compelling interest it purports to further. In Citizens United, the Court spent most of its energy determining whether a compelling interest existed rather than analyzing whether the law was narrowly tailored.85 In order to serve a compelling interest in this

---

81 Id. at 895-96 (citations omitted).
82 Id. at 898-99 (“Prohibited, too, are restrictions distinguishing among different speakers, allowing speech by some but not others. As instruments to censor, these categories are interrelated: Speech restrictions based on the identity of the speaker are all too often simply a means to control content.” (citation omitted)).
85 One commentator has noted that “[t]he Court did not deny that corporate political speech does present [certain] problems but believed that the statutory ban was both over- and under-inclusive. . . .” Melone, supra note 15, at 86. The Court’s focus on whether the law served a compelling interest may indicate that the Court was less
case, the Court suggested that the regulation would have to restrict speech or expenditures that "interfered with governmental functions," such as protecting a school or prison's ability to operate. The Court found that this regulation failed to support a government function and actually undermined the political process by preventing voters from "obtain[ing] information from diverse sources," including corporations, "in order to determine how to cast their votes." Therefore, the Court concluded that there is "no basis for the proposition that, in the context of political speech, the Government may impose restrictions [such as section 441b] on certain disfavored speakers," such as corporations.

Finally, the Court sought to provide clarity regarding its decision to overrule Austin and partially overrule McConnell. The Court based its departure from Austin and McConnell on the ground that "[t]he ongoing chill upon speech that is beyond all doubt protected makes it necessary in this case to invoke the earlier precedents that a statute [that] chills speech can and must be invalidated where its...invalidity has been demonstrated." Specifically, the Court held that the pre-Austin line of precedent, as exemplified by Bellotti, forbids any restriction on political speech on the basis of the speaker's corporate identity. In addition, the Court found that the post-Austin line of precedent would have erroneously permitted such a restriction as justified by an interest in preventing distortion of speech within the political process.

As in previous cases, the Court focused on the compelling governmental interest prong of the strict scrutiny test rather than on the breadth prong. In Austin, the Court identified the "compelling governmental interest" as "preventing the corrosive and distorting effects of immense aggregations of wealth that are accumulated with the help of

---

86 Citizens United, 130 S. Ct. at 899. The Court noted that it had "upheld a narrow class of speech restrictions that operate[d] to the disadvantage of certain persons, but th[o]se rulings were based on an interest in allowing government entities to perform their functions." Id. (citing Bethel Sch. Dist. No. 403 v. Fraser, 478 U.S. 675, 683 (1986); Jones v. N.C. Prisoners' Labor Union, Inc., 433 U.S. 119, 129 (1977); Parker v. Levy, 417 U.S. 733, 759 (1974); Civil Serv. Comm'n v. Letter Carriers, 413 U.S. 548, 557 (1973)).
87 Id.
88 Id.
89 Id. at 896 (citing FEC v. Wis. Right to Life, 551 U.S. 449, 482-83 (2007)).
90 Id. at 898-99 (citing First Nat'l Bank of Bos. v. Bellotti, 435 U.S. 765, 784 (1978)).
91 Id. at 903-04.
the corporate form and that have little or no correlation to the public's support for the corporation's political ideas." The Court did not take issue with the justification itself but instead focused on the practical outcomes of accepting that justification as a compelling governmental interest. For example, the Court stressed that if "the antidistortion rationale were to be accepted, . . . it would permit Government to ban political speech simply because the speaker is an association that has taken on the corporate form." The Court cast aside the Austin Court's reliance on the antidistortion rationale as "an aberration" and held that the true "purpose and effect of [section 441b] is to prevent corporations, including small and nonprofit corporations, from presenting both facts and opinions to the public." Therefore, the Court overruled Austin and part of McConnell because section 441b was an "unlawful" and "troubling assertion of brooding governmental power." Instead, the Court chose to "return to the principle . . . that the Government may not suppress political speech on the basis of the speaker's corporate identity."

---

92 Id. at 903 (quoting Austin v. Mich. Chamber of Commerce, 494 U.S. 652, 660 (1990)).
93 Id. at 904. The Court also noted the Government's argument "that Austin permits [the Government] to ban corporate expenditures for almost all forms of communication stemming from a corporation." Id. (citation omitted). The Court brought the Government's argument to what it viewed as its logical conclusion and found that "[i]f Austin were correct, the Government could prohibit a corporation from expressing political views in media beyond those presented here, such as by printing books." Id. For an extensive discussion of the antidistortion rationale, see Richard L. Hasen, Citizens United and the Orphaned Antidistortion Rationale, 27 GA. ST. U. L. REV. 989 (2011).
94 Id. at 907. The Court also noted that under section 441b, "certain disfavored associations of citizens—those that have taken on the corporate form—are penalized for engaging in the same political speech" as wealthy individuals and unincorporated associates, who could "spend unlimited amounts on independent expenditures." Id. at 908 (citation omitted).
95 Citizens United "overruled the portion of McConnell facially upholding the electioneering-communication prohibition. As a result, the [Court's] language about government being able to prohibit 'functional equivalent of express advocacy' is gone." Bopp & Coleson, supra note 38, at 58 (quoting Citizens United, 130 S. Ct. at 913); see also Wilson, supra note 9, at 2365 (Citizens United overruled "the portion of [McConnell] that restricted independent corporate expenditures, as codified in section 203 of [BCRA]" (citations omitted)).
96 Citizens United, 130 S. Ct. at 908.
97 Id. at 904.
98 Id. at 913. The Court based this conclusion on its holding that "[n]o sufficient governmental interest justifies limits on the political speech of nonprofit or for-profit corporations." Id.
III. Mandatory Shareholder-Approval Proposals

After Citizens United, lawmakers, journalists, and think tanks proposed several versions of a mandatory shareholder-approval scheme in an attempt to rein in the new political rights granted to corporations. Three of the most prominent proposals include the End the Hijacking of Shareholder Funds Act, the Shareholder Protection Act of 2010, and the Brennan Center for Justice’s proposal. Although the three proposals are similar, they have important distinctions that may bear upon the constitutionality of their limitations on corporate speech under the First Amendment.

A. End the Hijacking of Shareholder Funds Act

On January 21, 2010, the same day the Court handed down its decision in Citizens United, U.S. Representative Alan Grayson responded by introducing the End the Hijacking of Shareholder Funds Act (Shareholder Funds Act). At barely

\[\text{\textsuperscript{99} See, e.g., End the Hijacking of Shareholder Funds Act, H.R. 4487, 111th Cong. § 2 (2010); Shareholder Protection Act of 2010, H.R. 4537, 111th Cong. § 3 (2010).}\]

\[\text{\textsuperscript{100} See, e.g., Alter, supra note 19 ("New laws regulating corporate governance are also essential. Britain requires shareholders to vote on corporate political expenditures. We should do the same . . . .").}\]

\[\text{\textsuperscript{101} See, e.g., Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 1; Craig Holman, House Committee Passes Important Checks on Corporate Spending on Elections, PUB. CITIZEN (July 29, 2010), http://citizenvox.org/2010/07/29/house-committee-passes-important-checks-on-corporate-spending-on-elections/ ("Shareholders should have a say on how their money is spent.").}\]

\[\text{\textsuperscript{102} One commentator’s analysis of past corporate political spending “suggest[s] that corporations and unions were major players on the political stage even before Citizens United, and it is unclear how much the Court’s ruling is likely to change that influence.” Gilpatrick, supra note 27, at 412.}\]

\[\text{\textsuperscript{103} End the Hijacking of Shareholder Funds Act, H.R. 4487, 111th Cong. (2010).}\]

\[\text{\textsuperscript{104} Shareholder Protection Act of 2010, H.R. 4537, 111th Cong. (2010).}\]

\[\text{\textsuperscript{105} Torres-Spelliscy, Corporate Political Spending, supra note 23; Torres-Spelliscy, Corporate Campaign Spending, supra note 18. Since Professor Torres-Spelliscy’s book chapter details the general proposal of the published report, this note will consider both publications in tandem as one proposal.}\]

\[\text{\textsuperscript{106} This note will only discuss the shareholder approval proposals. It will not discuss any proposed disclosure requirements because they are not relevant for the purposes of the First Amendment analysis under Citizens United. In addition, it is of interest that these proposals were put forth during a “trend toward shareholder empowerment.” See Martin Gelter, Taming or Protecting the Modern Corporation? Shareholder-Stakeholder Debates in a Comparative Light, 7 N.Y.U. J. L. & Bus. 641, 655-58 (2011).}\]

two pages, the Shareholder Funds Act would, if enacted, require that “[a]ny expenditure by a public company to influence public opinion on matters not related to the company’s products or services” have the approval of a “majority of the votes cast by shareholders.” Endnote 108. Expenditures without this approval would “be considered a breach of a fiduciary duty of the officers and directors who authorized such an expenditure.” Endnote 109. For such a breach, those officers and directors would be personally liable to the company’s shareholders for the amount of the expenditure. Endnote 110. Although seemingly simple and straightforward, the proposed bill is too brief for Congress to take it seriously as a piece of legislation that could be enacted in its current form. It has many definitional issues and details that require resolution before enactment could occur. For example, the bill defines only two terms—“public company” and “shareholder.” A “public company” is defined as “any issuer that is required to submit periodical or other reports under section 13 of the Securities Exchange Act of 1934 (15 U.S.C. § 78m)” and “shareholder” is defined as “any person who owns or holds a share of stock in a public company.” Endnote 111. In addition, the Shareholder Funds Act does not define those matters that “influence public opinion” or those matters that do—or do not—“relate[] to the company’s products or services.” Endnote 112. These definitional deficiencies raise many questions. Are institutional shareholders, such as mutual or pension funds, given votes under the definition? Do shareholders have just one vote, or one vote for each share of stock they own? Despite its incompleteness, the bill is important because it likely provided the basis for the Shareholder Protection Act of 2010, Endnote 113 later proposed by U.S. Representative Michael Capuano.

Endnote 109. Id.
Endnote 110. Id. (stating that officers and directors would be jointly and severally liable).
Endnote 111. Id. § 3.
Endnote 112. Id. § 2. Not only would these definitional issues be difficult for a corporation, an attorney, or a court to interpret, but they may very well render this bill unconstitutionally vague, even without reference to the inevitable First Amendment strict scrutiny analysis that it would have to survive.
B. Shareholder Protection Act

Introduced six days after Citizens United was decided, H.R. 4537, or the Shareholder Protection Act of 2010 (Shareholder Protection Act), is just over eight pages long—four times as long as the Shareholder Funds Act. The Shareholder Protection Act would amend section 14 of the Securities Exchange Act of 1934 in order to require any issuer of securities to obtain “written affirmative authorization” from a majority of its shareholders before spending more than $10,000 in a given fiscal year on political expenditures. Like the Shareholder Funds Act, the Shareholder Protection Act would consider a violation of the prior-approval requirement as a “breach of fiduciary duty of the officers and directors who authorized” the political expenditure, and those officers and directors would be “jointly and severally liable . . . for the amount of such expenditure.” The bill defines political expenditures very broadly—beyond the types of activities that section 441b covered before Citizens United. In addition to covering corporate independent political expenditures targeted by section 441b, this proposal would also cover expenditures for voter registration campaigns and trade association dues.

---

114 Id.
115 "Issuer" is defined by the federal securities laws as “any person who issues or proposes to issue any security [or] . . . the person or persons performing the acts and assuming the duties of depositor or manager pursuant to the provisions of [a] trust or other agreement or instrument under which such securities are issued . . . ." Securities Exchange Act of 1934 § 3(a)(8), 15 U.S.C. § 78c(a)(8) (2006). A “person” is defined broadly in federal securities law as “a natural person, company, government, or political subdivision, agency, or instrumentality of a government.” Securities Exchange Act of 1934 § 3(a)(9), 15 U.S.C. § 78c(a)(9); 2 U.S.C. § 434 (f) (2006).
116 Shareholder Protection Act of 2010, H.R. 4537, 111th Cong. § 3(d) (2010). The bill defines “affirmative authorization” as “the full, free, and written consent of a shareholder, obtained without intimidation or fear of reprisal, and shall not include votes made by a broker or any other representative.” Id. By not including brokers or representatives, it is unclear whether the bill would allow for institutional investors, mutual funds, or hedge funds to vote either for or against approval of political expenditures. The bill defines “majority of all shareholders” as the “number of shareholders that combined own more than 50 percent of all outstanding shares. Shareholders not casting votes shall not count toward such a majority.” Id.
117 Id.
118 Id. The bill defines “expenditure for political activities” as (i) expenditures in support of, or opposition to, any Federal, State, or local candidate; (ii) contributions to or expenditures in support of any political party, committee, electioneering communication, voter registration campaign, ballot measure campaign, or an issue advocacy campaign; and (iii) dues or other payments to trade associations or other tax exempt organizations that are, or could reasonably be anticipated to be, used for the purposes described in subparagraphs (ii) and (iii).
In recognition of the special treatment by the Court of media corporations in Bellotti and Citizens United,\textsuperscript{119} the Shareholder Protection Act includes an exception for issuers “whose sole business is the publication or broadcasting of news, commentary, literature, music, entertainment, artistic expression, scientific, historical or academic works, or other forms of information.”\textsuperscript{120} The bill would require the Securities and Exchange Commission to issue “guidance as it determines necessary or appropriate regarding the extent of the exemption.”\textsuperscript{121} Although more detailed than the Shareholder Funds Act, the Shareholder Protection Act would actually cover a broader group of corporations. While the Shareholder Funds Act would only cover public companies,\textsuperscript{122} the Shareholder Protection Act would cover all issuers of securities, including some private corporations and many small businesses.\textsuperscript{123}

C. Brennan Center for Justice Proposal

On the same day that Representative Capuano introduced the Shareholder Protection Act of 2010, the Brennan Center for Justice published a proposal (Brennan Center Proposal) that would “require shareholder authorization of future corporate political spending.”\textsuperscript{124} The Brennan Center Proposal has the same two essential elements as the Shareholder Funds Act and the Shareholder Protection Act; it would require “corporations [to] obtain the consent of shareholders before making political expenditures” and it would “hold corporate directors personally liable for violations of these policies.”\textsuperscript{125} In addition, the Brennan Center Proposal

\textsuperscript{120} Shareholder Protection Act of 2010, H.R. 4537, 111th Cong. § 3 (2010).
\textsuperscript{121} Id. This provision has the potential to bring the Securities and Exchange Commission into conflict with the FEC, which is tasked with providing guidance on federal campaign-finance law.
\textsuperscript{122} End the Hijacking of Shareholder Funds Act, H.R. 4487, 111th Cong. § 2 (2010).
\textsuperscript{123} H.R. 4537 § 3. However, the bill would carve out some limited exceptions. See id. § 3(d).
\textsuperscript{124} Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 4.
\textsuperscript{125} Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 21. This proposal would not hold corporate officers liable, as would be the case under the Shareholder Funds Act or the Shareholder Protection Act. See Torres-Spelliscy, Corporate Political Spending, supra note 23, (manuscript at 74) (proposing that “[i]f a corporation makes an unauthorized contribution or expenditure for a political activity,
includes disclosure requirements that are undoubtedly constitutional and helpful to shareholders that would facilitate the BCRA’s central goal of a transparent democracy. Like the proposed bills’ provisions, these requirements would amend federal securities law.

The Brennan Center Proposal would require an “annual shareholder vote to authorize any spending of $10,000 or more” by a corporation for any political activity, but the corporation may also request authorization more frequently. The proposal defines “political activity” as “any contributions or expenditures made directly or indirectly to, or in support of or opposition to, any candidate, political party, committee, electioneering communication, ballot measure campaign, or . . . issue advocacy campaign.” For shareholder approval, the corporation’s proposed political spending must receive a majority vote from the corporation’s shareholders, defined then the directors at the time that the unauthorized contribution or expenditure was incurred are jointly and severally liable to repay to the corporation the amount of the unauthorized expenditure, with interest at the rate of eight percent per annum” (emphasis added)).

For example, the Brennan Center Proposal urges Congress to require disclosure of political spending ... frequent enough to notify shareholders and the investing public of corporate spending habits, and yet with enough of a time lag between reports so that corporations are not unduly burdened. To accommodate these two competing goals, disclosure of political expenditures should occur quarterly to coincide with company’s filing of its Form 10-Qs with SEC. Because the political disclosure will be contemporaneous with the 10-Q filing, transaction costs can be minimized.

Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 21; see also Citizens United v. FEC, 130 S. Ct. 876, 914 (2010) (upholding disclosure requirements); WHITAKER, supra note 13, at 5. For a discussion supporting increased disclosure requirements, see Francis Bingham, Note, Show Me the Money: Public Access and Accountability After Citizens United, 52 B.C. L. REV. 1027, 1061-64 (2011). However, one commentator has argued that many disclosure regimes, including the DISCLOSE Act, may not be constitutional under Citizens United due to the Court’s concern about the complexity of campaign finance regulation. See Briffault, supra note 68, at 645-46, 668-70.

As the author of the Brennan Center Proposal has stated, “disclosure of past political expenditures empowers the shareholder to anticipate whether future political spending is likely to conform with his or her political views.” Torres-Spelliscy, Corporate Political Spending, supra note 23, (manuscript at 66); see also WHITAKER, supra note 13, at 7.

This portion of the proposal would amend section 14 of the Securities Exchange Act of 1934. See Torres-Spelliscy, Corporate Political Spending, supra note 23, (manuscript at 73).

This proposal would cover “[a]ny corporation where proxies are solicited in respect of any security registered under section 12 of the Exchange Act occurring on or after the date . . . on which final rules are issued . . . .” Id.

Id. at 72. This proposal expressly excludes “activities defined as lobbying under any local, state or federal law.” Id.
explicitly to include institutional investors. In addition, the Brennan Center Proposal includes a provision that would require institutional investment managers to publicly disclose how they voted on such political spending authorizations. By contrast, neither of the proposed bills addresses institutional investment managers—such as the managers of mutual, pension, or hedge funds—that often play crucial roles in securing shareholder majorities.

IV. FIRST AMENDMENT ANALYSIS OF MANDATORY SHAREHOLDER-APPROVAL PROPOSALS

After these proposals were put forth, the Congressional Research Service raised the question of whether lawmakers could draft a mandatory shareholder-approval law in a way that would not “violate[e] the corporation’s free speech rights as described by Citizens United.” In order to pass constitutional muster and comply with Citizens United, the law or regulation must not infringe upon the corporation’s right to free speech as guaranteed by the First Amendment.

---

131 Id. at 74. “Majority of all shareholders” is defined as “a vote of shareholders that combined own 50 percent plus one of all outstanding shares. Shareholders not casting votes shall not count toward affirmative authorization under this [proposal].” Id. at 72.

132 Id. at 74 (proposing that “[e]very institutional investment manager subject to section 13(f) of the Exchange Act shall report at least annually how it voted on any shareholder vote... unless such vote is otherwise required to be reported publicly by rule or regulation of the Securities and Exchange Commission.”).

133 As of 2002, 58 percent of U.S. equities were owned by institutional investors, a percentage that has risen nearly every year. Murat Binay, Performance Attribution of US Institutional Investors, FIN. MGMT., Summer 2005, at 127, 128. For further discussion of institutional shareholders in light of Citizens United, see Melone, supra note 15, at 85; Osterlind, supra note 16, at 281.

134 Congressional Research Service, LIBRARY OF CONGRESS (Nov. 2, 2010), http://www.loc.gov/ CRSinfo (“The Congressional Research Service (CRS) works exclusively for the United States Congress, providing policy and legal analysis to committees and Members of both the House and Senate, regardless of party affiliation. As a legislative branch agency within the Library of Congress, CRS has been a valued and respected resource on Capitol Hill for nearly a century.”).

135 WHITAKER, supra note 13, at 8; see also GARRETT, supra note 13, at 6 (stating that this proposal “could raise questions about whether the requirements were essentially stifling corporate political speech”). The Congressional Research Service also notes that “[t]he practicalities of... enforcement of this kind of legislation may need to be carefully considered in order not to run afoul of corporate freedom of speech rights defined by the Supreme Court in Citizens United.” WHITAKER, supra note 13, at 8.

136 Although the “Constitution’s Commerce Clause may arguably provide Congress with authority to enact legislation of the type in question,” the First Amendment must also be satisfied. WHITAKER, supra note 13, at 7 (citing U.S. CONST. art. I, § 8, cl. 3).
The First Amendment states, in relevant part, that “Congress shall make no law . . . abridging the freedom of speech, or of the press . . . .”\footnote{U.S. Const. amend. I.} All three proposals would create significant barriers, constitutional or not, to corporate political speech on multiple levels. When a law has the potential to burden political speech, it is subject to strict scrutiny.\footnote{Citizens United v. FEC, 130 S. Ct. 876, 898 (2010).} To survive this test, the government must show that the regulation advances a compelling governmental interest and is narrowly tailored to further that interest.\footnote{Id.} Therefore, the proposals must meet this exacting standard to pass muster.\footnote{In addition to a constitutional law analysis, corporate law considerations must be considered when evaluating the proposals discussed in this note. For arguments against the proposals on corporate law grounds, see Stephen A. Yoder, Legislative Intervention in Corporate Governance Is Not a Necessary Response to Citizens United v. Federal Election Commission, 29 J. L. & Com. 1, 15-21 (2010).} In order to overcome this burden, three interests have been set forth to justify the proposals—anticorruption, anticorrosion, and shareholder protection.

A. Anticorruption

The restrictions on speech embodied in the three proposals must advance a compelling governmental interest in order to survive strict scrutiny. The first potential compelling governmental interest is anticorruption. In Austin, Justice Marshall, writing for the majority, found that independent expenditures had the potential to cause “real or apparent corruption.”\footnote{Austin v. Mich. Chamber of Commerce, 494 U.S. 652, 659 (1990) (citing First Nat’l Bank of Bos. v. Bellotti, 435 U.S. 765, 788 (1978)).} Although the Austin Court acknowledged that direct contributions more naturally raised a concern about “‘financial quid pro quo’”\footnote{Id. (quoting FEC v. Nat’l Conservative Political Action Comm., 470 U.S. 480, 497 (1985)) (emphasis omitted).} than independent expenditures, it recognized that the prevention of this type of corruption provided a “‘compelling governmental interest . . . [for] the restriction of the influence of political war chests funneled through the corporate form.’”\footnote{Id. (quoting Nat’l Conservative Political Action Comm., 470 U.S. at 500-01).} Accordingly, the Austin Court held that corporations had no constitutional right to make independent campaign expenditures out of their general treasuries.\footnote{Id. at 655.
In McConnell, Justices Stevens and O'Connor, writing for the majority, expanded the breadth of anticorruption as a compelling governmental interest. The McConnell Court stated that Congress had a legitimate interest in stopping corruption, or the appearance thereof, well beyond eliminating mere “cash-for-votes” schemes. The Court held that this rationale extended to preventing “improper influence” and “opportunities for abuse,” including those that arose from elected officials who were “too compliant with the wishes of large contributors.” In their decision, Justices Stevens and O'Connor also dug deeper than the Austin Court to uncover what they essentially cast as the “real world” of political campaign financing. For example, they quoted many former lobbyists and members of Congress when they discussed the high-level influence and access gained through large independent corporate expenditures. The majority’s emphasis on the practical effect of the campaign-financing scheme reflected the Court’s concern with current political realities rather than with the theoretical implications of such regulations.

Unlike the majority opinions of the Austin and McConnell Courts, the anticorruption rationale as a compelling governmental interest found popularity only with the dissent in Citizens United. Justice Stevens, writing in dissent, argued that the Citizens United Court construed the anticorruption interest too narrowly and that it actually encompasses more than just quid pro quo corruption. Rather, he argued that the compelling governmental interest also included the prevention of independent expenditures from having an “undue influence on an office holder’s judgment” or from generating an “appearance of such influence.”

146 Id. at 143 (quoting Nixon v. Shrink Mo. Gov’t PAC, 528 U.S. 377, 389 (2000)). One commentator has noted that “Justice Kennedy distinguished corruption from ‘favoritism and influence,’ which were ‘not . . . avoidable in representative politics.’” Eliza P. Nagel, Note, For the People or Despite the People: The Threat of Corporations’ Growing Power Through Citizens United and the Demise of the Honest Services Law, 63 Rutgers L. Rev. 725, 760 (2011) (quoting Citizens United v. FEC, 130 S. Ct. 876, 910 (2010)).
147 McConnell, 540 U.S. at 150-52.
149 Id. (citations omitted). Justice Stevens elaborated that corruption can take many forms. Bribery may be the paradigm case. But the difference between selling a vote and selling access is a matter of degree, not kind. And selling access is not qualitatively different from giving special preference to those who spent money on one’s behalf. Corruption operates along a spectrum, and the majority’s apparent belief that quid pro quo
Justice Kennedy’s view of the anticorruption rationale was far narrower in Citizens United. As expressed in the majority opinion, Justice Kennedy defined the anticorruption rationale as only applying to quid pro quo corruption—essentially only bribery—and not as encompassing the broader definition that Justice Stevens advocated. In addition, Justice Kennedy articulated what appeared to be a different conception of democracy than that held by Justice Stevens. Rather than viewing politicians’ reactions to corporate expenditures as a sign of corruption, Justice Kennedy viewed the give-and-take between politicians and corporations as the relationship that should exist between representatives and their constituents. Instead of perceiving responsiveness to corporate interests as corruption, Justice Kennedy found such responsiveness to be of a democratic nature—which should be enjoyed by all constituents. Most importantly, the Citizens United Court, through Justice Kennedy, concluded that independent expenditures by corporations, by their nature, “do not give rise to corruption or the appearance of corruption.”

There is a plausible argument that mandatory shareholder approval would further the anticorruption interest, as broadly defined by Justice Stevens in his dissent. The premise of the proposals is that they will likely reduce the amount of money that corporations, in the aggregate, spend on political causes. However, no plausible argument exists that

arrangements can be neatly demarcated from other improper influences does not accord with the theory or reality of politics.

Id.

150 For a critical view of Justice Kennedy’s definition of corruption in Citizens United, see Andrew T. Newcomer, Comment, The “Crabbed View of Corruption”: How the U.S. Supreme Court Has Given Corporations the Green Light to Gain Influence over Politicians by Spending on Their Behalf, 50 WASHBURN L.J. 235, 267-71 (2010).

151 See Kang, supra note 9, at 250 (“Justice Kennedy’s view of corruption may limit campaign finance restrictions to not much beyond the regulation of contributions to candidates and officeholders.”); Peterman, supra note 12, at 1175-78.

152 Citizens United v. FEC, 130 S. Ct. 876, 910 (2010) (“The fact that speakers may have influence over or access to elected officials does not mean that these officials are corrupt.”).

153 Id. (“It is in the nature of an elected representative to favor certain policies, and, by necessary corollary, to favor the voters and contributors who support those policies . . . . Democracy is founded on responsiveness.” (quoting McConnell v. FEC, 540 U.S. 93, 297 (2003) (Kennedy, J., concurring in part))).

154 Id. at 909. “By definition, an independent expenditure is political speech presented to the electorate that is not coordinated with a candidate. The fact that a corporation, or any other speaker, is willing to spend money to try to persuade voters presupposes that the people have the ultimate influence over elected officials.” Id. at 910 (citation omitted).
these proposals would further the narrower definition of anticorruption stated in Citizens United by Justice Kennedy's majority opinion. The inferential connection between shareholder approval of independent political expenditures and the prevention of quid pro quo corruption is tenuous. As the Citizens United Court expressly stated, the independent character of the expenditures makes them—by the expenditures' very nature—in capable of creating a quid pro quo relationship, because, by definition, it is unlawful for those who make independent expenditures to coordinate with a candidate for federal office.\textsuperscript{155} Therefore, the anticorruption rationale cannot serve as a compelling governmental interest for the mandatory shareholder-approval proposals.

B. Anticorrosion

The second potentially compelling governmental interest is anticorrosion. In Austin, Justice Marshall, writing for the majority, found that in addition to the government's interest in preventing real or apparent corruption, the government had a compelling interest in preventing corporations from unfairly influencing elections through independent expenditures.\textsuperscript{156} The Austin Court differentiated between independent expenditures by individuals and those by corporations on the ground that corporations benefited from "the unique state-conferred corporate structure that facilitates the amassing of large treasuries."\textsuperscript{157} Based on this finding, the Court held that corporations' "immense aggregations of wealth" had the potential to be "corrosive" in the political arena because the independent expenditures funded by that wealth "have little or no correlation to the public's support for the corporation's political ideas."\textsuperscript{158}

Justice Scalia rejected the Court's categorization of anticorrosion as a compelling governmental interest in a dissenting opinion in Austin. Instead, Justice Scalia called corrosion the "New Corruption" and argued that the attempt to fight so-called corrosion was actually an effort by the Court to

\textsuperscript{155} Id. at 909-11.
\textsuperscript{157} Id.
\textsuperscript{158} Id. The Court further stated that the legislation in question "does not attempt to equalize the relative influence of speakers on elections, rather, it ensures that expenditures reflect actual public support for the political ideas espoused by corporations." Id. (internal quotation marks and citations omitted).
mandate that “[e]xpenditures must reflect actual public support for the political ideas espoused.” The Court could use this type of equalization rationale, Justice Scalia argued, to restrict “virtually anything [it] deems politically undesirable.” Relying on the Court’s holding in *Buckley v. Valeo* that independent expenditures raise little risk of corruption, Justice Scalia concluded that the Court’s reasoning cannot be reconciled “unless one thinks it would be lawful to prohibit men and women whose net worth is above a certain figure from endorsing political candidates.”

In *Citizens United*, the Court rejected Austin’s anticorrosion interest on the basis that it was essentially an interest in equalizing speech. Echoing Justice Scalia’s dissent in *Austin*, the Court noted that any interest in equalizing speech in the political realm had been rejected as early as *Buckley*. Chief Justice Roberts, writing in concurrence, argued that Austin’s reasoning was in conflict “with Buckley’s explicit repudiation of any government interest in equalizing the relative ability of individuals and groups to influence the outcome of elections.” The Court restored the clear distinction between independent expenditures and contributions with the potential for corruption or corrosion and noted that independent expenditures simply create little or no risk of either problem.

In the context of the mandatory shareholder-approval proposals, a plausible argument exists that the proposals will serve the anticorrosion interest, as explained by Justice Marshall in *Austin*. The premise of the proposals is that they will likely reduce the amount of money that corporations, in the aggregate, spend on political causes. After *Citizens United*, however, there is little possibility that the Court will recognize an anticorrosion interest, much less one that would justify restrictions on independent expenditures. Justice Kennedy, writing for the

---

159 Id. at 684 (Scalia, J., dissenting) (citation omitted).
160 Id.
161 424 U.S. 1 (1976). *Buckley* involved a constitutional challenge to the Federal Election Campaign Act by various candidates for federal office, political parties, and organizations. The Court upheld provisions limiting individual contributions to campaigns but invalidated provisions that limited the amount a candidate could spend on his own election and the total expenditures for particular campaigns. Id. at 58-59.
164 Id. at 904.
165 Id. at 921 (Roberts, C.J., concurring) (internal quotation marks and citation omitted).
majority, made it clear that the Court has found the anticorrosion interest as the equivalent of speech equalization, and Buckley strongly prohibits speech equalization as a compelling government interest. Therefore, the anticorrosion rationale cannot serve as a compelling governmental interest for the mandatory shareholder-approval proposals.

C. Shareholder Protection

In the wake of Citizens United, it is unlikely that the Court will uphold a mandatory shareholder-approval law, or any restriction on corporate independent expenditures, on the basis of the anticorruption or anticorrosion rationales. The Citizens United Court construed the anticorruption rationale narrowly and dismissed the anticorrosion rationale entirely. Because neither of these interests has the potential to function as the compelling governmental interest needed to justify the mandatory shareholder-approval proposals, another potential interest must exist for the proposals to pass constitutional muster. Those in favor of limits on corporate expenditures, the supporters of the BCRA and lamenters of Citizens United, have now focused their attention on the one rationale left standing after Citizens United: the shareholder-protection interest.

The mandatory shareholder-approval proposals are based on the premise that shareholders need new regulations to protect them in light of the rights afforded to corporations under Citizens United. In particular, the concern is that publicly traded companies will spend money invested by shareholders in a way that may overlook the interests of the shareholders and instead support the managers' personal

---

166 Id. at 904 (majority opinion).
167 See Kang, supra note 9, at 245 (noting “the fact that a government restriction makes shareholder speech more difficult is obviously insufficient by itself to justify a constitutional prohibition of that restriction”).
168 See Torres-Spelliscy, Corporate Political Spending, supra note 23, (manuscript at 69). However, one commentator has argued that in the Citizen United Court’s “majority opinion, Justice Kennedy reject[ed] the shareholder protection interest as a reason for restricting corporate speech” and “that corporate democracy mechanisms should be even more effective today than they would have been at the time the [BCRA] was passed . . . .” Yoder, supra note 140, at 11. For a policy-based criticism of the shareholder-protection rationale, see David G. Yosifon, The Public Choice Problem in Corporate Law: Corporate Social Responsibility After Citizens United, 89 N.C. L. Rev. 1197, 1228-30 (2011).
political agendas or otherwise damage the corporation. The proponents of mandating shareholder approval of such expenditures hope to curb “unfettered corporate political spending” and eliminate the risk of corporate “managers’ potentially profligate spending on politics.”

The Court first identified shareholder protection as a governmental interest in Bellotti, the same case that recognized the political speech rights of corporations. In Bellotti, the Court struck down a state statute that banned corporate political expenditures on referenda unrelated to that corporation’s proprietary interests. Similar to the method for examining a statute for narrow tailoring, the Bellotti Court’s “fit” analysis essentially saw the potential for shareholder democracy as a less burdensome alternative to regulation and assessed whether the statute was over- or underinclusive. Because the Court based its holding on the statute’s failure under the “fit” analysis, the Court did not reach the issue of whether shareholder protection can serve as a compelling governmental interest.

The shareholder-protection issue was also raised but not addressed conclusively in MCFL. Unlike the for-profit corporation and referendum-campaign context in Bellotti, the MCFL Court had to resolve a First Amendment challenge in a situation that involved a nonprofit corporation and an election for political office. Because the entity at issue was a nonprofit corporation, which by nature does not have shareholders, no potential existed for board members to spend shareholder

---

169 Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 9-10; see also Lucian A. Bebchuk & Robert J. Jackson, Jr., Corporate Political Speech: Who Decides?, 124 Harv. L. Rev. 83, 90-94 (2010).
170 Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 8-9. However, “[j]ust making independent speakers jump through more regulatory hoops may not create any curbing effect in the real world . . . .” Note, Restoring Electoral Equilibrium in the Wake of Constitutionalized Campaign Finance, 124 Harv. L. Rev. 1528, 1546 (2011). Furthermore, one commentator has argued that “reformers who insist on limiting the political arena of corporate speech greatly exaggerate the effect that such speech has on political outcomes.” Melone, supra note 15, at 94; see also Bingham, supra note 126, at 1047 (“[I]t is questionable whether the U.S. Supreme Court’s 2010 decision in Citizens United v. FEC will alter the way corporations spend money on federal elections at all.”).
172 Id. at 780.
173 Id. at 791-92; see also Lambert, supra note 29, at 21.
174 Bellotti, 435 U.S. at 793-94.
175 Id. at 795.
177 Id. at 241-42.
investments on political expenditures without their consent.\textsuperscript{178} Instead, the members of MCFL joined the organization because they agreed with its political or ideological purpose.\textsuperscript{179} On that basis, the Court held that ideological nonprofits that do not take corporate or union donations could not be banned from spending their treasury funds on explicit political advocacy.\textsuperscript{180}

The Austin Court built on the shareholder-protection rationale discussed in Bellotti and MCFL and held that corporations had no constitutional right to make independent campaign expenditures out of their general treasuries,\textsuperscript{181} although they could support issue advocacy that mentioned specific candidates.\textsuperscript{182} In upholding a state law that banned nonmedia corporations from using corporate funds from their general treasuries to make independent expenditures on behalf of, or against, state electoral candidates,\textsuperscript{183} the Court found that the shareholders’ ability to withdraw from corporate association as a result of political expenditures would not sufficiently protect them.\textsuperscript{184} Thus, the Austin Court concluded there was a compelling governmental interest in “preventing a corporation . . . from exploiting those who do not wish to contribute to [its] political message.”\textsuperscript{185}

\begin{footnotesize}
\begin{itemize}
\item\textsuperscript{178} Id. at 241 (“MCFL was incorporated . . . as a nonprofit, nonstock corporation under Massachusetts law.”).
\item\textsuperscript{179} Id. at 242. MCFL’s “resources come from voluntary donations from ‘members,’ and from various fundraising activities such as garage sales, bake sales, dances, raffles, and picnics. The corporation considers its ‘members’ those persons who have either contributed to the organization in the past or indicated support for its activities.” Id. (footnote omitted).
\item\textsuperscript{180} Id. at 262-64. The shareholder-protection interest was not a basis for the decision because, as a nonprofit corporation, MCFL had “members,” not “shareholders.”
\item\textsuperscript{182} Bopp & Coleson, supra note 38, at 32.
\item\textsuperscript{183} Austin, 494 U.S. at 654.
\item\textsuperscript{184} See id. at 663. The Court did point out that similar to MCFL, the Chamber also lacks shareholders, [but] many of its members may be . . . reluctant to withdraw as members even if they disagree with the Chamber’s political expression, because they wish to benefit from the Chamber’s nonpolitical programs and to establish contacts with other members of the business community. The Chamber’s political agenda is sufficiently distinct from its educational and outreach programs that members who disagree with the former may continue to pay dues to participate in the latter . . . . Thus, we are persuaded that the Chamber’s members are more similar to shareholders of a business corporation than to the members of MCFL.
\item\textsuperscript{185} Id. at 675. On a practical level, the Court voiced concern that “shareholders in a large business corporation may find it prohibitively expensive to monitor the activities of the corporation to determine whether it is making expenditures to which they object.” Id. at 674 n.5.
\end{itemize}
\end{footnotesize}
In Citizens United, the Court returned to essentially the same shareholder-protection analysis set forth in Bellotti—a three-part “fit” inquiry that examined (1) whether corporate democracy can adequately protect shareholders, (2) whether the statute purportedly protecting shareholders is overinclusive, and (3) whether the statute purportedly protecting shareholders is underinclusive. If corporate democracy, as a less restrictive means, can adequately protect shareholders or if the statute is either over- or underinclusive, then the statute is constitutionally invalid. The Citizens United Court found that the relevant portion of the BCRA failed on all three parts of the inquiry because voluntary corporate democracy protected shareholders’ interests adequately in this situation, and the statute was both over- and underinclusive. The Court found the BCRA overinclusive in its coverage of all corporations—including those with merely one shareholder—and underinclusive in its ban on electioneering communications taking effect only during specific time periods.

In overruling Austin, the Citizens United Court also cast aside the Austin Court’s formulation of shareholder protection as a compelling governmental interest. Justice Kennedy stated that the current structure of corporate democracy is able to adequately protect shareholders’ interests. The Court did not suggest that legislation was needed to strengthen or mandate those procedures, nor did it suggest that the voluntary nature of corporate democracy created an independent issue.

The Shareholder Funds Act, the Shareholder Protection Act, and the Brennan Center Proposal are all aimed ostensibly at the protection of shareholders. Beyond the suggestion in Citizens United that the current structure of corporate democracy can serve to protect this interest without further legislation, the next barrier that the proposals must overcome is whether they are narrowly tailored to the interest of

---

187 Citizens United, 130 S. Ct. at 911.
188 Id.
189 Id.; see also John C. Coates, Corporate Governance and Corporate Political Activity: What Effect Will Citizens United Have on Shareholder Wealth? 16 (Harvard L. & Econ., Discussion Paper No. 684, 2010), available at http://ssrn.com/abstract=1680861 (“In Citizens United, the Supreme Court relaxed the ability of corporations to spend money on elections, and in so doing, the Court rejected a shareholder-protection rationale for restrictions on spending.”).
190 Citizens United, 130 S. Ct. at 911 (“There is . . . little evidence of abuse that cannot be corrected by shareholders through the procedures of corporate democracy.” (quotation marks and citation omitted)).
191 See id.
shareholder protection. In order to reach this analysis, this note assumes for the sake of argument that the Court could find shareholder protection to be a compelling governmental interest for this type of legislation.

1. End the Hijacking of Shareholder Funds Act

The Shareholder Funds Act will likely fail a “narrowly tailored” analysis because it contains three aspects that render it overinclusive. First, the Act covers a far broader scope of political spending than just independent political expenditures, which is the narrow class of speech that Citizens United suggested that shareholders may have an interest in voluntarily restricting. Instead, this bill would sweepingly cover all expenditures to influence public opinion in a way that only indirectly relates to that corporation’s products or services. This overbreadth has the potential to restrict speech well beyond the political realm and include environmentally supportive speech and many general goodwill advertising campaigns that do not actually “relate” to products or services. There is little basis to suggest that the government can restrict these types of speech properly under the First Amendment.

---

192 Some commentators have argued that rather than creating a new burden for corporate political speech, the mandatory shareholder approval proposals simply legislate the “corporate democracy” that Justice Kennedy referenced in his opinion. See Bebchuk & Jackson, supra note 160, at 90-94. However, Justice Kennedy’s opinion put strong emphasis on using the least restrictive means possible to ameliorate any issues that could potentially provide a compelling governmental interest. See Citizens United, 130 S. Ct. at 911. Justice Kennedy clearly stated that shareholder protection could not justify the BCRA regulations at issue in Citizens United because of the options already available through corporate democracy. See id. Rather than “enhance” or “enable” corporate democracy, the mandatory voting proposals would likely be viewed as an additional, and unconstitutional, burden by the Court.

194 See Citizens United, 130 S. Ct. at 911. End the Hijacking of Shareholder Funds Act, H.R. 4487, 111th Cong. § 2 (2010) (“Any expenditure by a public company to influence public opinion on matters not related to the company’s products or services that has not been approved by a majority of the votes cast by shareholders to approve or disapprove such expenditure shall be considered a breach of a fiduciary duty of the officers and directors who authorized such an expenditure.”).

195 Corporations often run general “Happy Holidays” or other types of advertisements that are “not related” to their products or services. For example, Kmart paid for an advertisement in the New York Times shortly after September 11, 2001 that was simply a full-page American flag. See N.Y. TIMES, Sept. 16, 2001, at 24.

196 But see Citizens United, 130 S. Ct. at 977-79 (Stevens, J., dissenting).
Second, the Act would cover media corporations because it provides no exemption for such entities.\footnote{See Jones, supra note 2, at 305-08. For a discussion of the media/non-media corporation divide in campaign finance jurisprudence, see Joel M. Gora, The First Amendment . . . United, 27 Ga. St. U. L. Rev. 935, 955-57 (2011).} The Citizens United Court cautioned that any bill that hinders the ability of a media corporation to engage in political speech will almost certainly fail the Court’s strict scrutiny analysis.\footnote{See Citizens United, 130 S. Ct. at 906 (“There is simply no support for the view that the First Amendment, as originally understood, would permit the suppression of political speech by media corporations.”).} This proposal runs afoul of that principle by defining “public company” in a manner that would encompass any media organization that is owned by a publicly traded corporation.\footnote{See id. For example, the New York Times, the Wall Street Journal, and all major American television news networks are owned by publicly traded corporations. The bill defines a “public company” as “any issuer that is required to submit periodical or other reports under section 13 of the Securities Exchange Act of 1934 (15 U.S.C. 78m).” H. R. 4487 § 3(1).}

Third, the Act would likely cover some independent expenditures made for the purpose of lobbying.\footnote{It is likely that courts would interpret the bill to cover some lobbying expenses since support of a tax-related bill, immigration reform, or healthcare legislation would presumably be “not related to the company’s products or services.” H.R. 4487 § 2.} The Citizens United Court stated that although mandating disclosure of lobbying expenditures is constitutional, the Constitution does not tolerate other restrictions on lobbying.\footnote{See Citizens United, 130 S. Ct. at 915.} Many corporations have in-house employees, consultants, and lawyers that lobby federal, state, and local governments on a range of regulatory and appropriations-related issues. In addition, corporations often contribute funds to trade associations, such as the Recording Industry Association of America or the Pharmaceutical Research Manufacturers of America, in order to support lobbying efforts on behalf of their industry. The Court would likely find a barrier of mandatory shareholder approval between a corporation and policymakers or regulators impermissible.

Therefore, the Act is overinclusive in that it encompasses media corporations and covers vastly more speech than independent political expenditures, including goodwill advertising campaigns and lobbying activities. Due to this overinclusiveness, the Shareholder Funds Act fails the “fit” analysis set forth in Citizens United and is not “narrowly tailored” to the shareholder-protection interest.
2. Shareholder Protection Act

Similar to the Shareholder Funds Act, the Shareholder Protection Act will likely fail the “narrowly tailored” analysis, because it contains two aspects that render it overinclusive. First, the exemption for media corporations is extremely narrow. While the New York Times Company’s “sole business” is the publication of news (and so it would be exempted from the Act), the companies that own and operate ABC News or The Colbert Report, for example, would not be exempted, because they operate businesses that are not dedicated to publication of news or related purposes. Therefore, the Shareholder Protection Act would impermissibly subject many significant media outlets in the United States to rigorous shareholder-approval procedures. As noted above, the Citizens United Court made it clear that even the potential for regulation of political speech or expenditures of media corporations would almost certainly fail the Court’s strict scrutiny analysis.

Second, the Shareholder Protection Act is overinclusive because it would apply not only to independent expenditures in federal elections, but also to independent expenditures made with regard to state or local elections. This broad and explicit application reaches far beyond the expenditures that the BCRA sought to restrict and raises serious federalism concerns. The Court, under Citizens United or other Supreme Court precedent, would likely invalidate this provision because it exceeds the power granted to Congress. The BCRA did not venture to regulate corporate expenditures made with regard

\[\text{References}\]

204 The New York Times Company is the owner of the New York Times and several other newspapers. ABC News is owned by The Walt Disney Company and The Colbert Report is aired on Comedy Central, which is owned by Viacom. The Walt Disney Company and Viacom both have divisions or subsidiaries that would not be considered publishers or broadcasters of “news, commentary, literature, music, entertainment, artistic expression, scientific, historical or academic works, or other forms of information.” Id. For example, The Walt Disney Company owns and operates theme parks and Viacom owns video-gaming websites. Neither of these examples constitutes publishing or broadcasting of any kind.
205 See Citizens United, 130 S. Ct. at 906; see also supra Part IV.C.1.
206 H.R. 4537 § 3(e)(4)(A)(i).
207 But see Torres-Spelliscy, Corporate Political Spending, supra note 23, (manuscript at 66-69).
208 Burdick v. Takushi, 504 U.S. 428, 433 (1992) (holding that the power to regulate state elections is vested in the states, not Congress, pursuant to the Elections Clause); see also U.S. CONST. art. I, § 4, cl. 1.
to state or local elections, and the broad sweep of this proposal crosses the line between what Congress can and cannot regulate.\footnote{209} Therefore, the Shareholder Protection Act is overinclusive because it would not exempt many media corporations and would unprecedentedly expand federal campaign-finance law to regulate corporate expenditures in state and local elections. As a result of this overinclusiveness, the Shareholder Protection Act fails the “fit” analysis set forth in \textit{Citizens United}, and the Court will likely find it not “narrowly tailored” to the shareholder-protection interest.

3. Brennan Center for Justice Proposal

The Brennan Center Proposal has the same core element as the Shareholder Funds Act and the Shareholder Protection Act—it would require “corporations [to] obtain the consent of shareholders before making political expenditures.”\footnote{210} The proposal defines “political activity” as “any contributions or expenditures made directly or indirectly to, or in support of or opposition to, any candidate, political party, committee, electioneering communication, ballot measure campaign, or an issue advocacy campaign.”\footnote{211} While the Brennan Center Proposal does exempt expenditures on lobbying from coverage, it makes no exception for media corporations.\footnote{212} Although it is the best crafted of the three proposals, the Brennan Center Proposal is still likely to fail the “narrowly tailored” analysis because it contains elements that render it both over- and underinclusive.

Two aspects of the Brennan Center Proposal are overinclusive. First, the proposal would apply equally to publicly owned nonmedia and media corporations alike.\footnote{213} In \textit{Citizens United}, the Court found that even the mere possibility of restricting media corporations would ensure constitutional

\begin{itemize}
  \item \textit{BCRA} prohibited corporations from expending general treasury funds for “electioneering communications,” which were defined as broadcast advertisements that “refer to” a federal candidate for a period of time directly before a primary or general election. L. \textsc{Paige} \textsc{Whitaker}, \textsc{Cong}. \textsc{Research} \textsc{Serv.}, 1898025, \textsc{Campaign Finance: Constitutional and Legal Issues of Soft Money} 1 (2003), available at http://fpc.state.gov/documents/organization/28106.pdf. It is likely that it is beyond the enumerated powers granted to Congress by the U.S. Constitution to regulate campaign financing with regard to state or local elections. See \textsc{U.S. Const. art. I, § 4, cl. 1}; \textsc{Burdick}, 504 U.S. at 433.
  \item \textsc{Torres-Spelliscy}, \textsc{Corporate Campaign Spending}, supra note 18, at 21.
  \item \textsc{Torres-Spelliscy}, \textsc{Corporate Political Spending}, supra note 23, (manuscript at 72).
  \item See id.
  \item See id. at 71-77.
\end{itemize}
invalidation of any campaign-finance legislation.\textsuperscript{214} The proposal
does not attempt to exempt media corporations nor does it
address this point of overinclusiveness as a known shortcoming.
This aspect of the proposal casts too wide a net and would
encompass every major media corporation in a way that the
Court likely will find impermissible under the First
Amendment.

Second, the Brennan Center Proposal would restrict a
corporation’s ability to conduct an “issue advocacy campaign”
which, as defined, is overly broad because it could potentially
include activity from which shareholders need no protection.\textsuperscript{215} The
proposal would define “issue advocacy campaign” as “contributions
or expenditures for any communication to the general public
intended to encourage the public to contact a government official
regarding pending legislation, public policy or a government rule or
regulation.”\textsuperscript{216} Although corporations and free speech advocates
may raise issues based on the straightforward result of this
definition, the potentially unintended—but nonetheless likely—
results are of greatest concern.

For example, the Brennan Center Proposal would not
allow the host of The Colbert Report to spend time on his show to
discuss the “Don’t Ask, Don’t Tell” law in a way that might be
reasonably interpreted as encouraging the viewing public to
contact their government representatives.\textsuperscript{217} Certainly, the host of
the show is well-compensated, and Comedy Central along with its
corporate parent, Viacom, has spent more than $10,000 to produce
and air the show. Therefore, in a fair reading of the proposal’s text,
it would seem that this expenditure would need shareholder
approval.

Another example of this likely result could involve a
situation where a corporate leader speaks out on behalf of (or
in opposition to) proposed legislation. On March 27, 2009, chief
executive officers (CEOs) of the nation’s several major banks
met with President Obama at the White House.\textsuperscript{218} If there had

\textsuperscript{215} See Torres-Spelliscy, Corporate Political Spending, supra note 23,
(manuscript at 72). This may also be a problem with the other two proposals, although
their definitions are not clear enough for a proper analysis. This definitional problem
may also raise an issue of potential constitutional vagueness.
\textsuperscript{216} Id.
\textsuperscript{217} See The Colbert Report (Comedy Central Television broadcast Dec. 7,
december-07-2010/poll-to-repeal-don-t-ask--don-t-tell.
\textsuperscript{218} Eamon Javers, Inside Obama’s Bank CEOs Meeting, POLITICO (Apr. 3,
been a press conference afterward and one or more of the CEOs had spoken in support of a regulation or proposed legislation, such speech may have constituted a violation of the proposal. Although news networks may broadcast the message free of charge, the CEO likely flew, at the company’s cost, on a very expensive private plane to the White House and spent corporate time and money preparing for the meeting and the press conference. As long as the CEO speaks in his or her official corporate role, the corporation pays him or her to deliver that message. Again, a literal and straightforward reading of the proposal yields results that go well beyond addressing independent expenditures and are far broader than requirements intended to protect the interest of the shareholders.

In addition to the Brennan Center Proposal’s overinclusiveness, another aspect of the proposal is underinclusive. The proposal does not cover charitable contributions, independent charitable expenditures, or speech supporting a charitable cause, all of which share many of the same characteristics as independent political expenditures and contain a similar risk of contravening shareholder interests. Like independent political expenditures, charitable expenditures are an activity in which “managers of publicly-traded companies spend . . . using other people’s money—in part, money invested by shareholders.”

In addition, charitable expenditures are similar to political expenditures, because they are unlikely directly related to the products or services offered by the corporation. At a high level, the aims of charitable and political expenditures are similar—they impact arenas outside of the market in which the corporation operates. If the Brennan Center Proposal covered political expenditures while not regulating charitable expenditures, then it would specifically single out political expenditures for disfavor. Further, it would limit the scope of this regulation so that it would not be broad enough and therefore inappropriately tailored to the shareholder-protection interest.

219 The non-inclusion of independent charitable expenditures is potentially an exception that could swallow the rule. Many political expenditures could be cast as charitable expenditures, as can be seen in the complicated relationship (and overlap) between 501(c)(3) and 501(c)(4) nonprofit organizations.

220 Torres-Spelliscy, Corporate Campaign Spending, supra note 18, at 9 (citation omitted).

221 See Gilson & Klausner, supra note 24 (“Typically [charitable contributions] are uncontroversial, providing support to schools, art museums, United Way and the like in communities where the corporation does business. A willful chief executive can use corporate funds to make contributions to pet causes, but this type of behavior is rare and, like other self-interested dealing, constrained by ordinary board oversight.”).
Thus, the proposal is overinclusive because it would not exempt media corporations and it would inhibit corporations from undertaking typical public-policy advocacy. In addition, the proposal is underinclusive because it would not regulate charitable speech, contributions, or expenditures. As a result of this over- and underinclusiveness, the Brennan Center Proposal fails the “fit” analysis set forth in Citizens United, and the Court is unlikely to find it “narrowly tailored” to the shareholder-protection interest.

D. Content-Neutrality Analysis

The “fit” analysis under Citizens United’s conception of shareholder protection as a compelling governmental interest is the most pertinent analytical tool available to critique the mandatory shareholder-approval proposals. However, a secondary constitutional argument may be made with regard to the proposals; namely, they violate the content-neutrality principle, a fundamental tenet of the First Amendment. The seminal case Police Department of Chicago v. Mosley articulated the content-neutrality principle. In Mosley, a protester challenged an ordinance that disallowed nonlabor picketing, but allowed labor-related picketing, in front of schools. The Court held the ordinance to be unconstitutional because it made an impermissible distinction between labor-related picketing and other peaceful picketing. Further, the Court stated that “the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.”

The mandatory shareholder-approval proposals violate Mosley’s content-neutrality principle because they restrict only certain speech based on content. For example, the proposals do nothing to inhibit charitable speech or other types of speech that may be unrelated to the products and services of the company. These proposals would not require shareholder approval for corporate management to spend money on contributions to

---

223 See Police Dep’t of Chi. v. Mosley, 408 U.S. 92 (1972). Although Mosley is a seminal case exemplifying the Court’s treatment of the content-neutrality principle, it has been repeatedly enshrined by the Court as a central tenet of the First Amendment. See Fallon, supra note 222, at 20.
224 Mosley, 408 U.S. at 92-93.
225 Id. at 94.
226 Id. at 95.
charitable organizations, encourage the public to donate or volunteer for charitable organizations, or even run advertisements supporting corporate management's favorite sports teams.\textsuperscript{227} Still, when it comes to political speech, under these proposals, corporate management would be required to seek shareholder approval and follow very specific regulations. The Mosley Court prohibited "[s]elective exclusions from a public form . . . based on content alone,"\textsuperscript{228} such as the prohibition on running advertisements about political issues without shareholder approval but allowing advertisements about charitable or other pet causes without restriction of any kind.

Although the content-neutrality analysis would vary slightly between the three mandatory shareholder-approval proposals, the main thrust of the proposals is identical. Because "[a]ny restriction on expressive activity because of its content . . . completely undercuts the profound national commitment to the principle that debate on public issues should be uninhibited, robust, and wide-open,"\textsuperscript{229} any significant restriction of a corporation's ability to make independent political expenditures is impermissible under the content-neutrality principle of the First Amendment. Because Citizens United placed political speech at the core of the First Amendment's protection and Mosley identified the content-neutrality principle as fundamental, these proposals are ripe for invalidation by the Court.

**CONCLUSION**

Many legislative and regulatory responses to Citizens United have been proposed and the three mandatory shareholder proposals discussed in this note are certainly among the most well-intentioned and carefully constructed. Although they would have the potential to ameliorate the effects of Citizens United regarding independent political expenditures by corporations, they run afoul of that very decision's interpretation of the First Amendment.

Even if the Court finds shareholder protection to be a compelling governmental interest, none of the three proposals is narrowly tailored to that interest. The shareholder democracy

\textsuperscript{227} While there may be limitations on such activities from corporate law, they are immaterial for this note's constitutional analysis.

\textsuperscript{228} Id. at 96.

\textsuperscript{229} Id. (quoting N.Y. Times Co. v. Sullivan, 376 U.S. 254, 270 (1964)).
and “fit” analysis, articulated with regard to the shareholder-protection interest in Citizens United, does not permit any of the three mandatory shareholder-approval proposals to exist within current First Amendment jurisprudence. In addition, the proposals likely violate the content-neutrality rule by placing significant burdens on corporations wishing to exercise their political rights while failing to place similar restrictions on other forms of corporate speech.

As a result of these constitutionally doomed proposals, PepsiCo may uninhibitedly support a potential campaign by Mr. Colbert for President in 2012 and similarly, Viacom will not have to seek shareholder approval before airing an episode of The Colbert Report that features Mr. Colbert’s candidacy. In the new legal and constitutional landscape of Citizens United, corporations have unprecedented constitutional protection for political activities. Although scholars and legislators may continue to craft creative responses, the barriers to restricting corporate political speech remain stronger than ever.

Robert B. Sobelman

230 Although a candidacy by Mr. Colbert in 2012 looks increasingly unlikely, he has remained active in presidential politics and devoted himself to pushing the boundaries of campaign-finance law and the First Amendment. See David Carr, Comic’s PAC Is More Than a Gag, N.Y. TIMES, Aug. 22, 2011, at B1.

† J.D. Candidate, Brooklyn Law School, 2012; B.A., Colgate University, 2008. I am grateful to Professor Stanley Brubaker of Colgate University for introducing me to the First Amendment, and to Paul Pimentel and former congressman Christopher Shays, co-author of the BCRA, for sparking my interest in campaign-finance law. I would like to thank Professor William Araiza of Brooklyn Law School for helping me explore this legal issue and for guiding me during the drafting of this note. In addition, I am thankful for the wise editorial advice of Lindsey Champlin, Kristie LaSalle, and Susan Nabet. I would also like to thank my mother, father, and brother, without whom none of my accomplishments would be possible. Most importantly, I am grateful for the unwavering support of my endlessly talented wife, Deborah.
Protecting a Right to Access Internet Content

THE FEASIBILITY OF JUDICIAL ENFORCEMENT IN A NON-NEUTRAL NETWORK

INTRODUCTION

If information is a weapon for change, then the Internet arms every man, woman, and child on the planet. Now more than ever, the disruptive power of viral mass-communication is palpable.

The spread of information networks is forming a new nervous system for our planet. When something happens in Haiti or Hunan, the rest of us learn about it in real time—from real people. And we can respond in real time as well. . . . As we sit here, any of you—or maybe more likely, any of our children—can take out the tools that many carry every day and transmit this discussion to billions across the world.¹

This cosmopolitan nervous system manifested itself in early 2011, when the Egyptian citizenry used Facebook to organize thousands in Tahrir Square to engage in antigovernment protest.² As a defensive measure, the Egyptian government took the rare and startling step of “switching off” Internet connectivity for its eighty million residents.³ Concurrently, embassy cables disclosed on WikiLeaks exasperated the uprising in Tunisia when the population

---

discovered the “mafia-esque” corruption that yielded “massive profits” for the nation’s government elite.  

Reliance on Internet connectivity obviously extends well beyond national protest; reliance presents itself several times over in varying facets of everyday life, especially in the nascence of ubiquitous computing. Over a quarter of the world’s population accesses the Internet.  

Unsurprisingly, an international sentiment has emerged clearly indicating that Internet access is desired by all. And of course, the public’s sentiment is not powerless: in June 2009, France’s Constitutional Council denied President Sarkozy’s power to create an Internet police force and ruled that Internet access is a basic human right. The Council opined, “In the current state of the means of communication and given the generalized development of public online communication services and the importance of the latter for the participation in democracy and the expression of ideas and opinions, this right implies freedom to access such services.” In recent history, the United Nations has warned the

---

7 Charles Brenner, Top French Court Rips Heart Out of Sarkozy Internet Law, TIMES (London) (June 11, 2009), http://technology.timesonline.co.uk/tol/news/tech_and_web/article6478542.ece.  
international community that cutting off Internet access to quell protest is a human rights violation.\(^9\)

While some members of the U.S. Supreme Court may be unwilling to recognize the weight international trends—even mandates—should have on U.S. lawmaking,\(^10\) U.S. citizens must begin to consider when a central function to daily living becomes something fundamental or guaranteed. Indeed, Americans consider the Internet as (if not more) important to their lives as the rest of the world. In 2009, the Federal Communications Commission (FCC) reported that the average American household user consumed over nine gigabytes of data per month.\(^11\) The number of Americans demanding high-speed Internet access appears to be accelerating very rapidly as well.\(^12\) In the wake of U.S. economic turmoil, Congress mandated the implementation of an initiative to make broadband services more accessible to Americans.\(^13\) Perhaps an indicator of increased broadband use, Facebook (an American-born company) has seen its membership base grow exponentially.\(^14\)

Beyond use as a forum for expression and a database for information, the Internet has become home to increasingly varied application across the spectrum of telecommunication. The Internet protocol (IP) suite is now used for phone services


\(^10\) See Lawrence v. Texas, 539 U.S. 558, 598 (2003) (Scalia, J., dissenting) (“Constitutional entitlements do not spring into existence because some States choose to lessen or eliminate criminal sanctions on certain behavior. Much less do they spring into existence, as the Court seems to believe, because foreign nations decriminalize conduct.”); Roper v. Simmons, 543 U.S. 551, 622-28 (2005) (Scalia, J., dissenting) (criticizing the majority’s use of international law and trend to reach its decision).

\(^11\) F ED. COMMC’NS COMM’N, supra note 5, at 4.


\(^14\) In July 2011, Facebook clocked 750 million registered users, skyrocketing from 500 million users only a year prior. Leena Rao, Zuck Confirms that Facebook Now Has 750 Million Active Users, TECHCRUNCH (July 6, 2011, 1:32 PM), http://techcrunch.com/2011/07/06/zuck-confirms-that-facebook-now-has-750-million-users/.
through Voice-Over Internet Protocol (VoIP) technology,\textsuperscript{15} IP television is quickly gaining its footing in the viewership market,\textsuperscript{16} and the demand for mobile broadband has greatly increased with the advent of smartphones, netbooks, tablets, and other ubiquitous computing devices.\textsuperscript{17}

One need not be terribly tech-savvy to conclude that much of humanity now relies upon Internet connectivity to retrieve and deliver content. The question emerges: ought the United States follow the worldwide trend to afford its citizens a right to access the lawful Internet content of their choice? Reserving for others the question of whether such a right should be fundamental under U.S. law, the most appropriate placeholder for a right to access Internet content, at least prima facie, is the First Amendment—the Internet is fundamentally a form of communication, after all.\textsuperscript{18} Theorists, scholars, and the citizenship-at-large must inquire, however, what the contours of such a recognized right would be, and—more importantly—how it can be preserved. The foundation of the latter inquiry goes beyond consumer-centered questions—how the government can deploy common carriage in the Internet age or how state actors can ensure that Internet service providers (ISPs) adhere to their terms of service. It lies in the more fundamental (and essential) question of how individuals can be assured access to the content of their choice when their ISPs and government have failed to account for that interest.

This fundamental question could not be riper for discussion. Current law and market realities do not easily accommodate a right or guarantee of access to Internet content.


\textsuperscript{16} Chris Davies, Smart TV Could Overwhelm the Internet Warns Analyst, SLASHGEAR (Dec. 30, 2010), http://www slashgear.com/smart-tv-could-overwhelm-theinternet-warns-analyst-30122009/.

\textsuperscript{17} FED. COMM'NS COMM'N, supra note 5, at 19. The FCC has recognized the increasing applications of the Internet, and has chosen to incorporate the above-mentioned applications in its rules on Internet openness. Preserving the Open Internet: Broadband Industry Practices, 25 FCC Rcd. 17,905, 17,932-33 (Dec. 21, 2010) (report and order) [hereinafter Net Neutrality Order].

\textsuperscript{18} U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech."); see also Reno v. ACLU, 521 U.S. 844, 868 (1997) (concluding that the limitations on Internet communication for decency under Congress's Communications Decency Act was a blanket prohibition that chills free speech). Courts and some scholars, unsurprisingly, have discussed this theory already. See generally Moran Yemini, Mandated Network Neutrality and the First Amendment: Lessons from Turner and a New Approach, 13 VA. J. L. & TECH. 1, 2 (2008); Christopher S. Yoo, Free Speech and the Myth of the Internet as an Unintermediated Experience, 78 GEO. WASH. L. REV. 697 (2010); see also infra note 23.
Such questions, which strongly seem to implicate constitutional liberties, are directly affected by network neutrality, an issue that is transforming Internet accessibility’s legal framework in the United States.\(^{19}\) Network (net) neutrality is a movement toward “the non-discriminatory interconnectedness among data communication networks that allows users to access the content, and run the services, applications, and devices of their choice.”\(^{20}\) The movement has emerged in the face of strong indications that the Internet is becoming (or has become) a privately regulated infrastructure where ISPs have the power to impede accessibility to the information exchanged over their networks.\(^{21}\) Bafflingly, the rise of private control over the Internet has been met with the fall of FCC oversight, even though the agency has for decades been tasked with maintaining nondiscriminatory access to telecommunications.\(^{22}\)

---


Think about the electricity grid. Alright, when you plug a television into the electricity grid, it doesn’t ask, “Is it a Sony television or a Panasonic television?” It doesn’t ask, “Is it a toaster made in America or a toaster made in Japan?” It just runs. And that’s because the electricity grid is a neutral network in this sense. You comply with the protocols—what the plug’s got to look like and how much power you’re taking—and it runs. That’s the way the internet was. It used to be it didn’t matter whether it was a browser made by Microsoft or a browser made by Netscape or a browser made by Mozilla. It just ran because the protocols said if you follow the rules, the system will run.


\(^{21}\) See infra Part I.

\(^{22}\) See infra Part I.A; see also Communications Act of 1934, 47 U.S.C. § 151 (2006) (“For the purpose of regulating interstate and foreign commerce in communication by wire and radio so as to make available, so far as possible, to all the people of the United States, without discrimination on the basis of race, color, religion, national origin, or sex, a rapid, efficient, Nation-wide, and world-wide wire and radio communication service with adequate facilities at reasonable charges,…there is created a commission to be known as the ‘Federal Communications Commission’. . .”).
The impending degradations to Internet services imperil the freedom to access the most widely used forum for public expression in the world.\(^{23}\)

This note tackles the net neutrality problem from a practical perspective by suggesting and assessing a countermajoritarian regulatory regime that checks administrative and congressional action—a private cause of action based in constitutional rights. Rather than to solely outline the constitutional theory that supports net neutrality, the overarching purpose here is to answer whether, assuming the regulatory status quo, the federal court system is equipped to take on the role as arbiter of Internet-content-access disputes.

In making that assessment, this note considers the multiple layers of constitutional protection that relate to the Internet-access issue—the freedom of expression and the right to access for end-users\(^ {24}\) as well as the freedoms of expression and property for ISPs.\(^ {25}\) The model of judicial enforcement assessed here can appropriately weigh individual and intermediary interests by adjudicating the reasonableness of network management on a contextual, case-by-case basis. This note concludes that there is ample existing doctrine to direct these case-by-case inquiries. But there are obvious hurdles to directly applying these constitutional standards to end-user litigation—namely, the state action doctrine and standing. Thus in conclusion, this note offers specific recommendations for legislation to facilitate the shift to judicial enforcement as a mode of Internet regulation that preserves the edges’ rights to access.\(^ {26}\) Part I discusses the contours of the net neutrality debate and summarizes the nation’s policy regarding Internet accessibility. Part II describes the shortcomings of other

\(^{23}\) Communication theory coined by Marshall McLuhan states “the medium is the message.” Mark Federman, What Is the Meaning of the Medium Is the Message?, UNIV. OF TORONTO (July 23, 2004), http://individual.utoronto.ca/markfederman/article_mediumisthemessage.htm. That is, a medium’s character is not the content it conveys, but the effect the medium itself has on society. Id. Here, content’s accessibility by Internet users is not the legal issue implicated; the issue is the functionality of the medium in conveying its impact. As related by Jerome Barron, “The new modes of communication engage us by their form rather than by their content; what captivates us is the television screen itself.” Jerome A. Barron, Access to the Press—A New First Amendment Right, 80 HARV. L. REV. 1641, 1645 (1967); see also Yemini, supra note 18, at 15 (“[I]n order to ‘reach’ the logical and content layers, one has to ‘pass through’ the physical layer; whoever controls the physical layer, unless restricted by law, becomes a gatekeeper for all other layers . . . .”)

\(^{24}\) See infra Part III.B.1.

\(^{25}\) See infra Part III.B.2.

\(^{26}\) See infra Part IV.
available enforcement mechanisms, including administrative rulemaking and legislative action. Part III outlines the legal standards potentially applicable to constitutional litigation over a theoretical right to access Internet content and highlights the benefits of judicial enforcement. Part III also recognizes the challenges end-user litigants face. Finally, Part IV proposes a private cause of action for end-user litigants.

I. THE FALL OF FEDERAL REGULATION, THE EMERGENCE OF NON-NEUTRALITY

As private control over the Internet’s architecture has increased, the federal entities that would typically check that control have taken the back seat. In the current deregulated environment, the Internet continues to increase in day-to-day importance. This part outlines the legal and historical developments that have created this seemingly paradoxical situation—the shift to broadband infrastructure and the definitional hocus-pocus that has freed ISPs from common carriage regulation—then goes on to discuss the main points of the net neutrality debate. Finally, this part will detail the non-neutral practices that have emerged since the onset of the FCC’s deregulatory approach to Internet access oversight.

A. Creating a Non-neutral Network

In the beginning the Internet was open.

From its inception, the Internet has used a packet-switching system, which was initially nondiscriminatory. Packet-switching has proven an extremely efficient mode of transfer because it allows information, divided into small pieces called packets, to exchange over any conceivable path of routers between Internet-connected terminals. In its early stages, the Internet consisted of a network of “narrowband,” packet-switched networks that were designed such that the

---

27 DAWN C. NUNZIATO, VIRTUAL FREEDOM: NET NEUTRALITY AND FREE SPEECH IN THE INTERNET AGE 19 (2009). During transfer, information sent from a computer terminal is separated into pieces that are then reassembled upon receipt. Id. The tags attached to packets dictate the packets’ destination as well as their source, in addition to identifying information to reassemble the information at the destination. Id.

28 The packet-switching system’s efficiency is due to the nondiscriminatory allotment of pathways between routers (to “empower the individual,” interconnecting router points as a “network of equals”) in contrast to the centralized, “hierarchical” circuit-switching infrastructure used by the AT&T telephone network. TIM WU, THE MASTER SWITCH 173 (2011).
long-distance infrastructure carried high-bandwidth traffic to limited access points. The connection to the end or individual users, though, was carried by separate, “last-mile” providers, which delivered content to central facilities through basic telephone call technology. The old regime was inherently nondiscriminatory before the advent of broadband. First, on the physical level, telephone technology and wire use a simple routing process without any need for intermediary traffic modification. Second, on the logical level, Internet code uses the transmission control and Internet protocols (TCP/IP), which automatically allot service resources to an end-user on a “first-in first-out” or “best efforts” basis, rather than by the needs of an interested intermediary. These structural characteristics limited the entities’ ability to modify and block the traffic running over the last mile.

The incumbent narrowband regime, which (at least initially) carried Internet providers’ services, was also subject to requirements designed to open competition under the Telecommunications Act of 1996. In 1984, the AT&T (Bell) telephone monopoly was ordered to divest its regional operating companies that provided local service. In turn, Congress passed the Telecommunications Act to combat the monopoly that “Baby Bells” enjoyed over the local exchanges and inject competition into those markets. Summarized succinctly, the Act forced the incumbent local exchange carriers

---

29 Last-mile providers were often the local telephone company. Id. But during the rise of the Internet, last-mile providers also successfully remained wholly independent from telephone carriers. See, e.g., id. at 262-63 (describing the AOL’s “walled garden” business model as wholly independent from the services provided by telephone-dial-up Internet). The separation between Internet transmission and services was implemented and maintained through the Nixon and Clinton administrations. Id. at 309.


34 The government argued in its 1974 antitrust suit against AT&T that the company unreasonably restrained trade in the telephone equipment markets in violation of the Sherman Act, Section Two. United States v. AT&T, 552 F. Supp. 131, 139 n.18 (D.D.C. 1982). Judge Greene approved, but modified, the consent decree to divest the Bell Operating Companies from AT&T long distance, noting that AT&T, for years, used its market power over local telephone services to prevent the entry of new competitors in the local exchange and equipment markets. Id. at 223.

35 Wu, supra note 28, at 194.

36 Id.
ILECs) to interconnect with other companies that wanted to use the network, resell their services at reasonable rates, unbundle network elements, and engage in what has become a complex form of payment between carriers called reciprocal compensation. All these requirements were passed upon findings that the local, last-mile providers sat on a natural monopoly, or bottleneck, of information that allowed them to discriminate against competitors trying to enter the market of data carriage. The 1996 Act therefore attempted to limit the extent to which the last mile could be dominated by a single entity—that is, as long as the last mile was a telecommunications service subject to the Communication Act's Title II regulation.

But today, the 1996 Act's common carrier obligations do not apply to home broadband Internet services. Whereas telecommunications services are normally subjected to common carrier requirements under Title II of the Communications Act, because broadband service providers have been deemed information services, they are only subject to the lighter touch of Title I. This definitional dichotomy between computing and telecommunications services was created over the span of several hearings held to address the burgeoning computer market and its convergence with telecommunications. The delineation that emerged in those hearings was between telecommunications and information services and was maintained in the Telecommunications Act of 1996.

In 2002, the information services designation was extended from computing to broadband cable modems in the

---

42 Time Warner Telecom, Inc. v. FCC, 507 F.3d 205, 210-11 (3d Cir. 2007) (quoting Regulatory and Policy Problems Presented by the Interdependence of Computer and Communication Services and Facilities, 28 F.C.C.2d 267, ¶ 8 (1971)). The earliest dichotomy contemplated was one between basic transmission and enhanced services, where basic services were regulated, and enhanced services were not. Id. at 211 (citing Amendment of Section 64.702 of the Commission's Rules and Regulations, 77 F.C.C.2d 384, ¶ 2 (1980)).
43 Brand X, 545 U.S. at 975 (citing 47 U.S.C. § 153(20), (44) (defining information service and telecommunications carrier, respectively)).
FCC’s Cable Broadband Order. The U.S. Supreme Court affirmed the permissibility of that extension in National Cable & Telecommunications Ass’n (NCTA) v. Brand X Internet Services. In Brand X, the Court indicated that the FCC’s decision to regulate cable ISPs as information service providers was an acceptable construction of the Communications Act because cable companies offer an integrated “offering” of Internet services and their transmission, rather than a “stand-alone” transmission service. The Court went on to find that the FCC’s decision to reduce cable Internet access regulation was justifiable due to the “fast-moving [and] competitive market” of Internet services. In other words, the FCC adopted the policy of deregulating the burgeoning Internet-services industry to avoid hindering its progress. As articulated by the FCC, “[E]xisting regulations constrain technological advances and deter broadband infrastructure investment by creating disincentives to the deployment of facilities capable of providing innovative broadband Internet access services.”

Soon after Brand X was handed down, the FCC issued its Wireline Broadband Order, which further extended the information service designation to telephone companies that provide DSL services. In turn, no broadband providers are currently subject to common carrier regulation under Title II. But, oddly, the information bottleneck has further spilled into the realm of hardware as networks have shifted into the broadband regime. This trend is problematic in the rise of broadband, where

---

44 Id. at 978-79 (discussing Inquiry Concerning High-Speed Access to the Internet Over Cable and Other Facilities, 17 F.C.C.R. 4798 (2002)).
45 Id. at 986.
46 Id. at 988.
47 Id. at 977.
48 Id. at 1001 (“The Commission concluded that ‘broadband services should exist in a minimal regulatory environment that promotes investment and innovation in a competitive market.’”).
50 Id. at 14,862.
51 Perhaps towards a consumer desire for easy and secure use, the bottleneck has further pervaded home and mobile information, where providers increasingly “appliancize” the devices we use to access Internet protocol. JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 3 (2008). That is, devices are purchased—often from the same company that offers access—to perform very specific functions with IP resources (consider your mobile phone or cable box). In turn, the ability of intermediaries to constrain end-user behavior is increased. Id. at 8-9. A new wave of proprietary networks in mobile broadband has created a separate chokepoint, where today’s popular mobile devices are only capable of accessing what is available on
intelligent data management hardware is increasingly used between the last mile and the core, and market power in the last mile of broadband continues to expand.

Though the United States has enacted no laws since Brand X that guarantee equal or open access to Internet content, the FCC has nonetheless clearly set a policy goal of digital connectivity. In 2005, the FCC issued a policy statement in an attempt to effectuate the goal of section 230(b) of the Communications Act of 1934, which states, "It is a policy of the United States...to preserve the vibrant and competitive free market that presently exists for the Internet and other interactive computer services, unfettered by Federal or State regulation." The policy statement sought "to encourage broadband deployment and preserve and promote the open and interconnected nature of the public Internet" by adopting four network management principles:

- Consumers are entitled to access the lawful Internet content of their choice...

- [T]o run applications and use services of their choice, subject to the needs of law enforcement...

the Apple App Store or Android Market. See Net Neutrality Order, supra note 17, at 17,925 (noting that a mobile wireless provider prevented users from using online payment options outside provider's contracted service). In turn, the constraints upon mobile hardware are particularly susceptible to the hardware or network provider's remote access to IP-enabled devices. See e.g., Timothy Karr, Is Apple Launching a Pre-Emptive Strike Against Free Speech?, HUFFINGTON POST (June 22, 2011, 8:33 AM), http://www.huffingtonpost.com/timothy-karr/is-apple-launching-a-pree_b_881940.html (reporting an Apple patent on technology that can shut down the iPhone camera remotely).

See, e.g., infra Parts I.C.-D.; see also Yoo, supra note 30, at 32-34 (describing the emergence of data-sorting technology between end-users to properly provide cable/Internet services over cable modems, and telephone/Internet services over DSL).


The notable exception is the FCC's recent Net Neutrality Order. See generally Net Neutrality Order, supra note 17.

See generally Appropriate Framework for Broadband Access to the Internet Over Wireline Facilities, 15 FCC Rcd. 14,986 (Aug. 5, 2005) (policy statement) [hereinafter Broadband Policy Statement]. The FCC's policy of promoting competition over the objection of incumbent technologies goes much further back than the Internet. See Use of the Carterfone Device in Message Toll Tel. Serv., 13 F.C.C.2d 420 (1968). Clearly, the FCC has taken a step back from this stance in the last ten years. Net Neutrality Order, supra note 17, at 18,045 (Statement of Commissioner Michael J. Copps Concurred, FCC 10-201, 2010 WL 5179798, at *7) ("Between 2001 and 2009...the FCC took American consumers on a dangerous deregulatory ride, moving the transmission component of broadband outside of the statutory framework that applies to telecommunications carriers.").


Id. § 230(b)(2) (emphasis added).
• [T]o connect their choice of legal devices that do not harm the network . . .

• [And] to competition among network providers, application and service providers, and content providers.\(^{58}\)

While these principles clearly demonstrate a commitment to the underlying goals of net neutrality, in reality they still pose several problems. First, the Broadband Policy Statement lacks the force of law.\(^{59}\) And second, even if the FCC’s policy statement did have such authority, it lacks any cognizable measures to implement the policies set forth.\(^{60}\) Therefore, as the law currently stands, there is little keeping ISPs from engaging in practices that discriminate against content to the detriment of end-users.

B. The Debate

The arguments for and against net neutrality can mostly be lumped into two fundamental, yet familiar, schools of thought.\(^{61}\) Proponents think that user competition benefits the path of innovation in Internet applications, and thus the government must preserve competition between empowered end-users in light of the growing market power network providers can leverage.\(^{62}\) Detractors believe that self-regulation, and even discrimination by private entities, will not adversely affect competition at the edges and—importantly—may more effectively preserve network economics.\(^{63}\) Some critics also argue that the incentives to engage in business models that discriminate against content or other Internet applications are not as readily obvious as proponents suggest.\(^{64}\) Indeed, AT&T

\(^{58}\) Broadband Policy Statement, supra note 55, at 14,988.

\(^{59}\) Comcast Corp. v. FCC, 600 F.3d 642, 654 (D.C. Cir. 2010) (“Policy statements are just that—statements of policy. They are not delegations of regulatory authority.”).

\(^{60}\) Yemini, supra note 18, at 5. Several developments in FCC action have cropped up since the implementation of the Policy Statement to fill this gap in regulatory authority. See infra notes 114-15, 131-34 and accompanying text. As discussed later, there is little to suggest that these rules will survive litigation. See infra notes 122-25 and accompanying text.

\(^{61}\) Unsurprisingly, the debate in Congress has remained consistent with party affiliations. Robert D. Atkinson & Philip J. Weiser, A Third Way on Network Neutrality, 13 NEW ATLANTIS 47, 49 (2006).


\(^{63}\) Yoo, supra note 30, at 56-59.

\(^{64}\) In part, the power to switch to other ISPs that do not discriminate against the content at issue is a viable choice for edge users, and would weigh against an ISP’s
has gone as far as claiming that the net neutrality issue is a "solution without a problem."\(^{65}\) Tim Wu, a proponent, characterizes neutrality's central premise as "Darwinian": only the "fittest" applications will survive the competition between developers.\(^{66}\) The argument comports with Schumpeterian "creative destruction," the frequently invoked theory in technology policy premised on the notion that competitive innovation tends to build on and destroy preceding norms chaotically, yet progressively.\(^{67}\) Schumpeterian
decision to engage in that type of practice. Becker et al., supra note 31, at 502. For this reason, among others, some describe neutrality regulation as "a solution in search of a problem." Lyons, supra note 62, at 67. The FCC notes, however, that the ability to switch providers may not truly remedy the problem, where users may have limited access to broadband providers, and the cost of switching may be prohibitive. Net Neutrality Order, supra note 17, at 17,921.

\(^{65}\) Grant Gross, AT&T Says It Didn't Censor Pearl Jam, PC WORLD (Aug. 9, 2007, 1:00 PM), http://www.pcworld.com/article/135767/atandt_says_it_didnt_censor_pearl_jam.html.

\(^{66}\) Wu, supra note 62, at 145-46; Ex Parte Submission in CS Docket No. 02-52 from Tim Wu, Assoc. Professor, Univ. of Va. Sch. of Law, and Lawrence Lessig, Professor of Law, Stanford Law Sch., to Marlene H. Dortch, Secy, FCC 3 (Aug. 22, 2003), available at http://www.timwu.org/wu_lessig_fcc.pdf. Google is an appropriate example. Google began in 1997 as a search engine quickly regarded as having "an uncanny knack for returning extremely relevant results." Google History, Google, http://www.google.com/about/corporate/company/history.html (last visited Sept. 9, 2011). Google now offers a wide array of Internet applications—including but not limited to search functions, word processing, e-mail, social networking, and mapping. See GOOGLE, http://www.google.com (last visited July 24, 2011). Google has, in turn, dramatically changed the way end-users access web content. For example, Google was in large part responsible for the paradigm shift to "cloud computing," which, as Google's current website structure demonstrates, allows users' applications and data to be stored remotely, then accessed from any location in the world with no more than a username and password. See Steve Lohr, Google and I.B.M. Join in 'Cloud Computing' Research, N.Y. TIMES (Oct. 8, 2007), http://www.nytimes.com/2007/10/08/technology/08cloud.html. Lawrence Lessig agrees that the value of Google is due in part to the Internet's neutrality:

Now, it's because at no stage did they have to ask permission from the network owner that they've been able to do this. If, at the very beginning, Larry—Sergey Brin and Larry Page had to go to the existing network owners at the time, AT&T, for example, and say, "May we develop this new technology for your network?" it would have taken years for the company, AT&T, to even figure out whether this was going to be permitted, just like if they had gone to a cable company and said, "We want to open a new cable station on your network," it would take forever to get that permission.

Lessig Interview, supra note 20; see also Net Neutrality Order, supra note 17, at 17,907 ("The Internet is a level playing field.").

economics appears to suggest that innovation and economic progress are protected through mandated neutrality.\(^{68}\)

According to neutrality activists, ISPs are incentivized to limit Internet access in order to prevent the utilization of competitive products or costly content, because they wield advantages in technology and law.\(^{69}\) These incentives trouble neutrality proponents in how the consequential practices would flout the benefits of an “end-to-end”\(^{70}\) design, undermining the “dumb” or nondiscriminatory Internet structure.\(^{71}\) Today, three industry practices stand at the forefront of neutrality literature: transparency, blocking, and tiering.\(^{72}\)

Professors Tim Wu and Lawrence Lessig note two important ways that mandated neutrality benefits the Internet as a medium. First, treating applications alike makes the market “predictable,” and therefore, incents the development of—and investment in—broadband applications.\(^{73}\) Like electricity, the Internet is a “general purpose technology.”\(^{74}\) Wu and Lessig note that in the current market for electricity, electronics manufacturers can design new products with peace of mind knowing that their products will work; “the uniformity of the electric grid is a safeguard against the risk of restrictions and uneven standards” that would give the electric company the power to discriminate against new products.\(^{75}\)

\(^{68}\) Wu, supra note 62, at 145 n.10.

\(^{69}\) Some congressional representatives have come to this conclusion: “Internet access service providers have an economic interest to discriminate in favor of their own services, content, and applications and against other providers.” Internet Freedom Preservation Act of 2009, H.R. 3458, 111th Cong. § 2(10) (2009). For example, several ISPs have engaged in full-scale blocking of VoIP technology, which allows users to make phone calls over the Internet. NUNZIATO, supra note 27, at 9-10.

\(^{70}\) The “end-to-end” argument was drafted by Jerome Saltzer, David Reed, and David Clark, and “counsels against introducing intelligence into the core of the Internet.” Yoo, supra note 30, at 41. Under the theory, the better system is one that checks for errors only at the origin and destination of packet transmission—end-to-end. Id. See generally J.H. Saltzer et al., End-to-End Arguments in System Design, 2 ACM TRANSACTIONS ON COMPUTER SYST. 277 (1984), available at http://web.mit.edu/Saltzer/www/publications/endtoend/endtoend.pdf.

\(^{71}\) See Yemini, supra note 18, at 1 (explaining that through “technological, economic, and legal factors,” ISPs can now control the stream of data transmission and that data transmission was formerly controlled by the end-users themselves).

\(^{72}\) Atkinson & Weiser, supra note 61, at 49-50.

\(^{73}\) Wu & Lessig, supra note 66, at 3. The FCC noted that “[n]ovel, improved, or lower-cost offerings introduced by content, application, service, and device providers spur end-user demand and encourage broadband providers to expand their networks and invest in new broadband technologies.” Net Neutrality Order, supra note 17, at 17,911.

\(^{74}\) Net Neutrality Order, supra note 17, at 17,909.

\(^{75}\) Wu & Lessig, supra note 66, at 3. Congress seems to agree with this analogy. See H.R. 3458, 111th Cong. § 2(2) (2009) (“The Internet is an essential
investment in new Internet applications and infrastructure developments will be stabler if the resource is open and ISPs cannot block “undesirable” applications.

Second, neutrality promotes the policy of innovation among applications. There is some speculation, given the current climate of law and technology, that ISPs will shift toward a tiered business model—one that charges fees when there is too much congestion on the network or that charges content providers for edge-user access to their sites. Such tiered access to content could prevent innovators from creating new uses for the Internet. For instance, if a search engine’s accessibility were treated more favorably than other sites on a network, developers would be incentivized to continue providing new Internet search features, without necessarily developing new media-streaming applications. The range of possible Internet applications would therefore be limited, and the benefits to end-user access decreased.

In response, those opposed to net neutrality regulations argue that “prophylactic” regulations could limit an ISP’s incentive and resources to invest in new infrastructure. Though the neutrality agenda would further a right to access Internet content, vying for complete neutrality ignores the fact that there are inherent trade-offs between mutually exclusive network design characteristics. That is, if internetworks prioritize connectivity, network providers may sacrifice the quality of service (QoS) applications necessary to access high-bandwidth content without latency. Some have therefore

---

76 Wu & Lessig, supra note 66, at 5.
77 Becker et al., supra note 31, at 501.
78 The World Wide Web, for example, was created almost twenty years after the development of Internet protocols. Net Neutrality Order, supra note 17, at 17,910. If the Internet’s initial development was stifled by network providers’ intermediation, the existence of the World Wide Web may not have become a reality. Furthermore, “[r]estricting [the] edge providers’ ability to reach end users, and limiting end users’ ability to choose which edge providers to patronize, would reduce the rate of innovation at the edge and, in turn, the likely rate of improvements to network infrastructure.” Id. at 17,911.
79 Professors Wu and Lessig note a similar example in online gaming. In short, under the current system ISPs are inclined to prohibit or disincentivize the use of popular online gaming applications because of the large amount of bandwidth they occupy. "If carriers choose to block online games in particular, this gives a market advantage to competing applications that have not been blocked." Wu & Lessig, supra note 66, at 15.
80 Atkinson & Weiser, supra note 61, at 49. But see ZITTRAIN, supra note 51, at 105 (noting that maintaining the “generative Internet” has historically allowed for technology to overcome “blunderbuss technology regulation”).
81 Wu, supra note 62, at 148-49.
argued that forms of data discrimination are a viable remedy toward efficiency. There is certainly very little debate over whether network providers ought to engage in discriminatory practices when it comes to detecting harmful information packets, such as viruses. The contentious question is what constitutes “reasonable network management” within the context of the current deregulated market, and how far an ISP may go in invading and prioritizing the content an end-user is uploading or downloading.

C. Comcast v. FCC

The landscape of Internet access law reached the apex of deregulation in 2010’s Comcast v. FCC decision from the D.C. Circuit Court of Appeals. The facts of Comcast show—unequivocally—that when left to its own devices, the corporate intermediary has the ability and the incentive to impede end-user access. The holding, on the other hand, may be the final straw in stripping the FCC of its power to regulate network management and, in turn, content access.

In 2007, the Associated Press released a report confirming through “nationwide tests” that Comcast was engaging in data discrimination. By impeding traffic, Comcast kept peer-to-peer applications from “swallowing” bandwidth and thereby limiting the Internet experience of other subscribers. Comcast admitted to prioritizing service for this reason, but only after investigation.

---

82 See generally Yoo, supra note 18.
83 See Wu, supra note 62, at 150-51; see also Internet Freedom Preservation Act, S. 215, 110th Cong. § 12 (2007) (Congress proposing to mandate neutrality except when “protecting the security of a user’s computer on the network”).
85 600 F.3d 642 (D.C. Cir. 2010).
87 Id.; Comcast, 600 F.3d at 644. One peer-to-peer application affected was BitTorrent. Id. BitTorrent, as explained by the FCC, puts strain on the network because of its untraditional method of sharing information. See Formal Complaint of Free Press and Public Knowledge Against Comcast Corporation, 23 FCC Rcd. 13,028, 13,029 (Aug. 1, 2008) (decision and order), vacated, Comcast, 600 F.3d 642 [hereinafter Comcast Order]. Rather than directly connecting a user’s computer directly to a shared
By order of the FCC, Comcast reported the following network management practices: Comcast’s subscribers had been grouped together and routed to hubs through a Cable Modem Termination System (CMTS). There had been approximately 3300 CMTS hubs functioning in the Comcast network, and they had served several million subscribers. Subscribers’ cable modems had shared upstream ports (content received from users’ cable modems) and downstream ports (content sent to users’ cable modems) on the hubs. In order to reduce congestion, Comcast installed hardware that analyzed the upstream traffic and managed information packets with characteristics that put undue strain on the network, in effect terminating the delivery of those packets. Comcast was careful to note that actual packet content was not inspected.

The FCC did not take sole issue with Comcast’s network management itself, however—what was more disconcerting was the fact that traffic-blocking targeted specific online conduct and “a customer ha[d] no way of knowing when Comcast . . . [had] terminate[d] a connection.”

After investigation, the FCC decided that Comcast was not engaging in “reasonable network practices,” and thus concluded that Comcast had violated the agency’s Broadband Policy Statement by “imped[ing] Internet users’ ability to use applications and access content of their choice.” The FCC ordered Comcast to file a disclosure statement with the Commission detailing its invasive network management practices and to suspend the unreasonable practices at issue. The FCC claimed jurisdiction to rule on Comcast’s conduct through multiple sections of the Communications Act by direct
and ancillary authority. Invoking Brand X, the Commission based its authority to regulate “facilities-based ISPs under its Title I ancillary jurisdiction.”

Comcast challenged the order in the D.C. Circuit, where the Court of Appeals held that the FCC failed to argue with specificity its statutory basis to regulate broadband data management practices. The FCC attempted to rely on its ancillary authority under Title I of the Act, as suggested regarding DSL service in Brand X, but the D.C. Circuit opined that the FCC would be stretching the Supreme Court’s precedent too far in arguing that this was a grant of “plenary authority over such providers....” The court also rejected any argument that the FCC could draw ancillary authority from policy statements such as Section 151 of the Communications Act, which states the purpose of the FCC: to regulate “interstate and foreign commerce in communication by wire so as to make available... a rapid, efficient, Nation-wide, and world-wide wire and radio communication service.” The court determined that legislative statements of policy—though conceivably declarations of the “legislative will”—“alone cannot provide the basis for the Commission’s exercise of ancillary authority”; the FCC needed a congressionally delegated power to which the administrative agency’s action could be “tethered.” The court went on methodically to decide that each section of the Act the FCC cited (including common

99 The Commission relied on section 230(b) of the Communications Act as well as six other sections of the Act to justify exercising jurisdiction. Id. at 13,036.
101 Comcast Order, supra note 87, at 13,035 (quoting Brand X, 545 U.S. at 996).
102 Comcast v. FCC, 600 F.3d 642, 661 (D.C. Cir. 2010). Despite the D.C. Circuit’s conclusion, this was not the first time the FCC had exercised authority in preventing a carrier from blocking Internet applications and content. In 2005, the FCC adopted a consent decree requiring a fine of Madison River Communications, which was blocking ports used for VoIP traffic. Madison River Commc’n, LLC & Affiliated Cos., 20 FCC Rcd. 4295, 4297 (Mar. 3, 2005) (consent decree).
103 47 U.S.C. § 154(i) (2006) (“The Commission may perform any and all acts, make such rules and regulations, and issue such orders... as may be necessary in execution of its functions.”).
104 The Supreme Court, in dicta, stated that the Commission may “reconsider[] its treatment of DSL service... when it decides whether, pursuant to its ancillary Title I jurisdiction, to require cable companies to allow independent ISPs access to their facilities.” Brand X, 545 U.S. at 1002.
105 Comcast, 600 F.3d at 650-51.
106 Id. at 651-52 (quoting 47 U.S.C. § 151).
107 Id. at 652.
108 Id. at 654.
109 Id.
carriage requirements\textsuperscript{110} and the mandate to implement broadband\textsuperscript{111} delegates no specific authority over the practice at issue.\textsuperscript{112} Thus, in one rap of the gavel, the D.C. Circuit created far-reaching consequences for the debate on net neutrality. More importantly, the decision has called into question the extent of a right to access Internet content—a guarantee that Internet users should expect to retain.

D. Post-Comcast Developments

To combat the trend of deregulation, the FCC has recently proposed and adopted rules for broadband management based on the Broadband Policy Statement,\textsuperscript{113} which explains the enforcement mechanisms that will attempt to aid in antidiscrimination measures.\textsuperscript{114} In the recent order, the FCC established three broad rules toward preserving an open Internet: transparency in broadband management, a prohibition against blocking “lawful content, applications, services, [and] non-harmful devices,” and a prohibition against “unreasonable discrimination.”\textsuperscript{115}

While an important step in FCC regulation of Internet practices, the rules stand on unstable legal foundation after Comcast. The FCC faced challenges immediately after it released its Net Neutrality Order. Verizon has brought an appeal to challenge the FCC’s authority to enforce the rules.\textsuperscript{116} The House of Representatives has also challenged the rules, moving to overturn the Net Neutrality Order through its powers under the Congressional Review Act.\textsuperscript{117} Contending that “the retail availability of Internet access service [has] never [been] regulated,”\textsuperscript{118} and noting the “sweeping” and “stifling” effect the rules would have,\textsuperscript{119} the House Committee on Energy

\begin{footnotesize}
\begin{enumerate}
\item See supra note 39 and accompanying text.
\item Comcast, 600 F.3d at 658-61 (discussing 47 U.S.C.A. §§ 201, 257, 301-99b, 543, 1302(a)).
\item See supra note 58 and accompanying text.
\item See generally Net Neutrality Order, supra note 17.
\item Net Neutrality Order, supra note 17, at 17,906.
\item Id.
\item Id. at 6.
\end{enumerate}
\end{footnotesize}
and Commerce voted to disapprove the Order.\textsuperscript{120} The House voted in favor of the committee’s resolution.\textsuperscript{121}

Textually speaking, Congress’s resolution appears to harp on a meritorious argument. Comcast’s holding is much broader than a statement that the FCC lacks authority because there is no specific rule on network management: the court’s language explicitly states that the FCC has not shown tethering for a statutory authority to regulate the activity of broadband network management.\textsuperscript{122} As noted by Commissioner McDowell, the lesson from Comcast was that Congress “has not established a new title of the Act to police Internet network management, not even implicitly.”\textsuperscript{123} Nonetheless, the FCC stated in its Net Neutrality Order that it has ancillary authority to pass the rules under several sections of the Act, including section 706.\textsuperscript{124} But the D.C. Circuit ruled unequivocally that section 706\textsuperscript{125} does “not delegate any regulatory authority” for broadband network management.\textsuperscript{126} Therefore, without congressional action, the FCC cannot properly impose regulatory obligations with respect to network management on ISPs. This shift in regulatory authority suggests that it is due time to impart greater weight on the discussion at hand.

* * *

Today, the Internet is non-neutral, privately regulated, and free from oversight protecting individual freedom. For now, Internet users are stuck in the bottleneck. Some suggest that the incentive to block or tier Internet access is not economically viable for the Internet gatekeeper.\textsuperscript{127} But if private ordering were a sustainable solution, the incidence of traffic-shaping,
blocking, and tiering should not have increased after Comcast. In fact, users continue to feel packet restrictions over home and wireless networks.\(^{128}\) In 2010, Level 3—the primary backbone provider for Netflix—engaged in heated negotiations with Comcast because Comcast began tolling Level 3’s traffic transmitted over the network.\(^{129}\) Even more recently, several ISPs have proposed and implemented tiered or capped access to the Internet in the mobile space.\(^{130}\) While there are economic arguments to support the beneficial aspects of private broadband, the public must begin to speculate as to its options should private intermediaries constrain the bottleneck to a point of no return.

II. THE PITFALLS OF OTHER MODES OF PROTECTION

Post-Comcast, regulators must look forward to fashion a regime that will better account for a right to access Internet content. Although the FCC may be disempowered to adjudicate the issues presented by non-neutral telecommunications under its current Title II authority, other federal powers with the ability to enforce individual liberties still exist. Considered a priori, a new approach to FCC oversight, antitrust litigation, or congressional legislation could each provide a meaningful method to regulate ISPs and account for end-users’ theoretical right to access. This section concludes, however, that these modes of protection face problems similar to, and even broader than, the now defunct FCC framework.

A. Administrative Law

This section proposes methods by which federal administrative agencies may step into the fray of Internet access regulation. First, this section will outline and criticize

\(^{128}\) See Net Neutrality Order, supra note 17, at 17,925-27.


\(^{130}\) AT&T has implemented a data-capping regime that charges a fee for every fifty GB of content over the 150 GB limit for mobile web users. Amy Lee, AT&T to Impose Broadband Data Cap, HUFFINGTON POST (Mar. 14, 2011, 11:21 AM), http://www.huffingtonpost.com/2011/03/14/att-data-cap_n_835318.html. Verizon also made the move to tiered data pricing in July 2011. Roger Cheng, Verizon's New Pricing Plan is a Godsend for Sprint, CNET (July 6, 2011, 10:26 AM), http://news.cnet.com/8301-1035_3-20077218-94/verizons-new-pricing-plan-is-a-godsend-for-sprint/?tag=rteol;pop. If the AT&T/T-Mobile merger takes effect, Sprint will be the only mobile wireless carrier with market power that does not restrict bandwidth usage. Id.
the FCC’s Third Way—a proposal designed to reconceptualize the categorical approach to telecommunications policy. Second, this section will present similar and broader problems presented by antitrust litigation as a mode of protection.

1. The FCC’s Third Way

Comcast and its background decisions have greatly weakened an enforceable regulatory scheme that protects Internet content access through the FCC’s administrative power. In a frenzy to fill the regulatory gap created, Chairman Genachowski has proposed building a legal foundation for the regulation of Internet access by bifurcating the classification between Internet access and Internet content itself, allowing the FCC to regulate access to the Internet as a “telecommunications service” under Title II common carriage requirements\(^{131}\) and to regulate the information layer under Title I.\(^{132}\) This strategy is aptly nicknamed the “Third Way”—a third method of regulation beyond staying the course or reclassifying broadband Internet to Title II regulation altogether.\(^ {133}\) The FCC bases its legal foundation for this reclassification on the dissent in Brand X, where Justice Scalia argued that transmission of broadband and computing were two separate “offerings” within the meaning of the Communications Act.\(^ {134}\)

But as long as rules are based solely on the power of the FCC to regulate toward an efficient communications network,\(^ {135}\) there is ample room for regulation to run astray from individual interests in accessing the Internet’s content. For example, the historical trend is that the FCC pushes the boundaries of regulating indecent speech over broadcast.\(^ {136}\) Without tying FCC action to principles of individual freedom, Chairman Genachowski’s proposal and the attempted passage of the Net Neutrality Order stop short of guaranteeing access to

---

\(^{131}\) See supra note 37 and accompanying text.


\(^{133}\) Id.

\(^{134}\) Id.; Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs., 545 U.S. 967, 1008 (2005) (Scalia, J., dissenting) (“T[he telecommunications component of cable-modem service retains such ample independent identity that it must be regarded as being on offer . . . .”).


That is, even if the FCC could regulate discriminatory practices limiting such access now, there is no guarantee that the FCC would do so in the future.

Despite the passage of the Net Neutrality Order, the FCC's regulatory capture and politicization present stark obstacles in appointing the FCC as the sole arbiter of end-users' individual rights going forward. In May 2011, after approving Comcast's massive acquisition of NBC Universal (NBCU), Commissioner Baker took a position with Comcast as senior vice president of government affairs. This development is problematic because the Comcast-NBCU merger goes against obvious policy considerations in promoting competition and common notions of First Amendment theory.

As noted by Judge Greene in precluding AT&T from entering the market of electronic publishing post-divestiture:

If, under these circumstances, AT&T were permitted to engage both in the transmission and the generation of information, there would be a substantial risk not only that it would stifle the efforts of other electronic publishers but that it would acquire a substantial monopoly over the generation of news in the more general sense. Such a development would strike at a principle which lies at the heart of the First Amendment: that the American people are entitled to a diversity of sources of information.

2. Antitrust Enforcement

The possibility of antitrust enforcement presents pitfalls equally deleterious to the end-user's right to access. In February 2011, the Federal Trade Commission (FTC) hired as

137 “There is a very strong presumption in most legal systems that other things being equal an interpretation which makes a law conform to a principle is to be preferred to one which does not.” Joseph Raz, Legal Principles and the Limits of Law, 81 YALE L.J. 823, 839 (1972).

138 Of interest, Rob Frieden explains that the FCC's attempts to gain greater flexibility in interpreting its statutory authority could be motivated by the following: “[T]he FCC engages in decision making with a predetermined outcome designed to accrue political dividends and support economic doctrine regardless of the facts and regardless of whether the decision unfairly and unlawfully tilts the competitive playing field in favor of one group of stakeholders over others.” Rob Frieden, Neither Fish Nor Fowl: New Strategies for Selective Regulation of Information Services, 6 J. TELECOMM. & HIGH TECH. L. 373, 415-16 (2008).


140 See infra Part III.B.1.b.

a senior advisor Tim Wu, a neutrality advocate and scholar in telecommunications policy. Beyond demonstrating the FTC’s intention to regulate the telecommunications industry, Wu’s appointment signals the FTC’s desire to gain consultation in regulating telecommunications economics. The FTC, along with the FCC and Department of Justice (DOJ), is vested with the power to require certain provisions in agreements between merging telecommunications companies. The FTC’s goal, of course, is to prevent unfair competition and deceptive acts in the marketplace.

The FTC’s role in regulating competition could potentially account for end-user interests through mandating competitive interconnection, and neutrality principles in the course of corporate mergers. But the degree to which interconnection requirements preclude other antitrust suits may be unclear after the Supreme Court’s decision in Verizon Communications v. Law Offices of Curtis V. Trinko. Furthermore, though the FTC may be qualified to monitor and enforce complex antitrust violations that may emerge in non-neutral conduct, the FTC—like the FCC—is no less susceptible to partisan and sometimes shortsighted goals. Just as the FCC has no obligation to preserve the free-expression interests of consumers, the FTC and DOJ may likewise maintain minimal oversight when the American economy benefits from a potential merger despite other harmful effects that merger may entail.

Furthermore, mandating neutrality provisions in merger agreements poses the possibility of piecemeal regulation. That is, under the FTC’s review, an ISP may voluntarily take on neutrality principles with respect to some

---

145 See Spulber & Yoo, supra note 38, at 1874.
147 Cf. id. at 415 (expressing disfavor in granting a general court’s authority to engage in regulatory practices more typical to an administrative agency).
149 Consider, for instance, the problems that have emerged from siloed treatment of information services as opposed to telecommunication services, where both in fact provide identical offerings. See supra notes 38-53 and accompanying text.
forms of transmission, but not others. In the Comcast-NBCU merger, for instance, Comcast agreed to abide by the FCC’s Net Neutrality Order, even if the order was overturned by a federal court.\textsuperscript{150} The agreement did not, however, prohibit Comcast from blocking Google TV, an emerging television service that provides video programming over Internet protocol, but on a digital television set.\textsuperscript{151} A lack of uniformity with respect to neutrality principles undermines Internet connectivity where the bottleneck problem is particularly constraining—regions of the country that have no choice in deciding which broadband provider to use.\textsuperscript{152} In those regions, the market cannot remedy an ISP’s lack of net neutrality through competing providers’ ability to offer greater packaged access.

B. Congressional Action

Congressional legislation could potentially fill the gap in broadband regulation to preserve user freedom, but Congress must be careful to avoid crafting the systemic problems that created the non-neutral network in the first place. Congress has attempted on several occasions to fashion bills that, in one way or another, proscribe acts of data discrimination and business models that discriminate against end-users. In 2007, the U.S. Senate introduced the Internet Freedom Preservation Act to amend the Communications Act.\textsuperscript{153} The bill proposed a new section, “Internet Neutrality,” to be appended to Title I of the Communications Act.\textsuperscript{154} The new section would address largely the same concerns presented in the FCC’s Broadband Policy Statement, but it would more explicitly proscribe the acts of blocking, discriminating against, or degrading broadband service for accessing lawful content.\textsuperscript{155} In addition, the bill requires ISPs to transmit content in a non-discriminatory manner that never “impose[s] a charge on the basis of the type of content.”\textsuperscript{156}

---


\textsuperscript{151} Id.

\textsuperscript{152} See Yemini, supra note 18, at 14 (“More than one quarter of consumers have only one choice between cable and DSL, and even in markets with both services available, customers usually face a duopoly . . . .”); see also supra note 64.

\textsuperscript{153} S. 215, 110th Cong. (2007).

\textsuperscript{154} Id. § 2.

\textsuperscript{155} Id.

\textsuperscript{156} Id.
The House of Representatives presented a similar bill in 2009 that proposed to amend the Communications Act to include a section called “Internet Freedom.”157 In addition to protections similar to those listed in the FCC’s Broadband Policy Statement, the bill proposed to prohibit both charging a fee to access lawful content and providing or selling devices that prioritize traffic for content or application providers; more broadly, the bill mandated “offer[ing] Internet access service to any person upon reasonable request therefor.”158 Significantly, the bill attempted to give the FCC power to make rules protecting against data discrimination and other anticompetitive practices.159 Legislatively empowering the FCC to make such rules would greatly help to fill the jurisdictional gap between the FCC’s Broadband Policy Statement and rules for network management.

But congressional legislation has two inherent limitations. First, with respect to rulemaking authority, mandating broad and unchecked regulatory power to the FCC has historically created the very threat that necessitates this writing.160 Second, legislation may codify overly specific legal regimes that cannot properly adapt to the dynamic technology that emerges in telecommunications. The Telecommunications Act, for instance, was created with the intention of opening the market for competition in telephony, but Congress could not adequately consider the emergence of Internet over broadband at the time of the Act’s passage. In turn, legislation left the medium untouched by the obsolete, siloed common-carriage requirements of federal law,161 and the populace continues to wait for a legislative solution.

***

Public sentiment and popular expectations do not control the federal regulatory powers-that-be; in fact, societal and governmental interests are sometimes in direct conflict. The FCC and its predecessor, the Federal Radio Commission (FRC), specifically, have been allowed leeway to infringe on normative expectations typically subject to constitutional protection.162 Therefore, if the only avenue to uphold citizens’

---

158 Id.
159 Id.
160 See supra Part II.A.1.
161 See supra notes 33-43 and accompanying text.
interests rests in courts of appeals' review of these infringing decisions, the Chevron standard categorically tips the scale towards affirming administrative regulatory decisions. Considering the rise of the non-neutral network, that framework is perilous to the establishment of a right to access Internet content. If, on the other hand, there is a competing forum for public outcry against regulatory action and communication-industry practices, ISPs and the government may be pressured to comport with individuals' interests. Supplementing the FCC's regulatory power with claims of individual right—in turn elevating the discourse on network management and pricing—thus avoids two pitfalls inherent in the other forms of regulation: first, users can be sure that they will have a claim available to them despite partisan effects on regulatory bodies; and, more importantly, the free flow of ideas can be secured in the Internet medium with a malleable standard despite the current lack of regulatory power under the Communications Act.

III. THE VENUE OF LAST RESORT: FEDERAL COURTS

In the absence of other oversight, federal courts can appropriately enforce a right to access Internet content; they have original jurisdiction over constitutional disputes and the ability to establish uniform rules. Though there are downsides to judicial oversight, the benefits are well-suited for the topic at hand. The underlying question, however, is what doctrinal "equipment" the courts can use to adjudicate these

---

Amendment standards applied to them.

See infra notes 175-76 and accompanying text.
disputes. This part will detail the constitutional doctrine applicable to an end-user’s challenge and address its current hurdles absent the existence of a private right of action—state action and standing. Incidentally, this part explores the contours of a theoretical right to access Internet content. To be clear, the goal of this section is not to argue that a non-neutral network violates the Constitution. Rather, in recognizing at the forefront that there is no express constitutional basis for litigation (though that point is arguable166), the goal here is to highlight the normative constitutional values that could inform a judicial avenue of redress to protect the end-user’s right to access Internet content. The considerations posed here will provide a basis for a legislatively created cause-of-action for end-users.

A. The Benefits and Drawbacks of Judicial Enforcement

Addressing Internet rights through the court system is appealing for two reasons. First, case-by-case adjudication will maintain the order of individual rights while considering the interests of ISPs. Contextual, fact-specific consideration will also account for new technologies that may modify individual and corporate interests without creating legacy limitations that entrench themselves in federal legislation or administrative rulemaking.167 Second, a broad power to adjudicate disputes between end-users and ISPs will avoid the constitutional problems that arise from overly specific congressional mandates168 and fill the regulatory gap left in administrative law.169 Furthermore, while net neutrality is an issue of technological infrastructure, the debate’s implications on personal liberty are so great that creating an additional, countermajoritarian remedial avenue based in constitutional doctrine may increase pressure on ISPs to properly consider individual freedom through their own self-regulation.170

As noted by the FCC, case-by-case adjudication is preferred when considering data management regulation and

---

166 See infra Part III.C.
167 See supra notes 41-50 and accompanying text.
168 See infra Part III.B.2 (discussing First and Fifth Amendment protections of ISPs).
169 See supra Part II.A.
170 It is worth noting that the FCC certainly has free expression in mind by imposing its net neutrality rules. See Net Neutrality Order, supra at 17, at 17,906. The concern posed here is whether that standard can be maintained by FCC oversight. The history of Internet regulation suggests that it cannot, and thus, regulators are left to consider whether self-regulation may provide the proper oversight. But see supra Part I.D.
common carrier–type problems.171 Indeed, the complexity of the medium begs that regulators weigh interests in every dispute arising over its access.172 The language of the Communications Act also suggests a fact-based inquiry when assessing the regulation of the telecommunications industry.173 Some may argue that the FCC’s Title I case-by-case adjudicative authority is better suited for the specialized knowledge of FCC commissioners. But in applying conservative and static constitutional jurisprudence,174 courts are equipped to ascertain whether data management practices are overly broad or burdensome on an individual’s right to access information through fact-based inquiries.175 Further, establishing a separate avenue of adjudicative remedy in the courts will put action directly into the hands of citizens.

Federal adjudication through constitutional discourse provides two further benefits. First, judicial interpretation of the Constitution can maintain uniform, binding precedent through the Supremacy Clause.176 Uniformity would therefore extend and preserve end-user rights to their maximum potential. Second, grounding ISP practices in limits delineated by the Constitution creates precedent that supersedes the actions of administrative agencies, and, further, signals congressional action.177 Decisions

---

171 Comcast Order, supra note 87, at 13,046.
172 See id.
173 See Communications Act, 47 U.S.C. § 201(a)-(b) (2006) (allowing for common carrier requirements when such action is “desirable in the public interest,” and as long as they are “just and reasonable”); see also Hush-A-Phone Corp. v. United States, 238 F.2d 266, 269 (D.C. Cir. 1956) (weighing the public and private detriment caused by federal regulation of a telephone invention).
174 For instance, the First Amendment precedent stating that strict scrutiny is triggered when the government restricts expression “because of its message, its ideas, its subject matter, or its content” traces its origins to the 1960s. Police Dep’t of City of Chi. v. Mosley, 408 U.S. 92, 96 (1972) (citing N.Y. Times v. Sullivan, 376 U.S. 254, 269, 270 (1964)). While the contours of this doctrine have changed over time, the origins remain in effect. See Brown v. Entm’t Merchs. Ass’n, 131 S. Ct. 2729, 2733 (2011).
175 Stated simply, a court need not know much about engineering or network design to determine whether there are other means to achieve an ISP’s “compelling purpose,” or whether there is any reasonable purpose at all. See infra Part III.B.1.a. Those means can be proposed by the litigants themselves.
176 U.S. Const. art. VI, cl. 2 (“This Constitution, and the Laws of the United States which shall be made in Pursuance thereof; and all Treaties made, or which shall be made, under the Authority of the United States, shall be the supreme Law of the Land; and the Judges in every State shall be bound thereby, any thing in the Constitution or Laws of any State to the Contrary notwithstanding.”).
177 For instance, in 1995, the U.S. Supreme Court ruled, in United States v. Lopez, that Congress exceeded its Commerce Clause power in legislating a federal offense for “knowingly . . . possess[ing] a firearm at a place that the individual knows, or has reasonable cause to believe, is a school zone.” 514 U.S. 549, 551, 561-63 (1995). The federal statute was later amended to include the “jurisdictional element” necessary
would therefore serve as a second line of defense should deregulation befall administrative action in the future.

Despite these systemic benefits, Lawrence Lessig argues that “[U.S. citizens] don’t want courts choosing among contested matters of values,” values that are clearly implicated in the Internet-access issues presented here. Courts are ill-fitted to determine these values, argues Lessig, because translating Internet issues into matters of constitutional law will inevitably result in “political” decision making that “makes,” rather than “finds,” cyberspace’s expressive characteristics. Put more concretely, when factual inquiries are left to judicial discretion, the possibility of directing cyberspace’s architectural realities increases. These decisions are perilous to Internet architecture, concludes Lessig, because they will dictate what cyberspace will become, perhaps in a manner contrary to end-user desires. For reasons discussed briefly below, Congress can circumscribe the courts’ value-oriented judgments through legislative specificity.

B. The Rights at Stake: The Contours of Constitutional Litigation

With the benefits of judicial oversight in mind, this section proposes some of the normative legal values that are relevant to end-user litigation. As discussed below, there is an initial hurdle to constitutional discourse because ISPs do not (at least ostensibly) seem to fit within the traditional state action doctrine. This section therefore, in part, applies constitutional precedent to ISPs by analogy, as if they were state actors. The end goal here is to provide and assess the relevant legal theory that may underpin a court’s decision in a hypothetical end-user challenge.

to “ensure... that the firearm possession in question affects interstate commerce.” United States v. Dorsey, 418 F.3d 1038, 1045-46 (9th Cir. 2005).


179 Id. at 316-17. “We have tools from real space that will help resolve the interpretive questions by pointing us in one direction or another, at least some of the time. But in the end the tools will guide us even less than they do in real space and time.” Id. at 25.

180 Id. at 317.

181 See infra Part III.C.1.
1. Edge Users' Rights

a. Traditional First Amendment Protection

There is no question that the Internet is a form of communication.\textsuperscript{182} The Supreme Court has confirmed that Internet content, if protected, can receive unqualified First Amendment scrutiny.\textsuperscript{183} Furthermore, Congress has found, as one of the bases of proposed legislation, that free speech is protected by “preserving the open nature of Internet communications.”\textsuperscript{184}

Impeding content due to agreement or disagreement with its message is a viewpoint-based regulation deserving heightened scrutiny under traditional First Amendment jurisprudence.\textsuperscript{185} Though the Internet has unique characteristics from other media, it does not have characteristics that set it so apart from the realm of speech that content- or viewpoint-based regulation would yield lesser scrutiny under First Amendment analysis.\textsuperscript{186} This is the case despite the fact that Internet “speech” can be omnidirectional—that is, without a specified geographical or personal recipient—and does not seem to fit neatly into the traditional, bimodal framework of speaker and government conceived by the Constitution.\textsuperscript{187} Therefore, as under traditional First Amendment protection, if restraining or degrading edge-user access to the core infrastructure were a form of content-specific regulation,\textsuperscript{188} the First Amendment would provide protection under


\textsuperscript{183} See Reno, 521 U.S. at 870.


\textsuperscript{186} Reno, 521 U.S. at 868-70 (noting that the Internet has not historically been regulated by the government and is not as invasive as radio or television, and still triggers unqualified First Amendment scrutiny).

\textsuperscript{187} See Ashcroft v. ACLU, 535 U.S. 564, 576-77 (2002) (ruling that the question of “contemporary community standards” in the discussion of obscenity can still be applied in the context of Internet speech, even though there is no targeted geographic community in posting on a website). That is not to say, however, that the Internet is a clean fit in the First Amendment, two-speaker framework. See Yemini, supra note 18, at 41-49.

\textsuperscript{188} Turner, 512 U.S. at 642.
heightened scrutiny.\textsuperscript{189} The Supreme Court has also taken the position, however, that when speech and conduct are “joined in a single course of action,” there must be a balancing between First Amendment protections and broader societal interests.\textsuperscript{190}

On its face, issues of accessibility fit more neatly into the classical sense of content-neutral regulation. But in fact, accessibility can create both content-specific and content-neutral restrictions. In Comcast, the network management technology at issue did not affect categories of content, but rather modes of transfer.\textsuperscript{191} In fact, Comcast is very clear that the content of information it transmits is not inspected.\textsuperscript{192} Comcast’s definition of content, however, may be too narrow in this discussion. Content’s definition is rapidly changing in the world of electronic files and digital conveyance of information. The authorship of an electronic document is said by many to yield a privilege under the doctrine of attorney work product.\textsuperscript{193} This would indicate that law regards protocol tags and metadata as document content, treating such tags the same way as a law firm’s internal memoranda.\textsuperscript{194} Comcast’s definition also fails to recognize that impeding one type of file transfer could disproportionately affect a category of content associated with that transfer. For example, ISPs could easily identify—and in turn limit or de-prioritize—online gaming tags, which are distinguishable among others.\textsuperscript{195} It would be difficult to argue that such a practice is “content-neutral”; the practice appears to directly target a category of speech.\textsuperscript{196} Furthermore,

\begin{flushright}
\textsuperscript{189} Id.
\textsuperscript{191} See supra note 93 and accompanying text.
\textsuperscript{192} Comcast Description,
\textsuperscript{193} supra note 88, at 7.
\textsuperscript{194} The American Bar Association has ethical rules regarding the inadvertent disclosure of metadata, or the “data about data” (author, date of authorship, etc.) in an electronic document. See generally Joshua J. Poje, Metadata Ethics Opinions Around the U.S., ABA, http://www.abanet.org/tech/lt/'ét/docs/metadatachart.html (last updated July 20, 2011).
\textsuperscript{195} Similar to metadata, the Second Circuit concluded in Universal City Studios, Inc. v. Corley, 273 F.3d 429 (2d Cir. 2001), that computer code is in fact “speech” under the First Amendment. Id. at 445. In determining whether restriction of code was content-neutral, the court also noted whether the “regulated activity is sufficiently imbued with elements of communication to fall within the scope of the First . . . Amendment.” Id. at 450 (internal quotation marks omitted). Information packets clearly fall within these parameters.
\textsuperscript{196} See Wu, supra note 62, at 168 (noting that ISPs could block online gaming through application information).
\textsuperscript{196} See Turner Broad. Sys. v. FCC, 512 U.S. 622, 642 (1994) (“Our precedents thus apply the most exacting scrutiny to regulations that suppress, disadvantage, or impose differential burdens upon speech because of its content.”). In fact, ISPs
lazy packet inspection to reduce congestion may sweep up unintended content. Courts should analyze any of these situations under a traditional, heightened scrutiny if the end-user does not explicitly agree to the specific practice.

b. Public Forums, the Right to Access Information, and the Free Flow of Ideas

In the modern age, where corporate media operate as gatekeepers to several of the most accessible means of relaying information, some have suggested a contextual approach to First Amendment protection, one—in contrast to the traditional view that the First Amendment is a restriction on what actors cannot proscribe—which asks informational gatekeepers to create opportunities for expression to be heard. Indeed, “[a] realistic view of the first amendment requires recognition that a right of expression is somewhat thin if it can be exercised only at the sufferance of the managers of mass communications.” This approach applies more neatly than the traditional approach discussed above.

The contextual, or affirmative, construction of the First Amendment was most famously declared by Justice Black:

[The First] Amendment rests on the assumption that the widest possible dissemination of information from diverse and antagonistic sources is essential to the welfare of the public, that a free press is a condition of a free society. Surely a command that the government itself shall not impede the free flow of ideas does not afford non-

including Comcast have not maintained complete integrity in shaping traffic without any consideration of content. In a clear example of content-based discrimination, Comcast was found censoring e-mails sent from antiwar groups on two separate occasions. NUNZIATO, supra note 27, at 5-7. In another instance, Comcast blocked access to Gmail and Google for their Boston subscribers, suggesting that subscribers switch over to Comcast e-mail. Id. at 11.

Comcast’s data regulation technology, for example, creates session thresholds for P2P applications specifically, without considering whether an overall traffic threshold is met. Comcast Description, supra note 88, at 8-9. In effect, P2P protocols are identified and limited without consideration of the overall strain on the network, simply because these protocols are known to cause congestion.

See Post, supra note 19, at 183 (“Even speech that seems on its surface irrelevant to politics... serves to focus and clarify public values and commitments. That is why constitutional protection should be extended to media of communication...”); Barron, supra note 23, at 1655-56 (“Today ideas reach the millions largely to the extent they are permitted entry into the great metropolitan dailies, news magazines, and broadcasting networks... As a constitutional theory of the communication of ideas, laissez faire is manifestly [sic] irrelevant.”).

Barron, supra note 23, at 1648 (1967). Barron’s work, though it precedes the issue addressed in this note, is eerily relevant.
governmental combinations a refuge if they impose restraints upon that constitutionally guaranteed freedom.\textsuperscript{200}

The role of government as a speech enhancer, contemplated by Justice Black, shifts the focus of the First Amendment from protecting the speaker to protecting the listener.\textsuperscript{201} It is supported in part by the doctrines of public forum, common carriage, and fairness.\textsuperscript{202} Henry Perritt has found additional support for a right to access cyberspace through common carrier requirements, the antitrust essential facilities doctrine, and contract law.\textsuperscript{203} Under this affirmative theory, preserving the free flow of ideas will differ when discussing radio as opposed to television, or newspaper as opposed to the Internet: each medium has a distinct abundance of resources and only a certain number of adequate alternative forms of expression that competently yield the same communicative effect.\textsuperscript{204} Therefore, the Internet's unique characteristics must be considered when determining the mode and extent of the government's intervention.

Although an intermediary's right to broadcast and editorialize the information it chooses conflicts with the right to access lawful content, such a right has been recognized and furthered by the FCC and the Supreme Court alike.\textsuperscript{205} In Red

\textsuperscript{200} Associated Press v. United States, 326 U.S. 1, 20 (1945) (emphasis added); see also Stromberg v. California, 283 U.S. 359, 369 (1931) ("The maintenance of the opportunity for free political discussion to the end that government may be responsive to the will of the people and that charges may be obtained by lawful means...is a fundamental principle of our constitutional system.").

\textsuperscript{201} "What is essential is not that everyone shall speak, but that everything worth saying shall be said." Post, supra note 19, at 181 (quoting ALEXANDER MEIKLEJOHN, POLITICAL FREEDOM: THE CONSTITUTIONAL POWERS OF THE PEOPLE 26 (1960)); see also Richmond Newspapers v. Virginia, 448 U.S. 555, 576 (1980) ("Free speech carries with it some freedom to listen.").

\textsuperscript{202} Nunziato, supra note 27, at 41. The constitutionality of the fairness doctrine is addressed at length in Red Lion v. FCC, 395 U.S. 367, 386-401 (1969). Though common carriage is not explicitly mentioned in the Telecommunications Act common carrier designation, it has been argued that Internet services meet the legal standard for a common carriage industry. See generally James B. Speta, A Common Carrier Approach to Internet Connection, 54 Fed. Comm. L.J. 225 (2002).

\textsuperscript{203} Perritt, supra note 162, at 61-62. For a brief discussion of the feasibility of federal protection under antitrust and common carriage doctrines, see supra Part II.A. The essential facilities doctrine, which explicitly emerged in the 1970s, allows courts to issue injunctive relief requiring monopolists to open "irreproducible bottleneck resources" to their rivals. Spulber & Yoo, supra note 38, at 1828-29.

\textsuperscript{204} Barron, supra note 23, at 1650-53; see also Kovacs v. Cooper, 336 U.S. 77, 97 (1949) (Jackson, J., concurring) ("The moving picture screen, the radio, the newspaper, the handbill, the sound truck and the street corner orator have differing natures, values, abuses and dangers."). Supreme Court jurisprudence supports this conception of the First Amendment. See Reno v. ACLU, 521 U.S. 844, 868 (1997) (discussing the roles history, scarcity, and invasiveness play in informing First Amendment protection).

\textsuperscript{205} See infra Part III.B.2.a.
Lion Broadcasting v. FCC, for example, a radio broadcaster challenged the constitutionality of the FCC's “fairness doctrine” and related regulations, which required “reply time” for those who were personally attacked over the airwaves. The doctrine was designed to further two main duties held by broadcast licensees: to “give adequate coverage to public issues” and to create fair coverage that “accurately reflects the opposing views.” Faced with a First Amendment challenge, the Supreme Court reasoned that Congress, and in turn the FCC, has the authority to regulate broadcast licensees’ conveyance of information due to the scarcity of the medium and “the legitimate claims of those unable without governmental assistance to gain access to [radio] frequencies for expression of their views.”

Though the fairness doctrine is no longer in effect, Red Lion implicitly upholds a normative right held by the public to access all viewpoints through commonly used media and government's authority to enact, in its power, what is necessary to effect that access given the nature of the medium. Boiled down, the fairness doctrine's practical ramifications is an extension of affirmative First Amendment obligations on private broadcasters. The Supreme Court would later find that the Internet lacks the scarcity concerns posed by the radio spectrum, but the effect on the right to access is nearly identical when Internet users are blocked from accessing online information. That is, limitation on the accessibility of an online discussion board or blog would be substantially similar to the limitations imposed by broadcasters that impede access to the free flow of ideas.

Legally speaking, however, the standard of review for an impediment to access is not entirely clear. And moreover, the Supreme Court has shown signs of scaling back its right-to-access interpretation of the First Amendment. In American Library Ass'n v. United States, the Supreme Court ruled that library-provided Internet was not a traditional public forum.

207 Id. at 373-75 (quoting 47 C.F.R. §§ 73.123, 73.300, 73.598, 73.679 (repealed 1987)).
208 Id. at 377.
209 Id. at 389, 400-01.
210 See id. at 390 (“It is the purpose of the First Amendment to preserve an uninhabited marketplace of ideas in which truth will ultimately prevail, rather than to countenance monopolization of that market.” (citing Associated Press v. United States, 326 U.S. 1, 20 (1945))).
and therefore held that congressional legislation imposing limitations on library Internet access was constitutional under reduced scrutiny.\textsuperscript{213} As the Court stated, a public library “provides Internet access, not to encourage a diversity of views from private speakers, . . . but for the same reasons it offers other library resources: to facilitate research, learning, and recreational pursuits.”\textsuperscript{214} The Court’s interpretation of public forums appears, however, not to be a wholesale exclusion of the Internet, but rather a narrow interpretation of library-offered Internet.\textsuperscript{215}

Indeed, when it comes to accessing primary sources online, outside of a library, recent events illustrate how the Internet may provide exclusive means to access some forms of information, thereby deserving a greater degree of protection.\textsuperscript{216} After the Iranian election of 2009 was met with popular unrest, Twitter, YouTube, and Facebook exploded with content exposing the violent turmoil, which was otherwise limited from exportation.\textsuperscript{217} In fact, some observers reported that news stations needed the Twitter and YouTube content to cover Iran’s protests.\textsuperscript{218} The Iran protests demonstrate that the Internet provides the potential for a direct conduit between public events and society—an important, unintermediated, and informative experience worthy of protection.\textsuperscript{219}

When adjudicating issues of free expression, courts sometimes weigh the information sought as well.\textsuperscript{220} It would flout the central policy of constitutional protection to contend

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{213} NUNZIATO, supra note 27, at 81-82 (citing Am. Library Ass’n, 539 U.S. 194).
\item \textsuperscript{214} Am. Library Ass’n, 539 U.S. at 206.
\item \textsuperscript{215} See id. at 205 (“Internet access in public libraries is neither a ‘traditional’ nor a ‘designated’ public form.” (emphasis added)).
\item \textsuperscript{216} Cf. Saxbe v. Wash. Post, 417 U.S. 843, 847-48 (1974) (reasoning that prison policy banning face-to-face meetings with inmates by unaffiliated individuals was not violation of First Amendment, in part, because journalists can attain information from prisons in other ways besides face-to-face interviews).
\item \textsuperscript{218} Twitter 1, CNN 0, ECONOMIST, June 18, 2009, available at http://www.economist.com/node/13856224?story_id=13856224.
\item \textsuperscript{219} In Richmond Newspapers v. Virginia, the Supreme Court held that the press’s access to a criminal trial is protected under the First Amendment. 448 U.S. 555, 580 (1980). Important to that holding was the Court’s reasoning regarding the role of the press, which “contribute[s] to public understanding of the rule of law” by “supplying the representations or reality of the real life drama once available only in the courtroom.” Id. at 572-73.
\item \textsuperscript{220} Compare Saxbe, 417 U.S. at 862 (Powell, J., dissenting) (noting the important press function in accessing information toward “preserving free public discussion of governmental affairs”), with Richmond Newspapers, 448 U.S. at 569 (concluding that the function of a public trial is indispensible in American law).
\end{itemize}
\end{footnotesize}
that Internet access should be guaranteed to access unlawful information.\footnote{221} Furthermore, unfettered access to unlawful or harmful content would create congestion over ISP networks that could substantially decrease accessibility to other, lawful content.\footnote{222} Even in the context of free expression, the First Amendment does not offer protection to obscenity, defamation, and incitement.\footnote{223} In recent discussion, there has been some attack on the Internet site WikiLeaks\footnote{224} for how it potentially compromises national security.\footnote{225} But the general principle should hold for WikiLeaks and the like: the government has a “heavy burden” to justify prior restraint on the spread of information;\footnote{226} dissemination of such information should only be restrained if it would “gravely prejudice the defense interests of the United States or result in irreparable injury to the United States.”\footnote{227}

That said, the notion that courts ought to make value judgments with respect to speech is inherently suspect. Indeed, recent First Amendment jurisprudence seems to reject this approach outright.\footnote{228} But the separation of powers easily provides a counterbalance. In making a cause of action for end-users, Congress can use constitutional jurisprudence to circumscribe the “lawful content”\footnote{229} to which a litigant can seek access.\footnote{230} In turn, value judgments would not be left to judicial discretion, but rather to carefully detailed, bright-line rules.

\footnote{221} The First Amendment historically does not protect obscenity, Miller v. California, 413 U.S. 15 (1973), or “fighting words,” Chaplinsky v. New Hampshire, 315 U.S. 568, 572 (1942), for example.
\footnote{222} See infra notes 257-58 and accompanying text.
\footnote{223} United States v. Stevens, 130 S. Ct. 1577, 1584 (2010).
\footnote{224} WIKILEAKS, \url{http://www.wikileaks.org} (last visited Sept. 11, 2011).
\footnote{228} Stevens, 130 S. Ct. at 1585 (“The First Amendment’s guarantee of free speech does not extend only to categories of speech that survive an ad hoc balancing of relative social costs and benefits.”).
\footnote{229} This language is pulled from the Net Neutrality Order as a primer for the base-level protections the cause-of-action could seek to protect. See Net Neutrality Order, supra note 17, at 17,906.
\footnote{230} See infra notes 316-18 and accompanying text.
2. Rights of ISPs

a. First Amendment Protection for Conduit Speech

Just as the First Amendment affords protection to individual speakers, it also affords protection to commercial enterprises, or conduits, that carry individuals' speech. In cable television, for instance, the Supreme Court has held that editorial discretion in selecting content under its services is within the scope of First Amendment protection. Similarly, a public library's decision to exclude materials does not trigger heightened scrutiny. In Turner Broadcasting System v. FCC, a cable television case, the Supreme Court addressed the issue of whether a congressional act requiring cable operators to relay local broadcasts was an infringement on the freedom of speech or of the press. The Court applied intermediate scrutiny to the Act, stating first that laws singling out a medium are subject "to at least some degree of heightened First Amendment scrutiny," then concluding that the Act was content neutral.

The FCC and others have argued that intermediate scrutiny should likewise apply to a federal neutrality mandate. But cases dealing with media suggest that the First Amendment interests of media consumers outweigh the providers' interests at some point, proportional in part to how much of a speech "conduit" the provider is. Important to recall

---

231 See Turner Broad. Sys. v. FCC, 512 U.S. 622, 636 (1994) ("Cable programmers and cable operators engage in and transmit speech, and they are entitled to the protection of the speech and press provisions of the First Amendment.").

232 City of Los Angeles v. Preferred Commc'ns, Inc., 476 U.S. 488, 494 (1986) (concluding that expression by a cable operator includes "exercising editorial discretion over which stations or programs to include in its repertoire").


234 Id. at 622.

235 Id. at 626.

236 Id. at 661-62.

237 Id. at 640-41.

238 Id. at 622, 647. Some have argued that the standard of review was "intermediate plus," where it "decidedly privileges speech rights over values." Yemini, supra note 18, at 25 (quoting Ellen P. Goodman, Media Policy and Free Speech: The First Amendment at War With Itself, 35 HOFSTRA L. REV. 1211, 1219 (2007)).

239 Net Neutrality Order, supra note 17, at 17,983; Yemini, supra note 18, at 20-22. Yemini argues that, similar to a cable operator, an ISP would just as easily engage in protected expression under Turner by blocking a website, for example. Id. at 18-20. In fact, Yemini suggests, to argue that ISPs do not engage in protected editorial discretion would be a contradictory position for net neutrality activists, where such discretion is essentially the conduct in question when engaging in data discrimination. Id. at 18-19.

is that the First Amendment’s protections vary with the carrying medium. Unlike cable television, Internet content is primarily created by end-users. Even the largest corporate content providers are search engines and social networking sites, which derive most of their Internet traffic from end-user contributions. Furthermore, the type of “editorial discretion” argued by ISPs “bears little resemblance to an editor’s choosing which programs [like in cable television] . . . to carry.”

The FCC went further in its Net Neutrality Order to say that no court has ever “suggested that regulation of common carriage requirements triggers First Amendment scrutiny.”

Nevertheless, Wu and Lessig appear to be correct in that the governmental interests furthered by promoting neutrality are “important or substantial” enough to withstand intermediate scrutiny by the courts. As stated by the Turner Court, “assuring that the public has access to a multiplicity of information sources is a governmental purpose of the highest order, for it promotes values central to the First Amendment.” Indeed, when weighing conduits’ First Amendment interests against society’s in the free flow of ideas, the freedom of expression is at least as heavy as the freedom of editorial discretion under an affirmative conception of the First Amendment. But the cases also suggest that editorial discretion qualifies for greater protection as a service provider

[D]ifferences in the characteristics of new media justify differences in the First Amendment standards applied to them.” Red Lion Broad. Co., 395 U.S. at 386. Greater latitude is allowed in infringing on First Amendment rights in broadcast-radio regulation. Id. Very little latitude is granted in regulating print. Tornillo, 418 U.S. at 258.

Net Neutrality Order, supra note 17, at 17,983; see also Comcast Order, supra note 87, at 13,033 (“Unlike newspapers or radio or broadcast television . . . the Internet gives Americans a great degree of control over the information that they receive.”).

Net Neutrality Order, supra note 17, at 17,983. But see id. at 18,073-74 n.114 (dissenting statement of Comm’r Robert M. McDowell).


Wu & Lessig, supra note 66, at 10. The interests put forward include “promoting the widespread dissemination of information” and “promoting fair competition in the market.” Id. The FCC agrees with this reasoning in its Net Neutrality Order, but it provides somewhat different interests. Net Neutrality Order, supra note 17, at 17,984 (relating the important government interests that would pass intermediate scrutiny, such as “consumer choice, end-user control, free expression, and the freedom to innovate without permission”).

Turner, 512 U.S. at 663.

See Barron, supra note 23, at 1654-55 (quoting Associated Press v. United States, 320 U.S. 1, 20 (1945)). But see Yoo, supra note 18, at 702 (“[I]n terms of deciding how that balance [between edge users and providers] should be struck, the cases indicate that free speech considerations favor preserving intermediaries’ editorial discretion unless the relevant technologies fall within a narrow range of exceptions, all of which the Court has found to be inapplicable to the Internet.”).
becomes less interested in the content of the expression being conveyed. From the vantage point of an Internet subscriber, Internet carriers should be disinterested in lawful content transmitted over their networks.

b. Fifth Amendment Takings

Another argument presented against the neutrality mandate is that governmental action would be a confiscatory taking under Fifth Amendment doctrine. The argument suggests that a limitation on network management practices would in turn limit the profitability of the broadband industry, or the choice to engage in new business models. When assessing whether a taking occurs, courts look to interference with “investment-backed expectations,” the “economic impact of the regulation,” and “the character of the government action.” The FCC found in its Net Neutrality Order, however, that the Fifth Amendment challenges are without merit. The FCC stated that “takings law makes clear that property owners cannot, as a general matter, expect that existing legal requirements regarding their property will remain entirely unchanged.”

Case law seems to support the FCC’s finding as it relates to exercise of First Amendment rights. In assessing whether allowing demonstrators to handbill in a shopping center is the proper exercise of the property owner’s right under the First Amendment, the Supreme Court stated:

248 Henry Perritt observes:

The First Amendment permits forcing some information conduits to accept content generated by others, but only when such forcing is necessary to permit the content to find its audience. When the entity burdened by the duty has relatively little interest in expression, for example if it is simply a router, the First Amendment allows a broader range of legislative and regulatory discretion to impose a duty because the harm to First Amendment interests is minimal.

Perritt, supra note 162, at 94; see also Red Lion Broad. Co. v. FCC, 395 U.S. 367, 390 (1969) (“It is the right of the viewers and listeners, not the right of the broadcasters, which is paramount.”).

249 “[N]or shall private property be taken for public use, without just compensation.” U.S. Const. amend. V.


251 See, e.g., Lyons, supra note 62, at 95. Lyons also argues that the Net Neutrality Rules could be a per se Fifth Amendment taking under the permanent physical occupation doctrine of Loretto. Id. at 92-94.


253 Id.
center was a “taking,” the Supreme Court stated in Pruneyard Shopping Center v. Robins that the true test to determine whether the public has—through state-granted free speech protections—confiscated property rights of an owner is whether the restriction “forc[es] some people alone to bear public burdens which, in all fairness and justice, should be borne by the public as a whole.” Applying that test, the shopping center’s Fifth Amendment claim failed because the owner could not show unreasonable impairment on value or use of his property, especially where the demonstrators were orderly and remained in common areas.

In the case of Internet accessibility, the benefit lost from ISPs’ data management practices would burden the entire industry. The government’s intervention would also not substantially limit the capability of the providers to compete. Moreover, the obligation to carry content without discrimination does not infringe on any physical space, so the challenge is inherently limited.

That is not to say that a Fifth Amendment challenge is completely without merit. It is at least foreseeable that a court could find reasonable a business model where ISPs compete over providing greater access to Internet content. In fact, modern-day consumers analogously subscribe to cable packages that function similarly—providers compete to provide bundled packages of content that suit the needs of the consumer. Furthermore, an overbearing and blanket prohibition of data discrimination could actually serve the opposite effect it intends if ISPs do not deploy new Internet infrastructure: a pure nondiscrimination mandate could sacrifice the ability to access content effectively for outright connectivity.

If ISPs cannot meet their burden of providing the requisite QoS to their subscribers, then sites that demand high bandwidth will

---

255 Id. at 83-84.
256 Perritt, supra note 162, at 93-94 (citing Bell Atlantic Tel. Cos. v. FCC, 24 F.3d 1441 (D.C. Cir. 1994) (invalidating order by FCC to allot office space of local telephone carriers to competing carriers)). But see Lyons, supra note 62, at 93 (arguing that mandated neutrality would allow a continuous right by content providers to “physically invade broadband networks with their electronic signals and permanently occupy portions of network capacity”).
257 See Becker et al., supra note 31, at 502 (discussing the disincentive of network providers to engage in data discrimination).
258 In other words, at odds with one another are the rights to connectivity and a “public interest in quality infrastructure.” Perritt, supra note 162, at 58.
not be available to users on that network, implicating the very problem that net neutrality seeks to remedy and decreasing service efficiency for ISPs.\textsuperscript{258}

C. Hurdles to Litigation

The previous section shows that there is some doctrinal basis under which a court could adjudicate the facts of an end-user litigation. Though informative to norms that should be considered in constructing a cause of action, the jurisprudence on its own is not self-executing in this context. Two doctrines are particularly noteworthy hurdles to litigation: state action and standing. Their respective merits and problems will be assessed here.

1. Arguing State Action

As Christopher Yoo has argued, “[I]nvoking [the] First Amendment as requiring governmental intervention to redress private power would stand the First Amendment on its head.”\textsuperscript{260} In other words, the First Amendment is a protection against intrusion by the government, not by private actors.\textsuperscript{261} That is not to say, though, that constitutional restrictions cannot be imposed on private actors.\textsuperscript{262} Under Fourteenth Amendment jurisprudence, for example, the Court has held that private actors may be bound by constitutional obligations when the state judiciary enforces racially discriminatory restrictive

\textsuperscript{258} Put differently, the issue of Internet access as a right is enforced by blanket prohibitions at the risk of creating a critical mass: complete deregulation gives ISPs the power to unduly discriminate against content, while over-regulation may make the content inaccessible to begin with. This effect would undermine the service the ISPs intend to provide. “We have a public network that is indeed a great creative commons for data applications, but it is less so for any application that requires a minimum quality of service. True application neutrality may, in fact, sometimes require a close vertical relationship between a broadband operator and Internet service provider.” Wu, supra note 62, at 148.

\textsuperscript{260} Yoo, supra note 18, at 700.

\textsuperscript{261} Id.

\textsuperscript{262} In Burton v. Wilmington Parking Authority, the Supreme Court held that a restaurant that leased space from a Wilmington, Delaware, parking facility was a state actor because in leasing space from the city authority, and being maintained by public funds, the restaurant was “an integral part of a public building devoted to a public parking service.” 365 U.S. 715, 724 (1961); see also Jones v. Alfred H. Mayer Co., 392 U.S. 409 (1968) (applying the Thirteenth Amendment to private actors); Heart of Atlanta Motel, Inc. v. United States, 379 U.S. 241 (1964) (same). Warranting further discussion, courts have in the past considered First Amendment rights of defendants sued under right of publicity claims. See, e.g., Cardtoons, L.C. v. Major League Baseball Players Ass'n, 95 F.3d 959, 968 (10th Cir. 1996).
In the context of Internet litigation, however, there is some indication that judges are reluctant to extend the state action classification to ISPs outright. Nonetheless, there is an argument that ISPs may fall directly within the parameters of the state action doctrine, in turn allowing for direct constitutional enforcement. The Third and Fourth Circuits have stated that there are three distinct tests utilized by the Supreme Court to assess whether a private actor has crossed the line into state action. First, the court may consider whether the entity has “exercised powers that are traditionally the exclusive prerogative of the state” (or, in short, the “public function” test). Second, the court may ask whether “the private party has acted with the help of or in concert with state officials.” And in the final test, the court may determine whether “[t]he State has so far insinuated itself into a position of interdependence with . . . [the acting party] that it must be recognized as a joint participant in the challenged activity.”

Public “insinuation” may be present in the Internet’s origins, where the initial connection of networks creating the Internet was instituted in large part by the federal government. In Lebron v. National Railroad Passenger Corp., the Supreme Court ruled, for First Amendment purposes, that Amtrak was a government actor subject to the limitations dictated by the Constitution. In reaching its decision, the Court compared Amtrak—a statutorily created rail service held through private stock—with other government corporations like the Communications Satellite Corporation (Comsat). Similar to Amtrak, Comsat was created by the federal government, yet it is “capitalized entirely with private funds.” The Internet was developed under circumstances similar to those of Amtrak and Comsat. Through the 1960s, one of the initial networks giving rise to the Internet, the Advanced

---

263 See Shelley v. Kramer, 334 U.S. 1, 18-21 (1948) (holding that judicial enforcement of racially discriminatory restrictions is state action that warrants Fourteenth Amendment challenge). Here, portions of the Communications Act, 47 U.S.C. §§ 151, 157, 201, 230(b), 256, 601, as interpreted by the D.C. Circuit in Comcast, 600 F.3d 642, imposes similar restraint on the freedom of expression for Internet users.
267 Id. at 400.
268 Id. at 385.
269 Id. at 390-91.
270 Id. at 390.
Research Projects Agency Network (ARPANET), was funded and developed by the Department of Defense for research purposes.\footnote{A Brief History of NSF and the Internet, NAT’L SCI. FOUND., http://www.nsf.gov/news/news_summ.jsp?cntn_id=103050 (last visited July 31, 2011).} The National Science Foundation (NSF), a federal agency created by Congress, extended the network connection to U.S. universities and beyond the realm of defense research.\footnote{Id.} One of the first national backbone infrastructures was, in turn, created by the NSF in 1992.\footnote{Id.}

The government is also largely responsible for implementing the uniform system of packet-traffic management currently controlled by the Internet Corporation for Assigned Names and Numbers (ICANN).\footnote{Id.} The federal government established ICANN in 1998, privatizing the root server system that informs subordinate servers of IP addresses.\footnote{See STUART M. BENJAMIN ET AL., TELECOMMUNICATIONS LAW AND POLICY 914 (2d ed. 2006).} Without that arrangement, the Internet would have potentially grown into a set of redundant and conflicting internetworks.\footnote{Id.} This fact is particularly relevant when considering the constitutional treatment of packet and network traffic management; though Internet services are provided by private actors, this does not detract from the clearly public origins that suggest a “close nexus”\footnote{See Improvement of Technical Management of Names and Addresses, 63 Fed. Reg. 8826, 8827 (proposed Feb. 20, 1998) (proposing the coordination of the root server network to ensure the system “work[s] smoothly” and “preserve[s] the stability and interconnectivity of the Internet”).} between Internet services and federal action.

The public function test, considered in conjunction with the joint participation of government action in private enterprises, also provides meaningful guidance in determining the public nature of Internet control. Marsh v. Alabama\footnote{See Jackson v. Metro. Edison Co., 419 U.S. 345, 351 (1974) (“[T]he inquiry must be whether there is 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.”).} expresses the principle that the more a private party opens his property to the function of the public, “the more do his rights become circumscribed by the statutory and constitutional rights who use it.”\footnote{326 U.S. 501 (1946).} In simple terms, the public function test states that, at some point, private owners provide access to a property so fundamentally public in nature that the owners...
begin to owe an obligation to protect "identical interests" to those held by citizens of a state or municipality.\textsuperscript{280}

In Altmann v. Television Signal Corp.,\textsuperscript{281} the United States District Court for the Northern District of California ruled that the defendant cable company was a "state actor" for the purposes of a constitutional challenge. The court used language to suggest that it assessed state action based on the government's participation in empowering the company to censor indecent public-access programming.\textsuperscript{282} Interestingly, the Altmann Court also noted the important function of public-access channels "to serve as public forums, accessible to all interests, including those that may otherwise lack the resources to communicate through electronic media."\textsuperscript{283} The court, in turn, concluded that the effect of allowing cable operators to block indecent material on these stations would likely fail strict scrutiny.\textsuperscript{284} This decision seems to apply here on two different levels, suggesting, first, that the government has a duty to preserve public forums in broadcast media, or, alternatively, that the government has imposed enough obligation on the cable operator to subsume its private interests.\textsuperscript{285}

Marsh and Altmann considered in tandem present the possibility that courts may be willing to extend the state action doctrine in issues of First Amendment freedom where, as in Altmann, largely public functions are subjected to the requirements of facilitating public forums and where, as in Marsh, the private intermediaries that own the medium of expression have opened the medium to public discourse. To be sure, the Court has never ruled that the Internet is a traditional public forum,\textsuperscript{286} but, in any event, the unintermediated debate the Internet holds strongly indicates that it ought to be. Indeed, except to the extent limited by non-neutrality, the Internet is a wide-open resource of access to information—suggesting its similarity to a sidewalk or square

\textsuperscript{280} Id. at 507. But see Hudgens v. NLRB, 424 U.S. 507, 519 (1976) (narrowly construing Marsh).
\textsuperscript{281} 849 F. Supp. 1335 (N.D. Cal. 1994).
\textsuperscript{282} Id. at 1342-43.
\textsuperscript{283} Id. at 1340.
\textsuperscript{284} Id. at 1343.
\textsuperscript{285} See id. at 1342 ("Congress stripped cable operators of any editorial control over constitutionally protected speech.").
\textsuperscript{286} The Court touched on the issue in United States v. American Library Ass'n, but the holding appears limited to the circumstances of library-provided Internet. See supra notes 214-15 and accompanying text.
deserving of public-function protection, even though controlled by private actors.\textsuperscript{287}

That being said, and in all fairness, the appropriate application of state action to ISPs is one that may be novel to normal constitutional discourse. Lawrence Lessig writes:

Architectures constitute cyberspace; these architectures are varied; they variously embed political values; some of these values have constitutional import. Yet for the most part—and fortunately—these architectures are private. They are constructed by universities or corporations and implemented on wires no longer funded by the Defense Department. They are private and therefore traditionally outside the scope of constitutional review. The constitutional values of privacy, access, rights of anonymity, and equality need not trouble this new world, since this world is "private" and the Constitution is concerned only with "state action."

Why this should be is not clear to me. If code functions as law, then we are creating the most significant new jurisdiction since the Louisiana Purchase. Yet we are building it just outside the Constitution's review. Indeed, we are building it just so that the Constitution will not govern—as if we want to be free of the constraints of value embedded by that tradition.\textsuperscript{288}

But if courts were to extend constitutional review to the issue at hand without legislation, Lessig is right to note that constitutional theory may not clearly provide the proper context.\textsuperscript{289} Other modes of speech are easily divided into two distinct and opposing forces: the government and the speaker.\textsuperscript{290} The Internet, by contrast, is inherently a multispeaker,\textsuperscript{291} multi-interest, or multilateral environment, mediated by

\textsuperscript{287} The logical conclusion from this fact is that the Internet should be treated as a common carrier, which seems "presumptively" appropriate. See Speta, supra note 202, at 269.

\textsuperscript{288} Lessig, supra note 178, at 317-18.

\textsuperscript{289} Lessig points out that the architecture contemplated by the framers of the Constitution, based in natural law and the laws of economics, is directly opposed to the man-made architecture of sovereignty in cyberspace based in code. Id. at 318.

\textsuperscript{290} Amit M. Schejter & Moran Yemini, "Justice and Only Justice, You Shall Pursue": Network Neutrality, the First Amendment and John Rawls's Theory of Justice, 14 MICH. TELECOMM. TECH. L. REV. 137, 162 (2007), available at http://www.mttlr.org/volfourteen/schejter&yemini.pdf; see also U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech, or of the press . . . ." (emphasis added)).

nongovernmental actors. In the multispeaker environment of the Internet, the free flow of ideas cannot be protected if liberty interests are evaluated in a two-speaker forum, because there are tiers of speakers and expressive interests present: edge users submit and access content; corporate content and application providers engineer methods to access their content and that of edge users; and ISPs deliver the resources to make all the above interactions possible. The government, after the Comcast decision, is only a limited overseer in this situation.

Though perhaps an activist-oriented position, this multitiered system suggests that rather than using precedent based in a bilateral system, we should be evaluating competing communicative interests irrespective of the state-like characteristics of these intermediary entities. Addressing the state action doctrine under this method promotes beneficial cultural values, granting to legal discourse symbolic statements upon which to base decisions of right and wrong, and freeing the categorical approach of decisions from what, in turn, becomes an arbitrary public/private distinction.

2. Standing to Sue and Problems of Pleading

An Internet subscriber may also face problems bringing their claims in federal court due to issues of standing and pleading. Unless the recipient of information knows by some other means that he or she is expecting data, there is insufficient information to proceed with an action on the basis that an ISP is limiting or degrading service. Without factual support, a litigant would have difficulty meeting the heightened pleading standards of the federal court system.
which require a statement of the claim that is “plausible on its face.”298 Even if the case can survive pleading, a litigant must still show a “concrete and particularized injury” that is “not conjectural or hypothetical” in order to have standing to sue in federal court.299 This is a difficult burden in the context of Internet accessibility; any number of errors, circumstances, or other factors could contribute to the degradation or blockage of service. Then again, injury may be more easily shown in challenging an ISP’s tiering access, which today is the more widely utilized form of limitation.300

An end-user cannot, therefore, sue with an adequate basis if transparency is not preserved. Transparency, as noted by the FCC, “increases the likelihood . . . that the Internet community will identify problematic conduct and suggest fixes.”301 As suggested below, the FCC must maintain its role in issuing orders of transparency if litigants will have an opportunity to properly plead their cases and discover the harm giving rise to them.

IV. PROPOSED ACTION

If courts were to adjudicate a right to access Internet content through litigation, issues of state action and standing would likely stop the actions in their tracks. Congress must legislate outright that unreasonably impeding access to lawful Internet content gives rise to a private right of action in federal courts.302 Congress should be specific as to what constitutes a per se violation, using First Amendment jurisprudence and the FCC’s recent Net Neutrality Order as a guide in determining which practices should be banned outright.303 But the legislation should also give sufficient leeway to courts to adjudicate circumstances specific to the technology, conduct, and information at issue. Congress must also make explicit in this legislation that the courts have jurisdiction to issue a wide array

298 Ashcroft v. Iqbal, 129 S. Ct. 1937, 1949 (2009) (quoting Bell Atlantic v. Twombly, 550 U.S. 544, 570 (2007)). “A claim has facial plausibility when the plaintiff pleads factual content that allows the court to draw the reasonable inference that the defendant is liable for the misconduct alleged.” Id.
300 See supra note 130 and accompanying text.
301 Net Neutrality Order, supra note 17, at 17,936-37.
303 See id. at 17,941-51 (prohibiting blocking and unreasonable discrimination).
of equitable relief—an end-user litigant is unlikely to incur actual money damages. Equitable relief will serve well to take the place of the FCC’s power to issue orders. Beyond the federal court system, the FCC must maintain its role as a diligent overseer of transparency. Transparency will enable end-users to make informed, specific, and plausible claims against ISPs.

A. The Specifics of Legislation

Congress can certainly outlaw data management practices and pricing schemes that infringe on the interests of end-users through its power to regulate interstate commerce.\(^{304}\) Legislating on issues of constitutional right, even without an explicit provision in the Constitution allowing for this, is not a shaky proposition either.\(^{305}\) The legislation must guide a court hearing a network discrimination challenge, instructing the judge to weigh the individual’s right to access the information sought with the reasonableness of the data management practice or the tiering scheme.

In its Comcast Order the FCC sought to impose a new standard for adjudicating network management challenges: “[An ISP’s] practice should further a critically important interest and be narrowly or carefully tailored to serve that interest.”\(^{306}\) The FCC’s proposed standard, however, is too stringent to accommodate ISPs’ interest; any standard utilized should not sanction ISPs for failing to exercise the least restrictive means possible in every instance of managing their networks. In fact, the FCC would later scale back its “narrowly tailored” standard because it “overly constrain[ed] network

\(^{304}\) See U.S. CONST. art. I, § 8, cl. 3 (“The Congress shall have Power . . . To regulate Commerce with foreign Nations, and among the several States, and with the Indian Tribes”). “Congress has authority to regulate and protect the instrumentalities of interstate commerce, and persons or things in interstate commerce.” Gonzales v. Raich, 545 U.S. 1, 16-17 (2005). Packets cross state lines instantaneously and continuously through the packet-switching system. See supra notes 27-28 and accompanying text.

\(^{305}\) In the past, Congress has used its powers to codify Fourth Amendment rights that courts declined to incorporate into common law. See Daniel J. Solove, Fourth Amendment Codification and Professor Kerr’s Misguided Call for Judicial Deference, 74 FORDHAM L. REV. 747, 754-60 (2005) (discussing passage of federal statutes concerning electronic surveillance, government access to records, and searches involving communicative material).

\(^{306}\) Comcast Order, supra note 87, at 13,055-56.
engineering decisions” and failed to consider that “reasonable network management practices may differ across platforms."

To preserve the property and speech interests of ISPs and yet still keep users from facing undue discrimination, legislation should adopt an intermediate standard of review. That is, network providers must show an important interest in utilizing challenged data management practices, and the practice should be substantially related to that interest. Applying intermediate scrutiny comports with the conduit speech assessed in Turner and is consistent with the conclusions of the FCC in its recent Net Neutrality Order. What “reasonable” network management is—and conversely, what unreasonable discrimination is not—poses a question that cannot be addressed at length here. But even where a system of interest-balancing is appropriate, bright-line categories of conduct are inherently unlawful. The FCC, in its rules to preserve the open Internet, has laid out some methods that weigh towards unreasonable discrimination on the one hand and reasonable network management on the other. Congress should look to these definitions and set out conduct that is unlawful per se. Consistent with First Amendment principles, Congress should at the very least contemplate prohibiting ISP conduct that impedes access to lawful and

---

307 Net Neutrality Order, supra note 17, at 17,953.
308 Id. Wu agrees that the appropriate goal is to “strike a balance” between the opposing individual and ISP interests so as to preserve efficiency. Wu, supra note 62, at 165.
309 Moran Yemini agrees that a normative (non-evidentiary) weighing of interests is the appropriate model of preserving individuals’ Internet free speech interests. Yemini, supra note 18, at 36-37. However, Yemini rejects the traditional approach advocated here—for governmental “enhancement” of free speech—where such a model unduly focuses on potentially irrelevant governmental interests, and because governmental enhancement of free speech interests presumes “a right superior to a governmental interest.” Id. at 37. What Yemini fails to consider are the Fifth Amendment ramifications of mandated neutrality, which put a conduit and an individual speaker on equal footing in the Internet forum, thereby warranting a presumption of right superior to governmental interest for edge users and ISPs.
311 In determining the standard of review, the Turner Court concluded Congress’s “must-carry” provisions for cable providers fell within the intermediate scrutiny of the Court, where the regulation is a “content-neutral restriction[] that impose[s] an incidental burden on speech.” Id. at 662.
312 See supra notes 307-08 and accompanying text.
313 Net Neutrality Order, supra note 17, at 17,946-47 (discussing discrimination and its unreasonableness when it is anti-competitive, harmful to end-users, or impairs free expression).
314 Id. at 17,954-56 (discussing reasonable network management practices like congestion-management and Internet security).
315 See supra Part III.B.1.a.
nonharmful content and degrades provision of services that compete with services provided by the ISP.

Congress must also guide courts in considering the weight accorded to the information sought by plaintiffs. Needless to say, information that would not normally enjoy First Amendment protection should not fall within the statute. But the Internet is home to many frivolous—and perhaps harmful—enterprises that may be protected speech. While Internet resources are not as scarce as broadcast radio frequencies, for example, they are still not unlimited. Unimpeded access to pornographic content (though protected under normal First Amendment jurisprudence, if not obscene) may therefore be granted at the expense of access to other information that has a direct benefit to society. The end-user is left back at square one if courts consistently uphold these access challenges—a degradation of service to access the Internet content of one’s choice. Congress should therefore provide that impeding access to certain categories of content—carefully circumscribed in the legislation for being indecent, threatening, or harmful—is actionable only if the defendant ISP’s access restrictions also sweep up other lawful content that Congress does not set out.

No congressional bill to date has accounted for end-user and ISP interests to the extent advocated in this note. Nor has a bill constructed a private right of action that stands in federal courts. Congress should create that right of action and mandate that the following examples of ISP conduct are significant or important network management and pricing interests:

- Preserving the privacy and security of end-user information and hardware;
- Providing unimpeded access to the content of an end-user’s choice, including but not limited to network management to maintain standard QoS for high-bandwidth web applications, so long as providing such access does not throttle or prioritize one form of lawful content over another;

---

317 Sable Commc’ns of Cal., Inc. v. FCC, 492 U.S. 115, 126 (1989).
318 “With respect to the Internet, intermediaries help protect end-users from exposure to spam, pornography, and viruses . . . while helping them sift through the ever-growing avalanche of desired content that appears on the Internet every day.” Yoo, supra note 18, at 701.
319 This list is, in part, derived from the 2007 Senate bill. See Internet Freedom Preservation Act, S. 215, 110th Cong. § 12(b) (2007).
• Allowing end-users to attach devices and use software that limit their own Internet accessibility to the extent end-users desire;
• Complying with other applicable federal or state law.

This list is not exhaustive. To preserve this bill’s application to circumstances currently unforeseeable, the judiciary must be entitled to consider other important interests identified by ISP defendants. But under no circumstances should an ISP’s interest in procuring a profit be sufficient, unless the network management or pricing scheme at issue only incidentally gains a profit in furthering one of the above-stated objectives. Finally, Congress should, at the very least, indicate that impeding access to any content that is not protected under the First Amendment does not give rise to an end-user challenge.

B. The FCC’s Continued Role in Transparency

Congress should also grant the FCC the authority to serve as an arbiter of transparency in ISP data management. In the Comcast proceedings, the FCC issued orders asking for more detailed illustrations of network management practices.\[320\]

The FCC should continue to use this power to play network management inspector. The FCC has the technical knowledge and resources to monitor the transparency of ISPs’ network management practices, and they are therefore better equipped to determine whether ISPs are being fully forthcoming in informing the public. Continued transparency in the market will keep information regarding ISPs in the open so litigants can file in court with a basis upon which to make allegations.

* * *

In conclusion, the reader should note that the model proposed here is not one that would replace the regulatory power of administrative law. It is rather designed as a countermajoritarian check on legislative and administrative solutions that have provided insufficient consideration of individual interests in Internet-content access. This new cause of action can serve to supplement any regulation with individual action—granting an avenue of remedy that is based in the conservative and (more) stable doctrines upheld by the Constitution. This additional mode

\[320\] See Comcast Order, supra note 87, at 13,060.
of remedy will not only put greater pressure on ISPs to preserve individual rights, but it can also serve as a barometer of consumer welfare in the Internet market.

CONCLUSION

The trend of recognizing access to Internet content as a right is important if United States citizens want to hold the power to access information and use the Internet as a speaking platform. The fact that other nations have adopted such a right is informative of international sentiment, but does not necessarily reflect the reality of American law. Traditional constitutional protections of the First Amendment (though not readily applicable to the right-to-access issue) evoke a possibility that litigation based in constitutional norms may serve as a mode of individual remedy, and further, as a form of pressure to exert upon Internet service providers. There are indeed major hurdles to this form of regulation: the outdated and draconian state action doctrine; an individual litigant’s difficulty in alleging or proving a plausible scheme to reduce accessibility; and the First and Fifth Amendment rights of ISPs in protecting their interests as conduits and enterprises of speech. The model proposed here allows for litigation on this topic and preserves transitory regulatory power held by federal courts until a new regime is fashioned. But that power rests on legislative action. In the meantime, U.S. citizens are left with little recourse or remedy to protect any theoretical right to access Internet content. Therefore, the potential for disempowerment of the populace remains.

Philip F. Weiss†

† J.D. Candidate, Brooklyn Law School, 2012; B.A., Boston University, 2007. I dedicate this to Ryan Weiss. May his memory be a blessing. I must thank my classmates and especially my colleagues on the Law Review for their remarkable effort, dedication, and patience. I would also like to thank Derek Bambauer and Jonathan Askin for their invaluable guidance. Finally, I could never have reached this point without the love and unwavering support of my family—my motivation, my hope, and my inspiration.