TABLE OF CONTENTS

ARTICLES

Democracy and Criminal Discovery
Reform After Connick and Garcetti Janet Moore 1329

The Class Differential in Privacy Law Michele Estrin Gilman 1389

Democratic Inclusion, Cognitive Development, and the Age of Electoral Majority Vivian E. Hamilton 1447

NOTES

For Sale: The Threat of State Public Accommodations Laws to the First Amendment Rights of Artistic Businesses Susan Nabet 1515

Protecting the Border, One Passenger Interrogation at a Time Dina Kleyman 1557

Gods Behind Bars: Prison Gangs, Due Process, and the First Amendment Justin L. Sowa 1593

Homebuyer Beware: MERS and the Law of Subsequent Purchasers Joshua J. Card 1633

“The Power to Tax Involves the Power to Destroy”: How Avant-Garde Art Outstrips the Imagination of Regulators, and Why a Judicial Rubric can Save It Rachel J. Tischler 1665
INTRODUCTION

The nondisclosure of information beneficial to criminal defendants causes wrongful convictions, wasteful litigation, and uncertainty in criminal adjudications.1 Prosecutors are required to disclose this information under *Brady v. Maryland*2 and related cases,3 criminal discovery rules,4 and codes of

---


professional ethics. But two recent U.S. Supreme Court decisions, *Connick v. Thompson* and *Garcetti v. Ceballos*, underscore the weak enforceability of *Brady*-line authorities as mechanisms for criminal discovery reform. The cases also point out the contrasting fairness and efficiency of full open file discovery as an alternative reform model.

*Connick* arose after John Thompson spent eighteen years behind bars. For fourteen of those years, Thompson was on twenty-three-hour-a-day solitary confinement in a six-by-nine foot windowless death row cell at Angola Prison. A few weeks before his scheduled execution, a last-ditch defense investigation revealed what the Louisiana Court of Appeal described as the prosecutors’ “intentional hiding of exculpatory evidence.” This new information led to Thompson’s release. He then filed a federal civil rights action. The federal jury awarded him $14 million in compensation for his wrongful imprisonment.

---


8 N.C. GEN. STAT. §§ 15A-903 to -910 (2012); id. § 15A-1415(f); see Mosteller, *supra* note 5, at 307-16 (discussing fairness and efficiency enhancements under full open file discovery statutes).

9 *Connick*, 131 S. Ct. at 1355; Thompson v. Connick, 553 F.3d 836, 842-43, 865 (5th Cir. 2009), aff’d by divided en banc opinion, 578 F.3d 293 (5th Cir. 2009), rev’d, 131 S. Ct. 1350 (2011).


11 *Connick*, 131 S. Ct. at 1355.

12 Id.

13 Id. at 1355-56.
jury found that Orleans Parish District Attorney Harry Connick had been deliberately indifferent toward the need to train line prosecutors on their *Brady* discovery duties.14

By a five-to-four vote, the Supreme Court ordered Thompson’s award vacated and sharply restricted 42 U.S.C. § 1983 as an avenue for enforcing prosecutors’ constitutional disclosure duties.15 The Court had previously held individual prosecutors immune from personal liability for failing to conduct *Brady* training.16 *Connick* effectively immunizes municipalities for those failures, eliminating the taxpaying and voting public as a meaningful resource to compensate and deter *Brady* violations. Writing for the dissent, Justice Ginsburg cautioned that “prosecutorial concealment . . . is bound to be repeated unless municipal agencies bear responsibility—made tangible by § 1983 liability.”17

*Connick* illustrates the subordination of *Brady* enforcement to other interests—here, a concern to limit municipal liability, even when government employees confess to intentional misconduct. *Connick* also highlights a similar subordination of *Brady* enforcement by the same five-member Supreme Court majority in *Garcetti v. Ceballos*18. Richard Ceballos’s civil rights case was in some respects the mirror image of John Thompson’s. Ceballos was not a criminal defendant. He was a Los Angeles prosecutor. He alleged that his supervisors unconstitutionally retaliated against him not for withholding beneficial information from the accused but for bringing such evidence to light.19 In *Garcetti*, the majority held that the First Amendment provides no protection against such retaliation.20

*Connick* was promptly condemned for restricting 42 U.S.C § 1983 as an avenue for enforcing *Brady*.21 But scholars

14 Id. at 1356.
15 Id. at 1356, 1366.
17 Connick, 131 S. Ct. at 1370 (Ginsburg, J., dissenting).
19 Id. at 416, 421-22.
20 Id. at 426.
have not unpacked Garcetti’s full significance on that point. In part, this neglect is understandable. With few exceptions, Garcetti has been analyzed on its terms, as clarifying First Amendment doctrine regarding government employee speech.

Closer analysis reveals a previously unremarked due process shield that should have protected Richard Ceballos from retaliation for Brady compliance. At a deeper level, the silence on Garcetti’s implications for constitutional criminal discovery is emblematic of the short shrift often accorded to the enforcement of prosecutors’ constitutional discovery obligations.

Taken together, the two cases neatly illustrate Brady’s weak enforceability. By effectively eliminating municipal liability even when prosecutors deliberately suppress evidence, Connick gives a wink-and-nod to nondisclosure. By limiting section 1983 protection against retaliation for good faith compliance with discovery duties, Garcetti sends a chilling message that prosecutors may be damned if they do disclose beneficial evidence to the defense. By landing this one-two punch against enforceability of Brady-line duties, Connick and Garcetti invite a contrast with full open file discovery statutes as the optimal strategy for increasing the fairness, finality, and efficiency of criminal adjudications.

Among discovery reform statutes, North Carolina’s are the most expansive in the nation. They mandate the prosecution’s disclosure to the defense of all information obtained in a criminal investigation. They require recordation of oral statements and impose criminal penalties for willful

---


23 See id. at 2176 & n.5 (citing scholarly commentary).

24 Connick’s wink-and-nod was not undone by the terse majority opinion granting Brady relief in the Orleans Parish case of Smith v. Cain, 132 S. Ct. 627, 630-31 (2012), nor by the remarkable suggestion during oral argument that the prosecutor consider forfeiting the case. See Lyle Denniston, Disaster at the Lectern, SCOTUSBLOG (Nov. 8, 2011, 4:51 PM), http://www.scotusblog.com/?p=131456. The governing doctrine invited diverse views on the materiality of the undisclosed evidence, as was demonstrated by Justice Thomas’s close analysis of the evidence, Smith, 132 S. Ct. at 631-41 (Thomas, J., dissenting), by the state trial judge who rejected Smith’s Brady claim on the merits after hearing witness testimony over the course of four days, Respondent’s Brief at 20, Smith, 132 S. Ct. 627 (No. 10-8145), and by the numerous Louisiana appellate judges who voted (apparently unanimously) to deny Smith’s petitions for discretionary review, see id. and State v. Smith, 45 So. 3d 1065 (La. 2010). For discussion of the practical and doctrinal problems that lead to such disparate assessments of Brady-line duties, see infra Part II.


26 Id. § 15A-903(a)(1)(c).
violations.\textsuperscript{27} Available empirical evidence shows significant success in the statutes’ implementation and expansion, yet full open file discovery remains a rarity in the United States.\textsuperscript{28} And like the precise due process disclosure duty at issue in \textit{Garcetti}, this cutting-edge development in criminal procedure has received scant scholarly attention.

The silence may result in part from a trend, expressed in Marc Miller and Ronald Wright’s \textit{The Black Box},\textsuperscript{29} which privileges internal bureaucratic improvement over litigation and legislation as the only effective avenue toward criminal justice reform—at least with respect to prosecutors’ discretionary decision making. Full open file discovery reform bolsters skepticism toward that trend. Dogged case investigation and litigation raised the profile of criminal discovery issues for the public, the media, and key legislators. Increased public scrutiny led to hard-fought political compromises in the enactment and amendment of the reform statutes. Full open file reform vindicates law and politics as effective strategies—complementary to internal agency reform—for increasing transparency and accountability in prosecutorial decision making. Broader attention and closer study, ideally through the full open file statutes’ evaluation as a model for uniform legislation, should raise discovery reform to parity with other criminal procedure reforms. Comparable initiatives include improvements in eyewitness identification protocols, in the testing and retention of forensic evidence, and in interrogation methods.\textsuperscript{30} This article fills an important analytical gap by focusing closely on the litigation and legislation that drove full open file discovery reform, and by proposing that the \textit{Connick-Garcetti} one-two punch against enforcing discovery duties can and should energize nationwide efforts to obtain full open file reform.

\textsuperscript{27} Id. § 15A-903(d).
\textsuperscript{28} For examples of other relatively broad criminal discovery provisions, see MMN. R. CRIM. P. 9.01 (2010); N.J. COURT R. 3:13-3(a)-(c) (2011).
\textsuperscript{29} Marc L. Miller & Ronald F. Wright, \textit{The Black Box}, 94 IOWA L. REV. 125, 128-30 (2008).
Parts I and II set the stage for in-depth examinations of Connick and Garcetti as examples of Brady’s weak enforceability and for discussion of the litigation and legislation that generated the full open file reform alternative. Part I excavates the roots of due process discovery doctrine to identify core meanings and principles. Part II summarizes intractable doctrinal and practical problems that weaken enforceability of constitutional criminal discovery rights and duties. Part III examines Connick and Garcetti as recent exemplars of those problems. Part IV contrasts Connick and Garcetti as examples of the Brady regime’s complexity and costs with the simplicity and efficiency of full open file discovery. This part surveys reported case law, available legislative history, and observations from some frontline participants as the most readily accessible evidence of full open file discovery’s implementation and expansion.

Forthcoming research identifies conditions that enable such reform in some jurisdictions and impede it in others. But the story of full open file discovery reform recasts the core lesson of Connick and Garcetti. The cases need not reinforce despair of litigation and legislation as strategies for sustainable system improvement. They can introduce a new chapter in a broader reform story. Their holdings underscore Brady’s weak enforceability and its intolerable results in wasted lives, trampled liberty, and squandered criminal justice resources. Connick and Garcetti should motivate broad adoption of full open file discovery statutes as a prerequisite—a necessary, although not sufficient condition—for improving efficiency, fairness, and finality in the resolution of criminal cases.

I. DUE PROCESS DISCLOSURE DUTIES

Brady doctrine requires prosecutors to disclose certain beneficial information to the defense.31 It is helpful to excavate the historical roots of these duties, if only to counteract shorthand citations that elide or misstate core principles and holdings of Brady-line cases.32 Due process discovery duties are
grounded in the prosecutorial mandate to speak truth and seek justice. The duties encompass two types of evidence. The first tends to reduce the defendant’s culpability with respect to guilt or sentencing. Examples include witness statements that corroborate an alibi defense. The second category comprises impeachment evidence. A deal between a witness and a prosecutor to exchange testimony for leniency on a pending charge, for example, may support an inference that the witness is less credible due to pro-prosecution bias.

The history of due process discovery doctrine also reveals a handful of principles through which the core meanings of truth-speaking and justice-seeking are to be implemented. For example, there is no mens rea element in a Brady claim. Prosecutors must disclose favorable evidence whether or not defense counsel requests it. Defendants need not prove prosecutors’ acts or omissions were undertaken intentionally or in bad faith. Another core principle requires cumulative prejudice review. Courts must assess harm from nondisclosure by reviewing the strength and weakness of all evidence presented by both parties at trial and in postconviction proceedings.

These core principles of Brady doctrine began to emerge in 1935, when Thomas Mooney obtained Supreme Court review of his capital murder case. Mooney was a workers’ rights activist. He was sentenced to death in 1917 for a San Francisco bombing that killed ten people. Mooney fought his conviction for eighteen years before the Supreme Court accepted his petition for federal habeas review. In Mooney v. Holohan, the Court held—for the first time—that the Fourteenth Amendment’s Due Process Clause forbids prosecutors from

implementing training programs for defense counsel, I have often found comparable misunderstanding or misstatement of criminal discovery rights and duties.

36 Id.
37 For thoughtful analysis of the duty to disclose impeachment evidence, see R. Michael Cassidy, Plea Bargaining, Discovery, and the Intractable Problem of Impeachment Disclosures, 64 Vand. L. Rev. 1429, 1431 (2011).
obtaining convictions through the knowing use of perjured testimony.\textsuperscript{43} The constitutional values at issue were fairness and reliability in contests between concentrated government power and the individual.\textsuperscript{44}

A few months later, the Court decided \textit{Berger v. United States}.\textsuperscript{45} This mine-run counterfeiting case contains oft-cited descriptions of the prosecution’s unique purpose and power and the corresponding primacy of prosecutors’ duties to speak truth and seek justice. Under \textit{Berger}, the prosecutor is a minister of government and “servant of the law.”\textsuperscript{46} The prosecution’s interest “is not that it shall win a case, but that justice shall be done.”\textsuperscript{47} The prosecution must ensure that “guilt shall not escape nor innocence suffer.”\textsuperscript{48} It is as much the prosecutor’s duty “to refrain from improper methods calculated to produce a wrongful conviction as it is to use every legitimate means to bring about a just one.”\textsuperscript{49} This is so, the Court reasoned, because the fact finder in a criminal case reasonably expects the prosecutor to abide by the foregoing high principles. Therefore, evidence and arguments bearing a prosecutor’s imprimatur weigh strongly against the defendant.\textsuperscript{50}

In a key doctrinal development, the \textit{Berger} Court assessed prejudice cumulatively. Berger alleged that the prosecutor in his case had violated his due process rights in several ways. Instead of addressing each allegation piecemeal, the Court held that the prosecutor’s misstatements of fact, insinuations of facts not in evidence, and other “pronounced and persistent” misconduct, taken together, rendered the proceedings unfair and required a new trial.\textsuperscript{51} \textit{Berger’s}

\textsuperscript{43} Mooney, 294 U.S. at 112-13. The Court sent the case back to state court for exhaustion of the new due process claim. \textit{Id.} at 115. Despite pleas on Mooney’s behalf from Felix Frankfurter via President Wilson, the California governor refused to order a new trial. Mooney spent four more years in prison. He was released the same week that Frankfurter was appointed to the Supreme Court. Frankfurter’s intervention on Mooney’s behalf was cited against him during his confirmation hearings through, inter alia, a letter from Theodore Roosevelt deriding Frankfurter’s “besmirching the reputation of God-fearing, patriotic Americans . . . destroying respect for law and order, and coddling anarchist[s], bomb throwers, and cowards.” Ringhand, \textit{supra} note 42, at 805 & n.56 (citing \textsc{Michael E. Parrish, Felix Frankfurter and His Times: The Reform Years 99 (1982)}).

\textsuperscript{44} Mooney, 294 U.S. at 112-15.
\textsuperscript{45} 295 U.S. 78, 78 (1935).
\textsuperscript{46} \textit{Id.} at 88.
\textsuperscript{47} \textit{Id.}
\textsuperscript{48} \textit{Id.}
\textsuperscript{49} \textit{Id.}
\textsuperscript{50} \textit{Id.}
\textsuperscript{51} \textit{Id.} at 89; see \textit{Dye v. Hofbauer}, 546 U.S. 1, 4 (2005) (per curiam) (interpreting \textit{Berger’s} holding as sounding in due process).
cumulative-prejudice analysis has remained a core component of constitutional criminal discovery doctrine.\footnote{See Kyles v. Whitley, 514 U.S. 419, 439-40 (1995).}

In addition to\textit{ Mooney} and \textit{Berger}, a third major due process discovery case emerged in 1935. That year a Kansas jury convicted Harry Pyle of a terrible series of crimes involving murder, torture, and theft.\footnote{Kansas v. Pyle, 57 P.2d 93, 96-98 (Kan. 1936).} Despite the heinous nature of these offenses,\footnote{The murder victim’s brother committed suicide after testifying against Pyle, “apparently because of the aftereffects of the torture” the defendants had inflicted upon him. \textit{Id.} at 98.} the Supreme Court’s per curiam decision in \textit{Pyle v. Kansas} opened the door to a significant expansion in constitutional criminal discovery rights and duties.\footnote{317 U.S. 213, 216 (1942).}

The Court held that Pyle’s “inexpertly drawn” but unrefuted pro se habeas petition required a hearing on claims that prosecutors convicted him through the knowing use of perjured testimony.\footnote{\textit{Id.} at 215-16.} That holding was consistent with \textit{Mooney}, but the Court went further. Pyle articulated a separate and distinct claim that prosecutors engaged in “the deliberate suppression . . . of evidence favorable to him” beyond the facts relating to the allegedly perjured testimony.\footnote{\textit{Id.} (emphasis added).} Citing \textit{Mooney}, the Court held that Pyle’s claims, if proved, would require his release.\footnote{\textit{Id.} at 216. On remand, the Kansas Supreme Court found no evidence of perjury or suppression of favorable evidence. Pyle v. Amrine, 156 P.2d 509, 521 (Kan. 1945). Pyle’s son was pardoned in 1959 after his conviction for these crimes. Alvin Dumler, \textit{Mass Murder Happened Before in Southwest Kansas}, \textit{Hutchinson News} (Kan.), Nov. 22, 1959, at 2, available at http://hutchnews.com/www/clusterpdfs/clustermurders-day7-othermassmurders.pdf. But Bedau and Radelet appear mistaken in stating that “Pyle was released.” Hugo Adam Bedau & Michael L. Radelet, \textit{Miscarriages of Justice in Potentially Capital Cases}, 40 STAN. L. REV. 21, 153 & n. 791 (1987) (citing \textit{Pyle}, 317 U.S. at 214); see also Kansas v. Marsh, 548 U.S. 163, 190-93 (2006) (Scalia, J., concurring) (compiling and discussing criticisms of Bedau-Radelet study).}

The Court applied the \textit{Pyle} “favorable evidence” rule fifteen years later in another per curiam opinion. In \textit{Alcorta v. Texas}, a capital case, the Court granted a pro se petition for \textit{certiorari}, stayed execution, and reversed the Texas courts’ denials of postconviction relief.\footnote{355 U.S. 28, 28-32 (1957) (per curiam).} The reason: the prosecutor had suppressed evidence that Alcorta could have used to impeach a key state witness.\footnote{\textit{Id.} at 30-32.} \textit{Napue v. Illinois} followed \textit{Alcorta} by imposing a due process duty on the prosecution to correct testimony that is known to be false, even when the evidence is not directly exculpatory and instead can be used to
challenge the credibility of a prosecution witness. The Napue Court reasoned, “A lie is a lie, no matter what its subject, and, if it is in any way relevant to the case, the district attorney has the responsibility and duty to correct what he knows to be false and elicit the truth.”

Brady v. Maryland became the eponymous due process discovery case by building on the core meanings developed from Mooney and Pyle through Alcorta and Napue. First, Brady clarified prosecutors’ due process duty to disclose evidence favorable to the defense that is “material to guilt or punishment.” A second key Brady holding rendered prosecutors’ good or bad faith regarding nondisclosure irrelevant. Initial refinements of these two holdings focused on the elusive definition of materiality.

Brady arose after the defendant was convicted of murder and sentenced to death. He later learned that his codefendant had confessed to committing the murder. The Court held the new evidence material to Brady’s defense. But the Court did not order a new trial. Instead, the Court ordered resentencing and gave some hints about the meaning of materiality in due process discovery doctrine. A key fact was Brady’s own confession to complicity in the murder. The Court reasoned that, in light of Brady’s confession, the undisclosed evidence would have been unlikely to alter the jury’s verdict on Brady’s guilt. In contrast, the Court held, evidence of shared culpability between the codefendants could have affected jurors’ views on Brady’s eligibility for the death sentence.

---

62 Id. The prosecutor elicited the same lie on redirect examination. Instead of holding the elicitation of that perjured testimony an independent due process violation under Mooney, the Court, instead, took it as proof of prejudice from the due process violation that occurred when the prosecutor suppressed the truth during cross-examination. The prosecutor proved the lie’s significance, the Court held, by deliberately eliciting the same lie on redirect. Id.
64 Id. at 87.
65 Id.
66 Id.
67 Id. at 84.
68 Id. at 87.
69 Id. at 87-88.
70 See id. at 84.
71 Id. at 89.
72 Id. at 88-91. The state resentenced Brady to life eight years later. Brady v. Superintendent, Anne Arundel Cnty. Detention Ctr., 443 F.2d 1307, 1309-10 (4th Cir. 1971).
After *Brady*, the Court clarified that due process disclosure duties encompass all material impeachment evidence. But the Court took another twenty-two years to define the term *material*. In crafting that definition in *United States v. Bagley*, the Court borrowed from new jurisprudence governing the Sixth Amendment right to effective assistance of counsel. *Bagley* grafted the prejudice definition from the ineffective assistance test onto due process doctrine governing prosecutors’ duty to disclose evidence beneficial to the defense. Therefore, to prove that undisclosed evidence is material, defendants must show “a reasonable probability” of a different result in the case had the undisclosed evidence been available to the defense. A “reasonable probability [is] a probability sufficient to undermine confidence in the outcome.”

A decade passed before the next significant refinement in due process disclosure doctrine. In *Kyles v. Whitley*, the Court undertook a detailed assessment of the alleged disclosure violations. Such detail was necessary, according to the concurring Justices, because prosecutors serving under Orleans Parish District Attorney Harry Connick, Sr. committed “blatant and repeated violations of a well-settled constitutional obligation” to reveal favorable evidence. Ordering a new trial, the Court clarified due process disclosure doctrine in several important ways. First, *Kyles* expanded on the key *Brady* holding that prosecutors’ good or bad faith regarding undisclosed evidence is irrelevant. *Kyles* reiterated the longstanding requirement of *Brady’s* due process discovery doctrine, imposing a disclosure duty upon prosecutors even in the absence of any defense request for the information. *Id.*

---

75 *Id.* at 682-83 (citing Strickland v. Washington, 466 U.S. 664, 668 (1984)).
76 *Id.*
77 *Id.* at 682. In another key holding, *Bagley* eliminated the “demand” requirement of *Brady’s* due process discovery doctrine, imposing a disclosure duty upon prosecutors even in the absence of any defense request for the information. *Id.*
79 *Id.* at 455 (Stevens, J., concurring). Kyles’s first trial ended in a hung jury. His second trial began on December 6, 1984 and ended in his conviction and death sentence. Petitioner’s Brief at *22, Kyles v. Whitley, 514 U.S. 419 (1995) (No. 93-7927) [hereinafter Kyles Pet. Br.]. On the same morning that Kyles’s second trial began, Raymond T. Liuzza, Jr., “son of a prominent New Orleans business executive,” was murdered in front of his home. Connick v. Thompson, 131 S. Ct. 1350, 1371 (2011) (Ginsburg, J., dissenting). John Thompson was convicted and sentenced to death for Liuzza’s murder. *Id.* at 1374. Thompson’s trial, like Kyles’s, took less than three days, including death-qualification of the jury, the guilt/innocence phase, and the sentencing phase. Compare Kyles Pet. Br. at *6 (trial completed December 6-8, 1984), with Petitioner’s Brief at 12, Connick, 131 S. Ct. 1350 (No. 09-571) (trial completed May 6-8, 1985).
80 *Kyles*, 514 U.S. at 421.
81 *Id.* at 433, 437-38.
principle that prosecutors are responsible for disclosing evidence in their own files as well as evidence held by other prosecutors in the same office.82 But Kyles also explained that prosecutors must obtain and disclose evidence held by case investigators that is materially beneficial to the defense, whether or not prosecutors know the evidence exists.83

Kyles retained Berger’s cumulative-prejudice principle.84 But Kyles also focused closely on the link between the Brady due process materiality test and the Sixth Amendment prejudice test for ineffective assistance of counsel under Strickland v. Washington.85 There are three critical components of this doctrinal link. First, the defendant’s burden to prove a “reasonable probability” of a different outcome is less than a preponderance of the evidence.86 Putting the familiar preponderance test in the Brady context, defendants need not advance the prejudice ball past the fifty-yard line; they do not have to show that it is more likely than not that the cumulative effect of the undisclosed evidence would have led to a different verdict.87

Kyles also clarified the distinction between Brady materiality and Strickland prejudice, on one hand, and the test for sufficiency of the evidence, on the other.88 The latter test is governed by Jackson v. Virginia.89 Under Jackson, the defense must prove that no reasonable juror would vote for conviction when the evidence is viewed in a light most favorable to the prosecution.90 The Kyles Court held that the demanding Jackson standard is inappropriately onerous, in the context of prosecutors’ due process duties, to ensuring that defendants receive fair trials resulting in verdicts that reliably warrant public (and judicial) confidence.91 In the constitutional discovery and ineffective assistance settings, therefore, courts must weigh the strength and weakness of all the evidence that is presented by both parties in both the trial and postconviction proceedings.92

---

82 Id. at 432-37.
83 Id. at 437; see also Youngblood v. West Virginia, 547 U.S. 867, 869-70 (2006) (police suppression of exculpatory evidence violates Brady).
84 Kyles, 514 U.S. at 436.
85 Id. at 437.
86 Id. at 434-35.
87 Id.
88 Id.
90 Id.
91 Kyles, 514 U.S. at 437-38.
92 Id. at 434-35 & n.8.
Finally, Kyles emphasized the significant responsibility that accompanies prosecutors’ authority to make discretionary decisions about disclosing Brady-line evidence. The Court cautioned that “a prosecutor anxious about tacking too close to the wind will disclose a favorable piece of evidence.”

II. DUE PROCESS TAKES A DIVE

New Orleans prosecutors did not “tack[] too close to the wind” in John Thompson’s case. They sank the ship. The Louisiana Court of Appeal concluded that Orleans Parish prosecutors put Thompson on death row through the “intentional hiding of exculpatory evidence.” That evidence, when unearthed by a determined defense investigator, contributed to dismissal and acquittal on the robbery and murder charges, respectively, that the prosecutors had used to seek Thompson’s execution.

Thompson’s case is one of many in which nondisclosure of exculpatory evidence imposed unnecessary harm. Criminal discovery reform is necessary because the constitutional disclosure duties traced in Part I suffer from limited scope and weak enforceability. Noncompliance and significant system costs are the predictable results. To cite examples from John Thompson’s jurisdiction alone, numerous reported Brady cases from Hary Connick’s tenure as Orleans Parish District Attorney involved prosecutors failing to disclose evidence to the

93 Id. at 439. Kyles spent more than a decade on Louisiana’s death row before the Supreme Court ordered a new trial. Id. at 421. After three subsequent trials ended in hung juries, Connick’s office dismissed the charges. JED HORNE, DESIRE STREET: A TRUE STORY OF DEATH AND DELIVERANCE IN NEW ORLEANS 317-21 (2005). In July 2010, Kyles was charged with the first-degree murder and second-degree kidnapping of Crystal St. Pierre, arising from an alleged dispute over the value of a food stamp card. See Alan Powell II, Curtis Kyles May Have Killed Woman over Food Stamp Card, TIMES-PICAYUNE (July 7, 2010, 4:18 PM), http://www.nola.com/crime/index.ssf/2010/07/curtis_kyles_may_have_killed_w.html; see also If Kyles Killed . . ., JEDHORNE.COM, http://jedhorne.com/2010/06/did-kyles-kill/ (last visited July 3, 2011).
94 Kyles, 514 U.S. at 439.
97 See supra note 1; see infra note 100.
98 There is near unanimity among courts and commentators that enforceability of Brady-line disclosure duties has remained problematic from the outset. Johns, supra note 5, at 516-21; Daniel Medwed, Brady’s Bunch of Flaws, 67 WASH. & LEE L. REV. 1533 (2010); Ridolfi, supra note 32, at 2030-31 (existing disclosure rules “are not doing enough because they are inadequate and sometimes not enforced at all”); Stephanos Bibas, Brady v. Maryland: From Adversarial Gamesmanship Toward the Search for Innocence?, in CRIMINAL PROCEDURE STORIES 129 (Carol Steiker ed., 2006). For an expression of judicial frustration with the development of Brady doctrine, see United States v. Oxman, 740 F.2d 1298, 1309-11 (3d Cir. 1984).
defense for timely use at trial. Discovery violations required new trials in seventeen of those cases. In the most recent case, Smith v. Cain, the Court vacated a quintuple-murder conviction.

Such statistics, and the associated inefficiencies and unfairness, result from inherent flaws in the governing doctrine and from practical realities confronting those charged with implementation. At the doctrinal level, the first critical weakness is Brady’s mistaken assumption that prosecutors are as well equipped as defense attorneys to recognize the exculpatory or impeachment value of particular pieces of evidence. Second, Brady’s built-in materiality prejudice standard requires prosecutors to assess, ex ante, a question that often can be answered only ex post: whether the cumulative

---


100 Smith, 132 S. Ct. at 630-31; Kyles, 514 U.S. at 421; Mahler, 537 F.3d at 503; Monroe, 607 F.2d at 152-53; Davis, 479 F.2d at 453; Perez, 2008 U.S. Dist. LEXIS 1660, at *59-64; Bright, 875 So. 2d at 44; Knapper, 579 So. 2d at 961; Rosiere, 488 So. 2d at 970-11; Perkins, 423 So. 2d at 1107-08; Curtis, 384 So. 2d at 397; Falkins, 356 So. 2d at 416-19; Carney, 334 So. 2d at 419; Lindsey, 844 So. 2d at 969-70; Kemp, 828 So. 2d at 546; Thompson, 825 So. 2d at 557-58; Lee, 778 So. 2d at 667; In State v. Parker, 361 So. 2d 226, 227 (La. 1978), the Court ordered a new trial because it was impossible to reconstruct the record in order to litigate the Brady claim on appeal; the defendant then withdrew the appeal.

101 132 S. Ct. at 630. By an 8-1 vote, the majority tersely rejected Justice Thomas’s conclusion, based on a detailed assessment of the evidence, that the defendant failed to satisfy Brady’s materiality standard. Instead, the Court emphasized conflicts between the pretrial statements and trial testimony of the lone eyewitness to link the defendant to the murders.
impact of evidence beneficial to the defense would have created a reasonable possibility of a different result if it had been disclosed to the defense in time to be used during the investigation and litigation of the original proceeding.\textsuperscript{102}

The two flaws are linked. From some perspectives, \textit{Brady}'s materiality test imposes upon prosecutors as much a duty of divination as disclosure.\textsuperscript{103} Cognitive phenomena such as tunnel vision, groupthink, confirmation bias, and avoidance of cognitive dissonance raise additional psychosocial barriers to \textit{Brady} compliance.\textsuperscript{104} These pervasive, unconscious patterns of cognition can trump even the best of prosecutorial intentions.

\textit{Brady}'s enforceability took another hit at the doctrinal level when the Supreme Court shrank law enforcement's constitutional duties to investigate and retain exculpatory information. Investigators have no due process duty to investigate information that helps the defense.\textsuperscript{105} Although prosecutors are responsible for obtaining \textit{Brady} information from their investigative agents,\textsuperscript{106} officers can destroy potentially exculpatory evidence with impunity unless a defendant can

\begin{thebibliography}{100}
\bibitem{Kyles} \textit{Kyles}, 514 U.S. at 437 (“[T]he prosecution, which alone can know what is undisclosed, must be assigned the consequent responsibility to gauge the likely net effect of all such evidence and make disclosure when the point of [materiality] is reached.”); United States v. Agurs, 427 U.S. 97, 108 (1976) (“[T]he significance of an item of evidence can seldom be predicted accurately until the entire record is complete.”), overruled on other grounds by \textit{United States v. Bagley}, 473 U.S. 667, 682 (1985); see Cassidy, \textit{supra} note 37, at 1437-45.
\bibitem{Kyles2} \textit{Kyles}, 514 U.S. at 439 (“[T]he character of a piece of evidence as favorable will often turn on the context of the existing or potential evidentiary record.”).
\bibitem{Baker} Baker v. McCollan, 443 U.S. 137, 145-46 (1979). Several circuits have distinguished \textit{Baker}. See Everson v. Leis, 556 F.3d 484, 498-99 (6th Cir. 2009) (arresting officer must consider both “inculpatory and exculpatory evidence” in determining whether there is probable cause to arrest); Russo v. City of Bridgeport, No. 05-4302-CV, 2007 U.S. App. LEXIS 16171, at *23-30 (2d Cir. Feb. 27), \textit{cert. denied}, 552 U.S. 818 (2007) (allowing civil rights claim to proceed under Fourth Amendment where arresting officers detained plaintiff while failing to investigate readily available exculpatory evidence) (collecting cases and correcting \textit{Russo v. City of Bridgeport}, 479 F.3d 196 (2d Cir. 2007)); see also Fisher, \textit{supra} note 4, at 1399-1401 (describing police duty to investigate exculpatory evidence under England’s “tough on crime” Criminal Procedure and Investigation Act 1996, ch. 25 (Eng.)).
\bibitem{Youngblood} Youngblood v. West Virginia, 547 U.S. 867, 869-70 (2006) (per curiam) (“\textit{Brady} suppression occurs when the government fails to turn over even evidence that is ‘known only to police investigators and not to the prosecutor.’” (quoting \textit{Kyles}, 514 U.S. at 438)). On remand, the West Virginia Supreme Court ordered a new trial over heated dissents. State v. Youngblood, 650 S.E.2d 119, 132-33 (W. Va. 2007) (Davis, C.J.); \textit{id.} at 134-36 (Benjamin, J., dissenting); \textit{id.} at 137-40 (Maynard, J., dissenting).
\end{thebibliography}
prove bad faith.\textsuperscript{107} At a practical level, when law enforcement does obtain and preserve \textit{Brady} evidence, economic realities undercut the enforceability of prosecutors’ constitutional disclosure duties. Heavy caseloads create an independent hurdle to compliance, as prosecutors are charged with identifying and disclosing \textit{Brady} material held by all prosecutorial staff and their agents, including law enforcement.\textsuperscript{108} Political pressure can be a factor as well. Prosecutorial elections are often dominated by “tough on crime” ideologies, which tend to be unsympathetic toward procedures that help defendants avoid or reduce punishment.\textsuperscript{109} \textit{Brady}’s enforceability sustained another blow when the Supreme Court declined to impose a due process duty to disclose impeachment evidence at the plea-bargaining stage, where the overwhelming majority of cases are resolved.\textsuperscript{110} The Court also held that the right to \textit{Brady} material does not apply in the specialized pretrial setting of the grand jury hearing.\textsuperscript{111} At the postconviction stage, the same five-member majority common to \textit{Connick} and \textit{Garcetti} rejected a claim that due


\textsuperscript{108} Youngblood, 547 U.S. at 869-70. The Supreme Court has not ruled on the question whether law enforcement officers may face civil liability under 42 U.S.C. § 1983 for deliberately suppressing or otherwise failing to disclose \textit{Brady} material. For a canvassing of divided opinions on this issue in the federal courts of appeals, see Moldowan v. City of Warren, 578 F.3d 351, 377-81 (6th Cir. 2009), cert. denied, 130 S. Ct. 3504 (2010); id. at 401-04 (Kethledge, J., dissenting in part); see also Smith v. Almada, 640 F.3d 931, 939 (9th Cir. 2011) (explaining that police may face liability under 42 U.S.C. § 1983 (2006)); Jennifer E. Laurin, \textit{Rights Translation and Remedial Disequilibration in Constitutional Criminal Procedure}, 110 COLUM. L. REV. 1002, 1016-32, 1062-65 (2010) (analyzing \textit{Jean v. Rice}, 945 F.2d 82 (4th Cir. 1991)). For a discussion of “Brady” lists, which target law enforcement officers for discipline when they fail to provide prosecutors with exculpatory and impeachment evidence, see Nazir v. \textit{City of Los Angeles}, No. CV 10-06546 SVW (AGRx), 2011 U.S. Dist. LEXIS 26820 (C.D. Cal. Mar. 2, 2011) (prosecutor’s office shielded by Eleventh Amendment immunity from lawsuit by officer fired for nondisclosure of \textit{Brady} evidence); Walters v. Cnty. of Maricopa, No. CV 04-1920-PHX-NVW, 2006 U.S. Dist. LEXIS 60272 (D. Ariz. Aug. 22, 2006) (describing “Brady list” policy); Christopher N. Osher, \textit{Denver Cops’ Credibility Problems Not Always Clear to Defenders, Juries}, DENVER POST (July 10, 2011, 1:00 AM), http://www.denverpost.com/search/ci_18448755#ixzz1Rv2e571u (identifying “one out of every 17 Denver police officers as having discipline issues serious enough that their courtroom testimony may be suspect” and describing debate over effectiveness of procedures for notifying defense lawyers of such issues).


\textsuperscript{111} United States v. Williams, 504 U.S. 36 (1992).
process imposes an independent duty to produce evidence that can prove a defendant’s innocence, beyond the duty imposed by state statutes governing access to such evidence. As a practical matter, the majority of convicted defendants who are indigent lack access to counsel and other resources needed to investigate and litigate Brady claims. Thus, there is “little reason for these violations ever to come to light.”

The majority of postconviction Brady claims do not succeed, often because courts hold that undisclosed information was either immaterial or available to the defense through a reasonable investigation. Habeas jurisprudence combined with the 1996 Antiterrorism and Effective Death Penalty Act heightened procedural hurdles to federal judicial review of state-court Brady claims. Finally, even before the five-member Supreme Court majority restricted § 1983 liability in Connick and

---


114 Green, supra note 113, at 2175 n.68 (citing, inter alia, United States v. Jones, 620 F. Supp. 2d 163, 172 (D. Mass. 2009) (reported cases cannot measure scope of disclosure violations); United States v. Oxman, 740 F.2d 1298, 1310 (3d Cir. 1994) (confessing “the nagging concern that material favorable to the defense may never emerge from secret government files”).

115 See, e.g., supra notes 99-100 and accompanying text.

116 Rector, 120 F.3d at 560 (“evidence is not ‘suppressed’ if the defendant either knew, or should have known of the essential facts permitting him to take advantage of any exculpatory evidence” (citing West v. Johnson, 92 F.3d 1385, 1399 (5th Cir. 1998)); Barnes, 58 F.3d at 975 & n.4 (“Brady requires that the government disclose only evidence that is not available to the defense from other sources, either directly or through diligent investigation”); United States v. Zackson, 6 F.3d 911, 918 (2d Cir. 1993) (“Evidence is not ‘suppressed’ if the defendant either knew, or should have known, of the essential facts permitting him to take advantage of any exculpatory evidence.” (citations omitted)). Under this analysis, defendants must plead alternative ineffective assistance claims based on failure to investigate.

Garcetti, judicially imposed doctrines of absolute and qualified immunity rendered civil rights litigation nearly useless as a mechanism for enforcing due process disclosure duties.

The foregoing doctrinal and practical limitations seriously weaken the enforceability of prosecutors' due process disclosure duties. The next section focuses closely on the Connick and Garcetti cases as recent exemplars of the resulting inefficiency, uncertainty, and unfairness in the processing of criminal cases. Part IV contrasts the flawed Brady model with the relative simplicity and efficiency of full open file discovery. Highlighting the vitality of law and politics in this cutting-edge reform initiative, Part IV urges that the contrast between the two models should motivate broad enactment of full open file reform across jurisdictions.

III. DUE PROCESS SUBMERGED

Connick and Garcetti are linked in their subordination of prosecutors' constitutional discovery duties to other interests, including the protection of municipalities from financial liability for prosecutors' acts and omissions regarding Brady material. Analyzing the link between the two cases requires close scrutiny of their distinctive facts, procedural posture, and judicial reasoning. Subsections A and B focus on Connick v. Thompson. The more doctrinally complex Garcetti case is analyzed in Subsections C through E.

A. Connick v. Thompson: “Egregious” and “Intentional” Misconduct in Orleans Parish

John Thompson was twenty-two years old when he and codefendant Kevin Freeman were arrested for the murder of Raymond T. Liuzza, Jr., in New Orleans. Liuzza was shot to death in early December, 1984, during a robbery outside his home. “Because Liuzza was the son of a prominent executive,
the murder received a lot of attention.” 122 Three weeks after Liuzza’s murder, the three children of a man named LaGarde fought off a carjacking attempt near the Superdome. 123 After Thompson and Freeman were arrested in January for Liuzza’s murder, Thompson’s photo was published in the Times-Picayune. 124 LaGarde showed the photo to his children. They agreed that Thompson was the man who had tried to rob them. 125

Prosecutors charged Thompson with armed robbery. 126 Then they switched the order of the murder and robbery trials. 127 They hoped to use the robbery conviction to keep Thompson off the stand at the murder trial, to impeach him if he testified, and to win a death sentence. 128 One prosecutor told Thompson at the robbery trial, “I’m going to fry you. You will die in the electric chair.” 129 With the exception of that specific denouement, the prosecution’s strategy succeeded.

The strategy succeeded in part because Thompson’s lawyers and the jurors in his robbery trial did not know that a patch of bloody cloth exonerated him on that charge. During the investigation of the LaGarde robbery, law enforcement officers seized a blood swatch from the scene. 130 Prosecutors ordered the blood tested before Thompson’s trial for the LaGarde robbery. 131 The swatch tested type B, 132 but the prosecution disclosed neither the swatch nor the lab report to the defense. Thompson’s trial lawyer had asked the property technician before trial if there was any blood evidence. 133 He was told that “[t]hey didn’t have any.” 134 On the first day of the robbery trial, assistant prosecutor Gerry Deegan “checked all of the physical evidence in the case out of the police property room.” 135 The next day Deegan returned everything but the

---

122 Thompson, 553 F.3d at 843.
123 Id.
124 Id.
125 Id.
126 Id.
127 Id.
128 Id.
130 Id. at 1356 (majority opinion).
131 Id.
132 Id. at 1372-73 & n.5 (Ginsburg, J., dissenting).
133 Id.
134 Id.
135 Id. at 1356 (majority opinion).
blood swatch. At the time, Deegan was just out of law school and had worked in Connick’s office for less than a year.

During Thompson’s murder trial, the prosecution also failed to disclose several pieces of impeachment evidence. First, law enforcement had secretly taped the conversation between the Liuzza family and Richard Perkins, the man who originally incriminated Thompson in the Liuzza murder. At trial, Perkins denied seeking reward money. But on the tape, Perkins had said, “I don’t mind helping . . . but I would like [you] to help me and, you know, I’ll help you.” After the family had told Perkins that they wanted to try to help him, he implicated Thompson and Freeman. The prosecution also failed to disclose eyewitness testimony describing Liuzza’s killer as “six feet tall, with ‘close cut hair.’” That description matched Thompson’s codefendant, Freeman. In contrast, Thompson was five-feet, eight-inches tall and had a large Afro. The prosecution also failed to disclose pretrial statements attributed to the codefendant Freeman, the key witness against Thompson, which were materially inconsistent with Freeman’s trial testimony. Without the benefit of the foregoing evidence, and facing impeachment with the attempted robbery conviction if he took the stand, Thompson elected not to testify. He was convicted and sentenced to death.

Thompson was incarcerated for eighteen years after his arrest in January 1985. He spent fourteen of those years in twenty-three-hour-a-day solitary confinement in a six-by-nine foot windowless death row cell. He faced six different execution dates as his case moved through direct appeal, state postconviction, and federal habeas. A month before his final execution date, a last-ditch investigation in police archives unearthed a microfiche copy of the report documenting that the

---

136 Id.
137 Id. at 1372 n.3 (Ginsburg, J., dissenting).
138 Id. at 1371.
139 Id. at 1374.
140 Id. at 1371.
141 Id.
142 Id.
143 Id. at 1371-72.
144 Id. at 1374.
145 Id.
146 Id.
147 Id. at 1355 (majority opinion); Thompson v. Connick, 553 F.3d 836, 842, 865 (5th Cir. 2008).
LaGarde robber had type B blood. The defense tested Thompson’s blood. It was type O. The report proved Thompson innocent of the attempted robbery conviction that had kept him from testifying in his own defense at the murder trial and had helped the prosecution argue for his execution.

After the defense investigator uncovered the blood test results, a former prosecutor named Michael Riehlmann came forward. Riehlmann admitted that, five years earlier, Gerry Deegan had confessed to hiding exculpatory blood evidence in Thompson’s robbery case. Deegan was the recent law graduate who assisted with the prosecution of the robbery case; he confessed to hiding the evidence after learning that he was dying of cancer. Despite Riehlmann’s encouragement, Deegan did not come forward. For the next five years, neither did Riehlmann.

Based on the blood mismatch, the trial court vacated Thompson’s attempted robbery conviction and Connick’s office dismissed that charge. The trial judge in the murder case denied Thompson’s request for a new trial but vacated his death sentence. The appellate court ordered a new trial on the murder charge. The court found it unnecessary to rule on Thompson’s Brady claim but concluded that the prosecution’s “egregious” misconduct in the “intentional hiding of exculpatory evidence” caused violations of Thompson’s rights to present a defense and testify on his own behalf at the first trial. Connick’s office retried the murder charge but the codefendant, Freeman, had died in the interim. Without Freeman’s live testimony (the jury heard his prior testimony read back)—but with the benefit of the previously withheld evidence and Thompson’s testimony—the jury voted to acquit after thirty-five minutes’ deliberation.

---

151 Id.
152 Id.
153 Id. at 1356 & n.1.
154 Id.
155 Id. at 1374-75 (Ginsburg, J., dissenting).
156 Id.
158 Id.
159 Id.
160 Id. at 557-58.
161 Id. at 557.
163 Id.
B. Deliberate Indifference

The jury in Thompson's federal civil rights case found that the admitted *Brady* violation resulted from deliberate indifference to prosecutors' obvious need for training on their due process disclosure duties.\(^{164}\) Writing for the five-member majority, Justice Thomas limited the scope of the *Brady* violation at issue to nondisclosure of the lab report, which Connick conceded was a violation of Thompson's due process rights.\(^{165}\) The Court then held that Thompson had failed to meet his burden to prove deliberate indifference because he showed neither that there was a pattern of prior violations that should have notified defendants of the need for training nor that the need for training was otherwise obvious.\(^{166}\) On the latter point, the majority concluded that municipalities that might otherwise face liability for failure to train prosecutors on their due process discovery duties could reasonably rely on law schools, bar exams, continuing legal education programs, and professional disciplinary procedures to fill the breach.\(^{167}\)

In dissent, Justice Ginsburg found ample record evidence to support the jury’s finding of deliberate indifference, including multiple *Brady* violations by “no fewer than five prosecutors” who, “year upon year,” withheld evidence “vital to [Thompson’s] defense.”\(^{168}\) The dissenters also disagreed that Connick justifiably relied on law schools and other institutions

\(^{164}\) *Id.* at 1376.

\(^{165}\) *Id.* at 1357 & n.3 (majority opinion).

\(^{166}\) *Id.* at 1358-66 (citing *Canton v. Harris*, 489 U.S. 378 (1989); *Monell v. Dep’t of Soc. Servs.*, 436 U.S. 658 (1978)). The majority did not cite the unanimous 2009 decision in *Van de Kamp v. Goldstein* to apply absolute immunity to petitioners despite being urged to do so. *Petitioner’s Brief on the Merits* at 18, *Connick*, 131 S. Ct. 1350 (No. 09-571); *Amicus Curiae Brief of the Orleans Parish Assistant District Attorneys in Support of Petitioners* at 7-8, *Connick*, 131 S. Ct. 1350 (No. 09-571); *Brief of Amicus National District Attorneys Association at 16, Connick*, 131 S. Ct. 1350 (No. 09-571). *Van de Kamp* granted absolute immunity to prosecutors sued in their individual capacities for damages caused by their failure to sufficiently train and supervise their deputy district attorneys in preventing the nondisclosure of material impeachment evidence. *Van de Kamp v. Goldstein*, 555 U.S. 335, 339 (2009). On remand, the plaintiff settled for monetary damages from the city of Long Beach, *Goldstein v. City of Long Beach*, CV 04-9692, 2010 U.S. Dist. LEXIS 111195 (C.D. Cal. Aug. 13, 2010), after the District Court expressed “little doubt that Plaintiff w[ould] succeed in proving” that the withheld evidence satisfied *Brady* and sketched the likely jury instructions regarding the prosecutors' violations of their due process disclosure duties. *Id.* at *8-11.

\(^{167}\) *Connick*, 131 S. Ct. at 1361-63.

\(^{168}\) *Id.* at 1374-75, 1383 (Ginsburg, J., dissenting). Justices Scalia and Alito joined in a concurring opinion taking issue with Justice Ginsburg's analysis. *Id.* at 1366-70 (Scalia & Alito, J.J., concurring).
to fulfill the training requirements necessary to prevent due process disclosure violations in his office.\[169\]

Several aspects of the majority opinion elide core meanings of due process discovery doctrine. Greater fidelity to prosecutors’ truth-speaking and justice-seeking duties might have reinforced the doctrinal stability that encourages compliance with those duties.\[170\] Examples include the majority’s cabining of its analysis to what it defined as a “single \textit{Brady} violation” that occurred when one or more prosecutors failed to turn over the lab report.\[171\] Another is the conclusion that prior reversals of Orleans Parish convictions for \textit{Brady} violations were irrelevant to the jury’s finding of deliberate indifference because “[n]one of those cases involved failure to disclose blood evidence, a crime lab report, or physical or scientific evidence of any kind.”\[172\] Yet another is the empirical support (or absence thereof) in the majority’s assessment of whether the need for prosecutorial training was obvious.\[173\] Also noteworthy is the \textit{sotto voce} questioning in the majority opinion, rendered overt in the concurrence, of whether Orleans Parish prosecutors violated \textit{Brady} by failing to disclose the lab report since they did not know Thompson’s blood type.\[174\] These aspects of the majority opinion typify the diminution of core due process meanings that might influence adjudication of substantive constitutional criminal procedure claims, create uncertainty in the law, and reduce its deterrent effect.\[175\] This discussion focuses on the first and last points identified above.

1. \textit{Canton}-izing Cumulative Review

Cabining the misconduct in John Thompson’s case to the conceded nondisclosure of the lab report is inconsistent with the core due process principle that \textit{Brady} materiality

\[169\] \textit{Id.} at 1380-81, 1385-86 & nn.21-22 (Ginsburg, J., dissenting).
\[170\] See \textit{Laurin}, supra note 108, at 1058-72.
\[171\] \textit{Connick}, 131 S. Ct. at 1357-60.
\[172\] \textit{Id.} at 1390. At least two other Orleans Parish defendants sought \textit{Brady} relief after prosecutors failed to disclose lab reports documenting test results on biological samples; in both cases, courts held the undisclosed evidence to fail \textit{Brady’s} materiality test. Joseph v. Whitley, No. 92-2335, 1992 U.S. Dist. LEXIS 17385, at *5-12 (E.D. La. 1992); State v. Walter, 675 So. 2d 831, 833-35 (La. Ct. App. 1996).
\[173\] See supra note 21.
\[174\] Compare \textit{Connick}, 131 S. Ct. at 1357 & n.3 (majority opinion), with \textit{id.} at 1369-70 (Scalia & Alito, J.J., concurring).
must be assessed cumulatively.\(^{176}\) The majority subordinated that principle in favor of a 42 U.S.C. § 1983 remedial rule pulled from dicta in Canton v. Harris.\(^{177}\) To fit Canton’s Procrustean bed, the Connick majority trimmed Orleans Parish prosecutors’ acts and omissions down to a “single Brady violation” or a “single-incident.”\(^{178}\)

In considering the majority’s subordination of the cumulative materiality principle, it is important to acknowledge that no court ever expressly ruled that Connick’s office violated Brady in Thompson’s case. But the state Court of Appeal concluded that the prosecutors committed “egregious” misconduct through the “intentional hiding of exculpatory evidence.”\(^{179}\) That court also concluded that the “intentional hiding of exculpatory evidence” caused the violations of Thompson’s right to testify and present a defense, which in turn required a new trial on the murder charge.\(^{180}\) These tightly linked state court findings and conclusions comprise the essential elements of a Brady violation: Orleans Parish prosecutors failed to disclose evidence materially beneficial to the defense. The absence of an express judicial ruling on Thompson’s Brady claims should not have weighed against him in the subsequent civil rights litigation. And the state courts’ granting relief on alternate grounds certainly did not warrant subordination of cumulative prejudice analysis to Canton’s single-incident calculus.

Similarly, Connick’s admitting that his office violated Brady by failing to disclose the lab report did not support the majority’s elision of cumulative materiality. Both the content and the context of the admission make the point clear. Connick testified before the jury in Thompson’s civil rights case that he knew the lab report was Brady material.\(^{181}\) More specifically, he testified that he knew nondisclosure of the report was illegal because in a prior case he “got indicted by the U.S. Attorney” for failing to disclose a lab report to the defense.\(^{182}\)

\(^{176}\) Kyles v. Whitley, 514 U.S. 419, 436 (1995); see also Connick, 131 S. Ct. at 1377 & n.11 (Ginsburg, J., dissenting) (citing Kyles, 514 U.S. at 421) (jury not limited to nondisclosure of lab report).

\(^{177}\) Connick, 131 S. Ct. at 1359-69 (majority opinion) (discussing Canton v. Harris, 489 U.S. 378, 390-91 (1989)).

\(^{178}\) Id. at 1356, 1361.


\(^{180}\) Id.

\(^{181}\) Connick, 131 S. Ct. at 1380 n.13 (Ginsburg, J., dissenting).

\(^{182}\) Id. (“[A] prosecutor must disclose a crime lab report to the defense, even if the prosecutor does not know the defendant’s blood type: ‘Under the law it qualifies as Brady material. Under Louisiana law we must turn that over. Under Brady we must
Connick’s sworn admission grounds a reasonable inference, apparently reached by the federal jury, that he had ample notice of the need to train his line prosecutors on *Brady* compliance. Connick’s admission on the lab report expressly references a prior *Brady* allegation directly on point in Thompson’s civil rights case. The admission weighs in favor of incorporating the core cumulative prejudice principle into the section 1983 analysis and weighs against the majority’s subordinating that principle to *Canton*’s single-incident doctrine.

2. See No Evil: Incorporating a Subjective Knowledge Element into *Brady* Analysis

Connick’s admission regarding the lab report took an even stronger turn on appeal of the civil rights verdict. There, the defendants argued that “the blood evidence was obviously exculpatory” and that nondisclosure was “a clear violation of the law.”

Once again, the state Court of Appeal’s reasoning seems significant. Even setting aside the rather abundant impeachment evidence that prosecutors failed to disclose in addition to the conceded nondisclosure of the lab report, the record also contains the state Court of Appeal’s conclusion that prosecutors intentionally hid exculpatory evidence. It is not clear that prosecutors intentionally withheld the lab report. But Deegan admitted to deliberately hiding exculpatory evidence; he took the physical evidence out of police custody and returned everything but the blood swatch. Applying cumulative materiality analysis, the swatch had exculpatory value independent of the lab report, and the prosecution had a duty to disclose it regardless of the existence of any lab report.

---

183 *Connick*, 553 F.3d 836, 857 (5th Cir. 2008); see also *Connick*, 131 S. Ct. at 1372 n.4 (Ginsburg, J., dissenting) (“Any reasonable prosecutor would have recognized blood evidence as Brady material.... [a]ccordingly, the proper response was ‘obvious to all.’”). The court of appeals rejected this “obvious violation” defense not because it disagreed with the statement but because while Connick conceded his *Brady* duty to turn over the blood evidence, his line prosecutors did not. *Connick*, 553 F.3d at 857.

184 See supra notes 138-44 and accompanying text.

185 *Thompson*, 825 So. 2d at 557.

186 *Connick*, 131 S. Ct. at 1356 (lead prosecutor denied seeing the report).

187 *Id.*
This is so because the swatch, on its own, would have alerted the defense to the need to check Thompson’s blood type.188

On this point, and of the remaining highlighted aspects of the majority and concurring opinions, perhaps the most powerful evidence of Brady’s anemic enforceability is the conclusion of Justices Scalia and Alito regarding the lab report. Their concurring opinion reasoned that the failure to disclose the lab report revealing the blood type of the LaGarde robber was probably not a Brady violation because prosecutors did not know John Thompson’s blood type.189 Notably, the opening lines of the relatively terse majority opinion also emphasized the absence of proof that prosecutors tested Thompson’s blood or “knew what his blood type was.”190 Justices Scalia and Alito went further. They denied awareness of any Supreme Court case law supporting what they described as the dissent’s “sub silentio expansion of the substantive law of Brady” to encompass a duty to reveal “untested” evidence that could exonerate a defendant.191

To be fair, the concurrence tweaked the dissent for citing no case in which a Brady violation rests solely on a withheld lab report.192 Nevertheless, the fact that two United States Supreme Court Justices could view the lab report documenting the robber’s blood type as comprising untested evidence underscores the many profound problems undermining Brady enforceability. Similarly revealing is the concurring Justices’ citation of Arizona v. Youngblood193 to support their not-very-silentio circumscription of prosecutors’ constitutional disclosure duties to evidence that they subjectively know to be exculpatory. As the concurrence acknowledged, subjective knowledge is irrelevant to Brady analysis.194

---

188 See also id. at 1373 & n.5 (Ginsburg, J., dissenting) (noting that at the federal trial, the jury was told of the parties’ stipulation that the prosecution did not inform the defense “of the existence of the blood evidence, that the evidence had been tested, [or] that a blood type was determined definitively from the swatch” (emphasis added)).

189 Id. at 1369 (Scalia & Alito, JJ., concurring).

190 Id. at 1356 (majority opinion).

191 Id. at 1369 (Scalia & Alito, JJ., concurring).


194 Connick, 131 S. Ct. at 1369.
In fact, prosecutors’ disclosure duties were not squarely at issue in *Youngblood*. The case did not involve prosecutors’ deliberate suppression of biological evidence or nondisclosure of potentially exculpatory lab reports. The problem in *Youngblood* was the inability to test the biological evidence at issue (semen samples) because one set of samples was too small and investigators failed to preserve another sample for testing. The Court’s chief concern in *Youngblood* was to avoid burdening the nation’s myriad law enforcement offices by imposing a due process duty to preserve potentially exculpatory evidence in all cases.

The citation of *Youngblood* by the *Connick* concurrence is a transitive error. A forensic sample that cannot be tested, like the semen in *Youngblood*, is distinguishable from the tested blood swatch in John Thompson’s case. The blood swatch tests yielded dispositive results. Those results were recorded in a lab report. The lab report revealed what the defense, with minimal investigation, could establish to be wholly exculpatory evidence.

That type of tested forensic evidence and forensic report defines materiality under *Brady*. This is so in part because the full value of exculpatory evidence can only be determined ex post, in the context of an adequate defense opportunity to investigate and litigate its meaning at trial. Those opportunities were denied to both *Youngblood* and Thompson. But unlike *Youngblood*, Thompson’s only impediment to investigating and litigating the exculpatory value of the retained, tested evidence was the prosecution’s failure to disclose it.

The concurrence failed to mention another reason that *Youngblood* supports rather than refutes Thompson’s position: *Youngblood* was innocent. Like Thompson, he was imprisoned for years before evidence of his wrongful convictions led to his release. In *Youngblood*’s case, it took fifteen years before advances in DNA technology allowed testing on the remnants of the biological evidence. Those tests led authorities to

---

195 488 U.S. at 53-54 (discussing trial court’s denial of state’s motion to test defendant’s blood and saliva because semen sample was inadequate for valid comparison); cf. *State v. Youngblood*, 734 P.2d 592, 595-96 (Ariz. Ct. App. 1986), rev’d, 488 U.S. 51 (1988) (“This is not a case where the samples were available for defendant’s examination.”). *But see* *State v. Youngblood*, 844 P.2d 1152, 1154 (Ariz. 1993) (semen sample “was available to the defendant at trial and the defendant chose not to perform tests of his own”).

196 *Youngblood*, 488 U.S. at 58.


198 *Id.*
identify and convict the child sex offender who had kidnapped and sodomized the ten-year-old victim.\textsuperscript{199} The silence of the \textit{Connick} concurrence on the full story of \textit{Youngblood} tracks the corresponding subordination by both the majority and the concurring opinions of the prosecutors’ core due process duties, developed from \textit{Mooney} through \textit{Kyles}, to speak truth and seek justice. The strained reasoning in those opinions with respect to the blood swatch, the lab report, and the undisclosed impeachment evidence significantly weakens \textit{Brady} enforceability. The majority’s wink-and-nod to due process disclosure violations highlights the same majority’s elision of constitutional discovery duties in \textit{Garcetti}.

C. \textit{Garcetti} v. Ceballos: Riding the Rims to the Junkyard Gate

According to the defense attorneys for Michael Cuskey and Randy Longoria, the events that culminated in the U.S. Supreme Court’s decision in \textit{Garcetti} v. \textit{Ceballos} arose because Sheriff’s deputies in Pomona County, California lacked sufficient evidence to arrest Cuskey and Longoria for running a chop shop.\textsuperscript{200} After his arrest, Cuskey threw fuel on the fire by suing the Sheriff’s department.\textsuperscript{201} On this theory, when deputies spotted a stripped-down, stolen truck near Cuskey’s junkyard,\textsuperscript{202} they saw a chance to link the truck to the defendants. They asked their supervisor, Detective Wall, to obtain a warrant to search the junkyard.\textsuperscript{203} Based on the deputies’ statements, Wall swore out an affidavit describing “tire tracks which appeared to match the tread pattern” of the stolen truck leading to “the end of a long driveway” and the fence that marked Cuskey’s property line.\textsuperscript{204}

The magistrate issued the warrant. The deputies searched the property. They found no evidence to support their chop shop suspicions.\textsuperscript{205} But while the warrant focused on stolen vehicle parts, the deputies brought along a drug dog that sniffed out a small amount of methamphetamine.\textsuperscript{206} The

\begin{itemize}
  \item \textsuperscript{199} \textit{Id.}
  \item \textsuperscript{201} \textit{Id.} at 238.
  \item \textsuperscript{202} \textit{Id.} at 257 (where court asked whether the property was a “junkyard,” one deputy demurred, “every man, it’s his castle,” but agreed “there were car parts on that property, there were birds, there were animals, there were refrigerators . . . all kinds of stuff,” including what “looked like an alligator . . . ”).
  \item \textsuperscript{203} Joint App’x Vol. I at 28, \textit{Garcetti}, 547 U.S. 410 (No. 04-473), 2005 WL 1620385.
  \item \textsuperscript{204} Joint App’x Vol. III at 497, \textit{Garcetti}, 547 U.S. 410 (No. 04-473), 2005 WL 1620386.
  \item \textsuperscript{205} \textit{Id.} at 498.
  \item \textsuperscript{206} \textit{Id.}
\end{itemize}
deputies also seized some firearms. 207 Cuskey, Longoria, and a third defendant named Ojala were charged with felony drug and weapon possession offenses. 208 Ojala pleaded guilty and began serving a one-year jail sentence. 209

Meanwhile, the lawyers for Cuskey and Longoria investigated the allegations in the search warrant. Based on that investigation, they alleged that the deputies had lied to obtain the affidavit. The junkyard was not at the end of a long driveway but on a road that was the length of a football field. 210 The road was thirty- to forty-feet wide, “equal to if not greater than a standard residential street.” 211 At least nine other residences faced onto the same road; all of those homeowners had to use the road to enter and exit their properties. 212 The road was made of asphalt, gravel, and dirt. There was “absolutely no way” tire-tread patterns could lead from the truck to Cuskey’s junkyard fence. 213 In the one spot where tread marks appeared, it was “almost impossible” to tell from which of the ten properties they originated. 214

The defense lawyers filed a motion to traverse the warrant. 215 If granted, the motion would have required dismissal of the charges. 216 The defense lawyers gave their investigative materials to Richard Ceballos. As supervising prosecutor, Ceballos had authority to dismiss the case. 217 Ceballos investigated the defense allegations and consulted with supervisors and colleagues. 218 All the prosecutors agreed “there was a problem with the warrant.” 219

Ceballos then talked with Detective Wall, the officer who had sworn out the affidavit supporting the warrant. 220 Ceballos confronted Wall with the defense allegations that the tire tread marks running from the truck to Cuskey’s gate were “a figment of the deputy’s imagination, that they had lied, that

---

207 Id.
208 Id. at 406.
209 Id. at 496.
210 Id. at 499.
211 Id.
212 Id.
213 Id.
214 Id. at 500.
217 Id.
218 Id.
219 Id. at 39.
220 Id. at 37.
they had made up the fact that they had seen these tire tracks.” Ceballos later told Wall that the charges would likely be dismissed because of the affidavit’s “grossly inaccurate” description of the road.

Wall then made a crucial admission. He told Ceballos that he had spoken to the deputies. He said that the affidavit should be “modified.” He wanted to change “tire tracks” to “tire gouges," caused by the rims of the stolen truck scraping along the roadway.” Ceballos asked when the officers decided to modify the affidavit. Wall did not answer. Ceballos wrote a memorandum recommending dismissal of the charges based on the misrepresentations and omissions in the affidavit. He cited Franks v. Delaware in support.

Franks allows Fourth Amendment challenges to the veracity of affiants’ statements supporting warrants. The Franks Court reasoned that probable cause “would be reduced to a nullity” if an officer could use false evidence to obtain a warrant and “remain confident that the ploy’ could never be exposed through the adversarial process. Given the magistrate’s constitutional duties, the Court concluded, “it would be an unthinkable imposition upon his authority if a warrant affidavit, revealed after the fact to contain a

---

221 Id.
223 Id. at 503.
224 Id.
225 Id.
226 Id. The issue arose a few months after the Rampart investigation revealed large-scale corruption, including perjury, evidence planting, drug dealing, violent crimes, and other misconduct by the Los Angeles police department’s anti-gang unit. Erwin Chemerinsky, An Independent Analysis of the Los Angeles Police Department’s Board of Inquiry Report on the Rampart Scandal, 34 Loy. L.A. L. Rev. 545, 547-53 (as revised 2005). As the district court put it in ruling on Ceballos’s subsequent civil rights claim: [T]he code word “Rampart” says it all—there can be no doubt that, in Southern California, police misconduct is a matter of great political and social concern to the community. In recent years, various local law enforcement agencies have been severely criticized for what many believe to be serious misconduct on the part of police officers.
227 Joint App’x Vol. III, supra note 204, at 503 (citing Franks v. Delaware, 438 U.S. 154 (1978)).
228 Id.
229 Franks, 438 U.S. at 156.
230 Id. at 168.
deliberately or recklessly false statement, were to stand beyond impeachment.\textsuperscript{231}

\textit{Franks} challenges are equivalent to perjury allegations against law enforcement officers. Such challenges seldom succeed.\textsuperscript{232} In the rare cases in which \textit{Franks} relief is granted, however, it is based on the type of prevarication documented in Ceballos’s memorandum.\textsuperscript{233} Ceballos’s supervisor was concerned enough about the affidavit’s veracity that he authorized the release of the third defendant, who was already serving time on his guilty plea.\textsuperscript{234} But the supervisor also asked Ceballos to revise the memo. He wanted to give the Sheriff’s department a less accusatory explanation for dismissing the charges.\textsuperscript{235}

Although Ceballos edited the memo as requested, the deputies still accused him of acting like a defense lawyer. They worried that dismissal might give Cuskey grounds for another lawsuit.\textsuperscript{236} Ceballos’s supervisor decided against dismissing the charges. He chose instead to leave any decision on the \textit{Franks} allegations to the “black robe” presiding over the suppression hearing.\textsuperscript{237}

Before the \textit{Franks} hearing, but after consulting with his supervisors, Ceballos gave a redacted version of his memo to

\textsuperscript{231} Id. at 165. Further research is needed to trace the extent to which a concern to protect the court’s integrity connects doctrines underlying \textit{Brady}, \textit{Franks}, and “fraud on the court” analyses. For a definition of the latter allegation, see, for example, \textit{Herring v. United States}, 424 F.3d 384, 386-87 (3d Cir. 2005) (defining “demanding” burden of proving fraud on the court by “clear, unequivocal and convincing evidence” of intentional and successful deception of the court through “the most egregious conduct” by an officer of the court).


\textsuperscript{233} United States v. Whitley, 249 F.3d 614, 624 (7th Cir. 2001) (remanding case to determine if search warrant was deficient when officers altered versions of events and facts did not match description); United States v. Reinholz, 245 F.3d 765, 774-75 (8th Cir. 2001) (approving trial court’s refusal to “bolster” affidavit with new and different information beneficial to the prosecution); United States v. DeLeon, 979 F.2d 761 (9th Cir. 1992) (remanding case due to material omissions in affidavit).

\textsuperscript{234} Joint App’x Vol. I, supra note 203, at 42-44.

\textsuperscript{235} Id. at 113-18.

\textsuperscript{236} Id. at 49.

\textsuperscript{237} Id. at 117-18. Another supervisor reviewed the affidavit and the results of Ceballos’s investigation. He agreed that the affidavit contained “obvious material misrepresentations and omissions per \textit{Franks v. Delaware},” Joint App’x’x Vol. III, supra note 204, at 411-12. After the meeting, Ceballos confronted his supervisor about “kowtowing to the sheriff’s department, of ignoring the fact that these deputy sheriffs had lied, and of setting aside his obligations simply to appease the sheriff’s captain and lieutenant.” Joint App’x Vol. I, supra note 203, at 23.
the defense.\textsuperscript{238} He also insisted that he would have to testify for the defense if subpoenaed.\textsuperscript{239} His immediate supervisor warned that if he kept “thinking . . . [and] talking like that” he would get “in trouble.”\textsuperscript{240} Ceballos interpreted this warning as an attempt to dissuade him from testifying and as a threat to retaliate if he did so.\textsuperscript{241} He received the defense subpoena and decided that “regardless of the consequences” he was going to “tell the truth” at the suppression hearing.\textsuperscript{242}

Ceballos was not allowed to testify to his own conclusions about the deputies’ veracity. But he did testify that the deputies wanted to change the affidavit.\textsuperscript{243} Although the affidavit specifically stated that the deputies had followed “tire tracks matching the tread pattern” of the stolen truck, the deputies testified differently at the hearing. There, they described tracing a “scratch” or “indentation” made by a tire rim from the stolen truck to Cuskey’s gate.

Ceballos knew the presiding judge as “a good judge for prosecutors.”\textsuperscript{244} The judge recited the affidavit’s “specific” description of each of the truck’s four tires.\textsuperscript{245} The judge was unmoved by the affidavit’s failure to mention any flat tire or rim.\textsuperscript{246} He was similarly unmoved by disparities between the deputies’ descriptions of the flat tire.\textsuperscript{247} He credited the deputies’ testimony and denied the defense motion.\textsuperscript{248}

Shortly thereafter, Ceballos was demoted.\textsuperscript{249} He also was taken off a murder case and offered “freeway therapy” (a different job with a long commute).\textsuperscript{250} He exhausted the available grievance process and filed his civil rights action.\textsuperscript{251} He urged that his conduct was required by \textit{Brady v. Maryland} and that his supervisors’ retaliation violated the First Amendment and the Due Process clause of the Fourteenth Amendment.\textsuperscript{252}

\begin{footnotes}
\item[238] Joint App’x Vol. I, supra note 203, at 56.
\item[239] Id. at 57.
\item[240] Id.
\item[241] Id. at 58.
\item[242] Id. at 58-59.
\item[243] Joint App’x Vol. II, supra note 200, at 297-308.
\item[244] Joint App’x Vol. I, supra note 203, at 60.
\item[245] Joint App’x Vol. II, supra note 200, at 349 (“emergency tires . . . on the right front and left rear . . . a white spoke type wheel with a Sonic brand . . . on the left front . . . a Firestone . . . on the right rear”).
\item[246] Id. at 348-49.
\item[247] Id.
\item[248] Id.
\item[250] Id. at 13.
\item[252] Joint App’x Vol. I, supra note 203, at 139-40.
\end{footnotes}
D. Brady Duties in the Pretrial Context

With the core meanings of Brady-line duties firmly rooted in the mandate to speak truth and seek justice, Ceballos’s constitutional duty to speak out as he did might seem obvious. Contrary to the Supreme Court majority opinion, Ceballos did not write his memorandum because “he did not receive a satisfactory explanation for the perceived inaccuracies” in the affidavit.253 His independent investigation of the defense allegations confirmed that the affidavit could not be true. He confronted the lead detective and was told the sworn affidavit should be “modified.”254 Ceballos was not alone in concluding that this admission supported the perjury allegation and that he had a duty to turn the information over to the defense.255

The first set of federal judges to consider the issue agreed that Ceballos had a duty to inform his supervisors about what he had learned. The district court judge cited Brady to hold that when Ceballos wrote his memorandum, “he was complying with his (and the government’s) duties under the due process clause of the Fifth and Fourteenth Amendments not to introduce or rely on evidence known to be false.”256 On appeal, two Ninth Circuit judges concluded that, taking the evidence in a light most favorable to Ceballos, his memorandum comprised “[g]ood-faith statements made in pursuit of [the] obligation” imposed upon prosecutors by their “duty to disclose information favorable to an accused, including information relating to a witness’s veracity and integrity.”257 The concurring judge also viewed the memorandum as performing “the basic communicative duty Brady imposes on ‘the prosecution.’ . . . Ceballos was simply ‘doing what he was supposed to do’ as a deputy district attorney carrying out non-discretionary quintessentially ‘prosecutorial functions.’”258

The district court and Ninth Circuit drew upon the parties’ pleadings and summary judgment arguments. But if

253 Garcetti, 547 U.S. at 414.
254 Joint App’x Vol. III, supra note 204, at 503.
255 Id. at 411-12 (memorandum of Deputy District Attorney Michael Grosbard, concluding that Franks required disclosure of investigators’ “obvious material misrepresentations and omissions”); Kyles v. Whitley, 514 U.S. 419, 439 (1995) (“[A] prosecutor anxious about tacking too close to the wind will disclose a favorable piece of evidence.”).
257 Ceballos v. Garcetti, 361 F.3d 1168, 1179 (9th Cir. 2004).
258 Id. at 1189 & n.3 (O’Scannlain, J., concurring in the judgment).
(as it appears) the lawyers and judges were relying on the four corners of the Supreme Court’s *Brady*-line jurisprudence as the basis for Ceballos’s constitutional disclosure duty, they were overreaching. From *Mooney* through *Kyles*, the Supreme Court has consistently emphasized that constitutional criminal discovery duties and rights protect interests in a fair trial. None of the Supreme Court’s *Brady*-line cases impose a constitutional disclosure duty upon prosecutors in the pretrial setting.

To the contrary, by the time Ceballos wrote his memorandum, the Court had held that prosecutors have no duty to disclose exculpatory evidence before a defendant is indicted, at least in the specialized setting of the grand jury hearing. In *dicta* in *United States v. Agurs*, the Court also had indicated that there was no pretrial disclosure duty by stating that prosecutors could reveal *Brady* information “during the course of a trial.” The federal courts of appeals have generally agreed that prosecutors can wait until quite late in the game to comply with constitutionally mandated disclosure duties—even after defense cross-examination of a prosecution witness.

The litigants and the courts in Ceballos’s case were silent on the significance of this doctrinal problem in applying *Brady* to the disclosures prompted by Cuskey’s motion to traverse. On one hand, such silence may not have been wholly inappropriate. Ceballos had evidence indicating that law enforcement officers committed perjury. There were two opportunities for that evidence to have any meaningful effect: the memorandum recommending dismissal and, absent dismissal, the *Franks* hearing on the motion to traverse the warrant. On those facts, the core constitutional meaning of due process disclosure duties—speaking truth, seeking justice, disclosing material exculpatory and impeachment evidence—could have trumped the absence of any binding Supreme Court precedent on the need for pretrial disclosure.

On the other hand, the silence of the litigants and courts on the unsettled timing-of-disclosure doctrine may indicate inattention. A few weeks before the district court ruled

---

261 427 U.S. at 107.
on the summary judgment motions in Ceballos’s case, the Supreme Court granted certiorari in another Ninth Circuit case, United States v. Ruiz. That case raised the question whether Brady requires disclosure of impeachment evidence during plea bargaining. A few months later, and nearly two years before the Ninth Circuit ruled in Ceballos’s case, a unanimous Court held that no such duty exists. Since Ruiz arguably had implications for prosecutors’ pretrial disclosure duties beyond the plea bargaining context—potentially weakening Ceballos’s civil rights claim—one would reasonably expect to see some evidence that those implications were considered, had the litigants or judges done so.

This deepening silence raises the question whether any federal constitutional precedent imposed a specific disclosure obligation on Ceballos, in addition to his core due process duties to speak truth and seek justice, in the context of a pretrial Franks hearing. Although no court or commentator has discussed this question, the answer is yes.

In 1993—after the Supreme Court held that no disclosure duty applied in the grand jury setting, but long before Cuskey’s defense lawyer challenged Detective Wall’s affidavit—the Ninth Circuit became the only jurisdiction in the country to impose Brady-line disclosure duties on prosecutors in the latter, pretrial setting of a motion to traverse. In United States v. Barton, the defendant argued that his Brady rights were violated when law enforcement officers let his marijuana plants rot in their evidence locker. Detectives had seized the marijuana pursuant to a warrant. They obtained the warrant by swearing in an affidavit that they had smelled marijuana at Barton’s house.

265 Id. at 630-33. Specifically, the Ruiz Court held that due process does not bar requirements that defendants waive their rights to impeachment evidence as a condition of accepting a plea bargain; such a waiver does not render the plea involuntary. Id. at 630, 633.
266 United States v. Barton, 995 F.2d 931, 935 (9th Cir. 1993); cf. Mays v. City of Dayton, 134 F.3d 809, 816 (6th Cir. 1998) (where a Franks challenge alleges omission of disputed facts from warrant affidavit, defendants must make “a strong preliminary showing that the affiant with an intention to mislead excluded critical information from the affidavit, and the omission is critical to the finding of probable cause”). But see United States v. Stott, 245 F.3d 890, 902-03 & n.9 (7th Cir. 2001) (applying plain error analysis and finding no clear rule that Brady applies to suppression hearings; collecting cases).
267 995 F.2d at 932.
268 Id.
269 Id.
that statement. He argued that only marijuana plants with glandular trichomes emit any odor. Had his plants been preserved, he urged, he could have impeached the detectives with evidence that his plants had no glandular trichomes and were therefore odorless.

The Ninth Circuit affirmed the denial of Barton’s suppression motion. The panel reasoned that Barton failed to prove officers destroyed the plants in bad faith. But to reach that issue, the court first extended Brady protections to the pretrial setting of a Franks hearing. The panel concluded that, by violating due process discovery duties (in Barton, through destruction of material impeachment evidence), an agent of the prosecution “could feel secure that false allegations in his or her affidavit for a search warrant could not be challenged . . . [and] effectively deprive a criminal defendant of his Fourth Amendment right to challenge the validity of a search warrant.”

Excavating the due process disclosure rights and duties in Ceballos’s case leads to a previously unremarked and precise core constitutional meaning. At least in jurisdictions within the Ninth Circuit, prosecutors must seek justice, speak truth, and disclose material exculpatory and impeachment evidence—including when defendants challenge the veracity of affidavits supporting search warrants.

E. Discovery Duties Diminished

The Supreme Court’s five-member majority held that Ceballos had no First Amendment protection against retaliation for writing the memorandum recommending dismissal of the charges against Cuskey and Longoria because he conceded that such activity was part of his official duties as a government

270 Id. at 933.
271 Id.
272 Id. at 936.
273 Id. at 934-35 (citing Franks v. Delaware, 438 U.S. 154, 156, 168 (1978)).
274 Id. at 935.
employee. I will not recanvass the abundant commentary on what even the apparently lone scholarly supporter of Garcetti concedes to be the opinion’s undertheorized treatment of First Amendment workplace rights for government employees. The majority’s reasoning is formalistic and syllogistic. Ceballos recorded the Brady information in a charge disposition memorandum. Ceballos conceded that it was part of his job duties to produce charge disposition memoranda. The First Amendment does not protect communications that are part of a government employee’s job duties. Therefore, the First Amendment did not protect Ceballos from retaliation for writing his memorandum.

Concerns about federalism, separation of powers, and administrative efficiency played a strong role in Garcetti’s outcome. Those concerns rise to a peak in federal civil rights actions seeking oversight of local criminal justice systems. Garcetti also is part of a long and more particular trend toward concentration of power in, and deference to, the prosecution function as an administrative entity. Granting the effect of those concerns in Garcetti, the case carries important implications for prosecutors’ constitutional discovery duties—which, in the view of all of the lower federal court judges, required Ceballos to act as he did. The California Prosecutors Association and the Association of Deputy District Attorneys of

---

276 Garcetti v. Ceballos, 547 U.S. 410, 421-24 (2006). The Court did not address issues of qualified immunity. It would appear that the disclosure duty in the context of Cuskey’s Franks hearing was “clearly established” for purposes of section 1983 qualified immunity analysis under Barton. Wilson v. Layne, 526 U.S. 603, 617 (1999) (to establish defendant duty, plaintiffs may cite “controlling authority in their jurisdiction at the time of the incident which clearly established the rule on which they seek to rely”); Harlow v. Fitzgerald, 457 U.S. 800, 818 n.32 (1982) (declaring to determine “the circumstances under which the state of the law should be ‘evaluated by reference to the opinions of this Court, of the Courts of Appeals, or of the local District Court’” (quoting Procunier v. Navarette, 434 U.S. 555, 565 (1978) (reviewing federal district court and court of appeals decisions to determine whether the right at issue was “clearly established”))); Francis v. Coughlin, 891 F.2d 43, 46 (2d Cir. 1989) (“[T]he court must determine whether the decisional law of the Supreme Court or the appropriate circuit court has clearly established the right in question.”); see also Karen M. Blum, The Qualified Immunity Defense: What's “Clearly Established” and What's Not, 24 Touro L. Rev. 501, 515 (2008) (“controlling authority from the jurisdiction—that circuit’s court of appeals or the highest court of the state in which the case was sitting” may show law is “clearly established” for qualified immunity purposes).

277 See Tarkington, supra note 22, at 2176 & n.5 (citing scholarly commentary).

Los Angeles County urged the same position as *amici* before the Supreme Court. They begged the Court to protect prosecutors’ good-faith compliance with due process discovery duties.\textsuperscript{279} Yet the Justices gave the issue little attention. With few exceptions,\textsuperscript{280} scholars have followed suit.

Justice Kennedy’s majority opinion in *Garcetti* never referred directly to prosecutors’ constitutional discovery duties. A single citation to *Brady v. Maryland* was sandwiched between the opinion’s penultimate closing lines.\textsuperscript{281} Those lines referenced “safeguards in the form of . . . constitutional obligations apart from the First Amendment” and an assurance that such “imperatives . . . protect employees and provide checks on supervisors who would order unlawful or otherwise inappropriate actions.”\textsuperscript{282} The majority opinion was silent on Ceballos’s claim that it was precisely this imperative that required him to act as he did, and yet failed to protect him from alleged retaliation—an allegation that, like all other aspects of his complaint, had to be construed with every inference in his favor at the summary judgment stage.\textsuperscript{283}

Justice Stevens’s short dissenting opinion made no mention of prosecutors’ due process disclosure duties.\textsuperscript{284} Justice Souter’s dissenting opinion, in which Justices Stevens and Ginsburg joined, attributed to Ceballos the constitutionally protected “interest of any citizen in speaking out against a rogue law enforcement officer.”\textsuperscript{285} But these dissenters also viewed the interest at stake as that of a First Amendment right exercised by a government employee. Therefore, Ceballos’s free speech interest had to be balanced against a set of powerful competing interests.\textsuperscript{286} These dissenters would have weighed Ceballos’s interest in his favor only if his comments addressed “official dishonesty, deliberately unconstitutional action, other serious wrongdoing, or threats to health and safety.”\textsuperscript{287}

But Justice Souter’s dissent went further. That opinion defined Ceballos’s official duties as a prosecutor in a way that


\textsuperscript{280} Tarkington, supra note 22, at 2178 & n.13 (responding to Rosenthal, supra note 22, at 56-57).

\textsuperscript{281} *Ceballos*, 547 U.S. at 425.

\textsuperscript{282} Id. at 425-46.

\textsuperscript{283} Id. at 442 & n.13 (Souter, J., dissenting) (citing *Anderson v. Liberty Lobby, Inc.*, 477 U.S. 242, 255 (1986)).

\textsuperscript{284} Id. at 426-27 (Stevens, J., dissenting).

\textsuperscript{285} Id. at 431 (Souter, J., dissenting).

\textsuperscript{286} Id. at 419-20 (majority opinion).

\textsuperscript{287} Id. at 435 (Souter, J., dissenting).
could entail something more than the First Amendment interest of “any citizen.” Ceballos’s job was “to enforce the law by constitutional action” and “to exercise the county government’s prosecutorial power by acting honestly, competently, and constitutionally.”

Justice Souter’s dissent also cited Ceballos’s allegation that Brady required him to disclose the memorandum to the defense as exculpatory evidence. Finally, this opinion noted, Ceballos’s “claim relating to [his own] truthful testimony in court must surely be analyzed independently [of the other aspects of his claim] to protect the integrity of the judicial process.” These statements marked a path toward relief for Ceballos on remand.

Justice Souter’s opinion offered no substantive analysis of the constitutional doctrine that allegedly compelled Ceballos to ensure, among other things, that the judge considering Cuskey’s motion to traverse heard “truthful testimony.” In the lone opinion viewing Ceballos’s due process disclosure duties as dispositive, Justice Breyer’s dissent added little doctrinal analysis to Justice Souter’s. He would have held that Ceballos had a protectable First Amendment interest because, as a general matter, Ceballos was engaged in “professional speech—the speech of a lawyer,” and because, as a prosecutor, he had more specific speech obligations imposed by the Brady doctrine. In Justice Breyer’s view, those twin circumstances vested a First Amendment interest in Ceballos’s production of the memorandum because they simultaneously increased the need to protect the speech at issue, reduced government interests in controlling the speech, and lowered the risk of inefficiencies and overreaching that might occur through judicial trenching on managerial authority in the government workplace.

---

288 Id. at 437.
289 Id. at 442.
290 Id. at 444. Justice Souter’s opinion stated that Ceballos’s testimony at the motion hearing “stopped short of his own conclusions” that the deputies lied. Id. at 442. This was so because the judge sustained the prosecution’s objections to that testimony. Joint App’x Vol. III, supra note 204, at 298-303.
291 The case settled shortly after the Supreme Court remanded the case. Ceballos v. Gareatti, No. CV 00-11106 AHM (AJWx), 2002 U.S. Dist. LEXIS 28039 (C.D. Cal. Jan. 30, 2002). Orly Lobel reported that, although the terms of the agreement are not public, “the settlement may have been very favorable for Ceballos.” Orly Lobel, Citizenship, Organizational Citizenship, and the Laws of Overlapping Obligations, 97 CALIF. L. REV. 433, 454 & n.125 (2009).
292 Gareatti, 547 U.S. at 446-47 (Breyer, J., dissenting).
293 Id.
With respect to the professional speech prong of Justice Breyer's test, Margaret Tarkington has supplied much of the doctrinal analysis missing from his cursory dissenting opinion.\textsuperscript{294} Tarkington also is one of the few scholars to discuss the facts and law that gave rise to Ceballos's disclosure duties. Her work points to the core piece of \textit{Brady} evidence in the case: Ceballos's two conversations with "the police affiant for the warrant" mentioned in the majority opinion.\textsuperscript{295}

Those conversations comprised Ceballos's confrontation of Detective Wall with evidence that the affidavit contained falsehoods as well as Wall's subsequent admission about needing to change the affidavit.\textsuperscript{296} That evidence had both exculpatory and impeachment value. First, it tended to show that the truck could not be traced to Cuskey's junkyard. Second, it cast doubt on the investigating officers' veracity. Because the evidence called key components of the state's case into question, Ceballos's investigation and Wall's admission constituted \textit{Brady} material, inaccessible to the defense unless the prosecution complied with the constitutional duty to turn it over.\textsuperscript{297} Unsurprisingly, Ceballos, his supervisors, the federal District Court judge, three Circuit Court judges, and Justice Breyer all agreed on this point.\textsuperscript{298} The puzzle remains: why was

\footnotesize
\textsuperscript{294} Tarkington, \textit{supra} note 22. For California authorities governing Ceballos's conduct, see \textit{People v. Davis}, 309 P.2d 1, 9 (Cal. 1957) (in reference to defense attorney speech, stating that "counsel may not offer testimony of a witness which he knows to be untrue" since "[t]o do so may constitute subornation of perjury"); \textit{CAL. BUS. \\& PROF. CODE} § 6068(d) (2011) (attorneys have duty to employ "those means only as are consistent with truth, and never to seek to mislead the judge or any judicial officer by an artifice or false statement of fact or law").

\textsuperscript{295} Tarkington, \textit{supra} note 22, at 2178 \\& n.13 ("Ceballos, a member of the prosecution, talked twice with the police affiant for the warrant" (citing \textit{Garcetti}, 547 U.S. at 414 (majority opinion))).

\textsuperscript{296} \textit{See supra} notes 220-28 and accompanying text.

\textsuperscript{297} Without accounting for the impeachment value of the officers' admissions or the opinions of Ceballos's supervisors, the lower court judges, and Justice Breyer that Ceballos had a constitutional disclosure duty, Professor Rosenthal reached the opposite conclusion. \textit{See Rosenthal, supra} note 22, at 56-57. An experienced prosecutor and admitted advocate for the "managerial prerogatives" vindicated in \textit{Garcetti}, \textit{id.} at 33, Rosenthal is alone among commentators in acknowledging the relevance of \textit{Franks v. Delaware} to the case. \textit{Id.} at 40 \\& n.11. He also acknowledged prosecutors' duties to obtain and disclose \textit{Brady} evidence held by law enforcement, \textit{id.} at 56 \\& n.72, and concluded that prosecutors must be protected against retaliation for good-faith compliance with \textit{Brady}-line disclosure duties. \textit{Id.} at 67-68. He might concede that, regardless of the outcome at the motion hearing, the impeachment value of Wall's admission supported a good-faith belief in the prosecutors' constitutional duty to disclose it, and agree to the improbability that the officers would have volunteered that evidence to the defense lawyers who had publicly accused them of misleading or lying to a magistrate.

\textsuperscript{298} On the correlation of duties and rights, see generally, \textit{Osborn v. Bank of United States}, 22 U.S. 738 (1824) ("It is not unusual for a legislative act to involve
To be sure, the plaintiff’s pleadings and argument did not point the Court toward Ceballos’s disclosure duty under Barton. But it is hardly satisfactory to blame the relatively thin summary judgment record for the almost complete silence on the core due process principles implicated by the alleged facts. The case was argued twice. The Court had the benefit of several amicus briefs. Nor does it seem likely that the Justices gave Ceballos’s due process duties short shrift because the trial judge, in denying the Franks motion, effectively ruled that he had no duty to reveal evidence of police perjury. Brady materiality is determined ex post by estimating the cumulative value that withheld evidence would have had if the defense had been able to use it. The rejection of a Brady claim does not reach back to eliminate a prosecutor’s duty to reveal potentially exculpatory evidence in time for the defense to incorporate it into the case investigation and litigation. To the contrary, Brady requires disclosure, even in close cases, to prevent prosecutors from “tacking too close to the wind” and infecting criminal adjudications with unfairness and unreliability. Ceballos’s commitment to protecting the integrity of the judicial process appears to have been in good faith; he was willing to sacrifice his career for it.

Nor does it solve the puzzle to insist that the case is really about government employee speech under the First Amendment and not about due process discovery duties. On that point, it is noteworthy that the Garcetti majority did not adopt the view of the District Court and the Court of Appeals’ concurring opinion that Ceballos had no protectable interest in his communication precisely because his speech was required

consequences which are not expressed. An officer, for example, is ordered to arrest an individual. It is not necessary, nor is it usual, to say that he shall not be punished for obeying this order. His security is implied in the order itself. . . . [T]he judicial power is the instrument employed by the government in administering this security.”); Brewer v. Hoxie Sch. Dist. No. 46, 238 F.2d 91, 100 (8th Cir. 1956) (granting injunction restraining anti-desegregationists’ interference with plaintiff school board directors’ attempts to comply with Brown v. Board of Education, 347 U.S. 483 (1954)); id. (“[T]he existence of a Constitutional duty also presupposes a correlative right in the person upon whom the duty is imposed to be free from direct interference with its performance.”).

299 At oral argument, plaintiff’s counsel urged that government employees have a First Amendment right to speak on any topic that is newsworthy. Transcript of Oral Argument at 48-50, Garcetti, 547 U.S. 410 (No. 04-473) [hereinafter 2005 Transcript].

300 But see Rosenthal, supra note 22, at 45.


302 See supra notes 256-58 and accompanying text.
by due process disclosure doctrine. That logic, like the Supreme Court majority’s, was syllogistic—but more sweeping.

On the reasoning of the lower courts, Ceballos had no First Amendment protection against retaliation for performing his job duties. Those duties included notifying his superiors of the Brady information in Cuskey’s case. Therefore, Ceballos had no First Amendment protection against retaliation for communicating Brady information to his superiors. This reading of Ceballos’s due process disclosure duties submerges them within First Amendment doctrine. In turn, constitutional protection against retaliation for prosecutors who comply in good faith with Brady-line obligations sinks out of sight.

But the Supreme Court majority did not ride that wake (at least not to its most extreme conclusion) despite being urged to do so at oral argument. Instead, the majority focused exclusively on the form, not the content, of Ceballos’s memorandum. For the majority, it was the act of writing such memoranda, more than the content, that erased Ceballos’s free speech rights. With respect to the content of the memorandum, the majority threw due process disclosure duties what turns out to be, on closer analysis, a slender lifeline. The majority cited Brady not as the basis for submerging due process protections under First Amendment doctrine, but as exemplifying constitutional “safeguards” and “imperatives” that “protect employees and provide checks on supervisors who would order unlawful or otherwise inappropriate actions.” Those lines trace back to exchanges during oral argument hinting that Ceballos had his own due process interest at stake.

One interpretation of the majority opinion is that prosecutors’ truth-telling and justice-seeking duties are so fundamental to the proper functioning of justice systems that they comprise bedrock due process components warranting protection independent of the First Amendment. The majority’s implicit retention of independent due process protections on the facts alleged indicates that, even under a First Amendment-dominant analysis, Ceballos’s federal constitutional obligation to speak might have changed the outcome in the case had the litigants focused closely on the core due process principles at issue and their precise applications to the facts of the case.

---

305 Id.
under *Franks* and *Barton*. As discussed above, it is possible that the failure to undertake this analysis resulted from inattention. It is also possible that, had Ceballos’s constitutional duties been fully fleshed out, a majority might nevertheless have subordinated those duties to internal management interests. The Court might have declined to identify a due process right to be free from retaliation for complying with *Brady* in the context of internal agency communication. The Court might have held that a federal appellate court decision such as *Barton* does not clearly establish a right or duty for purposes of a claim filed under 42 U.S.C. § 1983.

But assuming that the due process lifeline traced here has some strength, focusing on core meanings in the context of the specific constitutional rights and duties implicated in *Garcetti* might have pulled Ceballos’s claim safely ashore. The claim could have rested directly in due process or undergone the more complex move from due process through the First and Fourth Amendments and back again. In either case, the due process analysis would dominate. But certainly for Richard Ceballos, *Barton* reinforced the core *Brady* duties to disclose Detective Wall’s statement. The duty triggered a corresponding right to disclose the information without retaliation.307 Indeed, as even Ceballos’s supervisors acknowledged, the entire office shared the same disclosure duty.308 Any remaining adjudicative work would comprise fact-based determinations to be resolved on remand, including the credibility of defense denials that any retaliation occurred.

**IV. FULL OPEN FILE DISCOVERY AS A MODEL FOR REFORM**

**A. Opening the Black Box: Law, Politics, and Bureaucracy in the Regulation of Prosecutorial Decision Making**

The Supreme Court acknowledged more than a decade ago that open file discovery can “increase the efficiency and the fairness of the criminal process.”309 The same concerns animate *Brady*-line requirements for prosecutors and investigative

---

307 See supra note 298.

308 See supra note 255.

agencies to disclose information beneficial to the defense.\footnote{310}{See supra notes 1-5 and accompanying text.}

Providing defendants with information obtained through government’s superior investigative resources levels the playing field. Fact finders make better decisions when adversaries present their strongest admissible evidence in the most compelling manner. The finality of reliable verdicts increases public confidence in the transparency and accountability of adjudicatory systems. Finality and reliability also reduce the significant costs resulting from alleged and actual error in criminal cases—costs borne by defendants, crime victims and survivors, their families, and the taxpayers who support prosecutors, public defenders, courts, and prisons.\footnote{311}{For data on prison costs alone, see, for example, \textit{Justice Reinvestment: Facts and Trends}, \textit{Justice Reinvestment}, \url{http://justicereinvestment.org/facts_and_trends} (last visited Jan. 27, 2012).}

Implementation and expansion of full open file criminal discovery provides an effective model for vindicating the foregoing interests. The statutes are short and simple. Prosecutors must provide the complete investigative files, including any material obtained by law enforcement, to the defense before trial.\footnote{312}{N.C. GEN. STAT. § 15A-903 (2012).} A file includes investigators’ notes, all oral statements (which are required to be recorded), and any other information obtained during the investigation.\footnote{313}{\textit{Id.} § 15A-903(a)(1). Particularly as modified by the 2011 amendments, the unique combination of provisions in the North Carolina statutes qualify that state’s reform as “full” open file discovery. Mosteller, supra note 5, at 263.} Oral statements need not be signed or adopted by potential witnesses to fall under the discovery requirement.\footnote{314}{State v. Shannon, 642 S.E.2d 516, 523 (N.C. Ct. App.), review granted, 649 S.E.2d 893 (N.C.), petition withdrawn, 654 S.E.2d 246 (N.C. 2007).} Work product privileges are narrowed to “protect the prosecuting attorney’s mental processes while allowing the defendant access to factual information collected by the state.”\footnote{315}{\textit{Shannon}, 642 S.E.2d at 525 (interpreting N.C. GEN. STAT. § 15A-904 and quoting \textit{John Rubin}, N.C. INST. OF GOV’T, ADMINISTRATION OF JUSTICE 8 (2006)).} Reciprocal discovery of specified material by the defense to the prosecution is mandated.\footnote{316}{\textit{Id.} §§ 15A-905 to -906 (2012).} Ex parte motions to restrict disclosure are allowed where necessary to prevent any “substantial risk to any person” of harm, intimidation, or even “unnecessary annoyance or embarrassment.”\footnote{317}{\textit{Id.} § 15A-903(d).} Willful violation of the statute is punishable as a felony.\footnote{318}{Id. § 15A-903(d).}
The readily available empirical evidence—including case law, legislative history, and observations of some key stakeholders—demonstrates that full open file discovery reform can succeed. The history of full open file reform also illustrates that litigation and legislation are viable strategies for regulating prosecutors’ discretionary decision making. That evidence is significant because full open file reform can help prevent debacles like John Thompson’s convictions and incarceration, protect prosecutors like Richard Ceballos from retaliation for complying with disclosure duties, and increase the fairness, finality, and efficiency of criminal litigation.

Yet despite the number and magnitude of restrictions inhibiting enforcement of Brady-line duties; despite the documented link between disclosure violations and wrongful convictions; despite the Supreme Court’s implicit acknowledgement that resources are wasted in litigating pretrial discovery motions and postconviction Brady claims; and despite the unfairness and uncertainty that the Brady regime creates for defendants, victims, and their families, full open file discovery remains a rarity in the United States. Reform on this front lags even as years of dogged litigation and media attention spur cures—including passage of model uniform statutes—for the mistaken eyewitness identifications, false confessions, and flawed forensic analysis that also contribute to unfair, unreliable, and inefficient outcomes in criminal cases.

Full open file discovery reform also has received little scholarly attention. Commentators have generally followed the courts in deferring to prosecutors, and any internal administrative changes they may choose to make in governing their own discretionary decision making, as the optimal prophylaxis against improper nondisclosure and its attendant costs. In its strongest recent iteration, this trend expressly despair of litigation and legislation as effective avenues for criminal justice reform, at least with respect to the prosecutorial function. Miller and Wright’s Black Box is

319 See Mosteller, supra note 5, at 306-16.
320 Id.
321 See supra note 1.
323 See supra note 30.
324 Exceptions include Mosteller, supra note 5. Other scholars recommending open file reform include Alafair S. Burke, Revisiting Prosecutorial Disclosure, 84 IND. L.J. 481 (2009); Medwed, supra note 98, at 1557-66 (noting benefits and risks); and Richard Rosen, Reflections on Innocence, 2006 Wis. L. REV. 237, 271-73.
325 Miller & Wright, supra note 29.
exemplary. Miller and Wright argue that development of internal data collection protocols is the best, if not the only mechanism for regulating prosecutorial discretion.326 They cite Orleans Parish District Attorney Harry Connick, Sr. as an example of such internal reform.327

The citation carries a Niebuhrian irony.328 On one hand, the Black Box embodies the opacity of discretionary decision making by the most powerful players in criminal justice systems.329 Miller and Wright cogently argue that data collection, assessment, and reporting can open this Black Box and bring greater transparency and accountability to charging decisions, including the identification and correction of any intentional or unconscious racial bias.330 But bureaucratic reform is no panacea. For example, Daniel Richman cautions that Connick’s charge-screening reform was, on closer analysis, “more like a case study in institutional disarray.”331 And Pamela Metzger reveals that, even pre-Katrina, Orleans Parish defendants were jailed without counsel for weeks under local rules and policies while Connick’s office decided whether and how to charge them.332

326 Id. at 128-30.
327 Id. (citing Ronald Wright & Marc Miller, The Screening/Bargaining Tradeoff, 55 STAN. L. REV. 29, 30-36 (2002)); id. at 186-87 (anticipating opportunities for increasing “external’ transparency” through internal agency data collection and assessment).
328 R EINHOLD NEIBUHR, T HE IRONY OF AMERICAN HISTORY (1958); Wright & Miller, supra note 29, at 59 (“New Orleans would probably not appear at the top of most lists of progressive criminal justice systems.”).
329 Morrison v. Olson, 487 U.S. 654, 727 (1988) (Scalia, J., dissenting) (“Only someone who has worked in the field of law enforcement can fully appreciate the vast power and the immense discretion that are placed in the hands of a prosecutor with respect to the objects of his investigation.”); Erik J. Luna & Marianne Wade, Prosecutorial Power: A Transnational Symposium: Prosecutors as Judges, 67 WASHT. & LEE L. REV. 1413, 1415 (2010) (“In many (if not most) American jurisdictions, the prosecutor is the criminal justice system. For all intents and purposes, he makes the law, enforces it against particular individuals, and adjudicates their guilt and resulting sentences.”); Stephanos Bibas, Prosecutorial Regulation Versus Prosecutorial Accountability, 157 U. PA. L. REV. 959, 960 (2009) (“No government official in America has as much unreviewable power and discretion as the prosecutor.”); William J. Stuntz, The Pathological Politics of Criminal Law, 100 MICH. L. REV. 505, 509 (2001) (“Law enforcers, not the law, determine who goes to prison and for how long.”); see generally DAVIS, supra note 109.
330 Miller & Wright, supra note 29, at 155, 193-95; see also Janet Moore, Causes, Consequences, and Cures of Racial and Ethnic Disproportionality in Conviction and Incarceration Rates, 3 FREEDOM CTR. J. 35 (2011) (introducing and moderating panel discussion with Wayne McKenzie, founding Director of the Prosecution & Racial Justice Program at the Vera Institute of Justice).
Moreover, it was during Connick’s nearly thirty-year tenure as District Attorney of Orleans Parish that John Thompson became one of several prisoners, including other death row inmates, whose cases were tainted by Brady violations. Until Thompson’s case was decided, Kyles v. Whitley was the leading national exemplar of Brady violations. This was due in part to the sheer scope of the “blatant and repeated” due process discovery violations by prosecutors in Connick’s office. More recently, the Court held that the latest Brady violation from Orleans Parish required a new trial on quintuple-murder charges for a due process violation that at least one Justice considered so egregious as to warrant asking the prosecutor during oral argument whether she had considered forfeiting the case.

To be clear, Miller and Wright offer no imprimatur for Connick’s leadership beyond the commitment to data collection for charge-screening purposes. They do argue, however, that Orleans Parish exemplifies, with respect to the exercise of prosecutorial discretion, how “internal regulation can deliver even more than advocates of external regulation could hope to achieve.” Legislators and judges, they explain, “have never answered the calls” for reform, “and the political dynamics of American criminal justice make it very unlikely that they will do so in the future.”

At first glance, the Supreme Court’s constriction of 42 U.S.C. § 1983 as a mechanism for enforcing prosecutorial disclosure duties in Connick v. Thompson and Garcetti v. Ceballos would seem to support this thesis. Nor are Miller and Wright alone in promoting a “physician, heal thyself” strategy for reform of discretionary prosecutorial decision making. Stephanos Bibas agrees that interbranch regulation and professional disciplinary oversight comprise a generally “ineffectual” check on prosecutorial power and joins those who

---

334 Id. at 455 (Stevens, J., concurring); see supra notes 98-100 and accompanying text.
335 See supra note 24.
336 Miller & Wright, supra note 29, at 129-33.
337 Id. at 128-29.
offer managerial blueprints for internal agency reform.\textsuperscript{339} A recent gathering of scholars and practitioners on criminal discovery reform also focused on prosecutors’ internal policy development as the optimal solution for the improper nondisclosure of information beneficial to the defense.\textsuperscript{340} For example, a working group charged with assessing disclosure processes assumed that “prosecutors will disclose appropriate information to the defense” and left the issue of nondisclosure “as a matter for discussion by other Working Groups.”\textsuperscript{341} The “Systems and Culture” group could not even reach consensus on training prosecutors to disclose “favorable” as opposed to “material” evidence.\textsuperscript{342}

The reticence of courts and scholars vis-à-vis external regulation of prosecutorial disclosure duties is unsurprising. Federalism and separation of powers concerns inhibit external oversight of criminal justice systems. And as a general statement of realpolitik it is certainly true, as the late William Stuntz and many others have noted, that the disproportionate impact of crime and criminal justice processes on low-income people and people of color is simultaneously a cause and a consequence of these systems’ resistance to reform through the traditional avenues of litigation and legislation.\textsuperscript{343}

Deference to prosecutorial agencies with respect to the vindication of defense discovery rights is also consistent with the unprecedented concentration of unchecked power in the prosecutorial function.\textsuperscript{344} From \textit{Brady}’s inception onward,\textsuperscript{339} Bibas, \textit{supra} note 329, at 978.


\textsuperscript{342} Wright, \textit{supra} note 104, at 1998.


\textsuperscript{344} See \textit{supra} note 329.
The constitutional doctrine has prioritized deference to prosecutorial discretion over enforceability. Case after case underscores the weakness of Brady-line authorities as enforceable mechanisms for criminal discovery reform and the prevention of wrongful convictions. In this context, Connick and Garcetti comprise the most recent illustration of the doctrinal subordination of prosecutorial disclosure duties, the related due process and fair trial rights of defendants, and the broader interests of criminal justice stakeholders in convictions and sentences that are worthy of confidence.

B. The Birth of Reform and the Presumptuousness of Despair

Connick and Garcetti could be taken as more evidence (if any were needed) justifying despair of litigation and legislation as effective strategies for criminal justice reform, particularly in the context of prosecutors’ discovery duties. That despair might deepen in light of the stalled progress on federal legislation designed to restore some of the protections taken from government employees by Garcetti. The ironic death of that legislation by the secret vote of a single senator might seem to undercut the thesis that law and politics remain vibrant and quintessentially democratic avenues toward criminal justice reform.

But unpacking the Connick-Garcetti interplay should encourage reformers, despite losses on the litigation and legislation fronts, to recover subordinated core meanings of relevant rights and duties and to seek new avenues for their vindication. I view the cases as holding this potential, perhaps because capital litigation taught me that despair is nearly always presumptuous. Certainly Connick and Garcetti illustrate Brady’s complexity and weak enforceability. But the cases also highlight the contrasting fairness, finality, and efficiency that can be obtained through full open file discovery.


346 Devine, supra note 345.

347 Cf. JULES LOBEL, SUCCESS WITHOUT VICTORY: LOST LEGAL BATTLES AND THE LONG ROAD TO JUSTICE IN AMERICA 1-10 (2003).
reform. Moreover, it was precisely the type of dogged investigation and litigation that led to the Connick and Garcetti rulings—and discontent inspired by similar examples of the Brady model’s failures—that motivated full open file reform. In this broader view, the Connick-Garcetti duo may open a new chapter in the reform story. To appreciate that possibility, one must consider the irony surrounding the advent of full open file discovery statutes.

Full open file discovery had its genesis in the “tough on crime” movement of the late 1990s and the specific desire to speed the pace of capital postconviction litigation and executions.\footnote{Mosteller, supra note 5, at 262-64.} Like states across the country, North Carolina was modeling new legislation on the federal Antiterrorism and Effective Death Penalty Act.\footnote{Email communications with Staples Hughes, N.C. Appellate Defender; Malcolm “Tye” Hunter, Dir., Ctr. for Death Penalty Litig., Durham, N.C.; and Robert Mosteller (on file with the author). I served as an Assistant Appellate Defender under the leadership of Hunter and Hughes, respectively, between 1998 and 2005.} During legislative hearings on the issue in 1996, members of the defense bar cited Robert McDowell’s capital case as an example of the increased efficiency that full open file discovery could bring to capital litigation.\footnote{Id.} McDowell, a black man, was convicted of capital murder by a North Carolina jury in 1979.\footnote{Id.} After his appeal was denied,\footnote{State v. McDowell, 271 S.E.2d 286, 289 (N.C. 1981).} his lawyers learned that the prosecution had failed to disclose evidence at trial that indicated, among other things, that the assailant was white.\footnote{Id.} The dogged investigation and litigation of McDowell’s Brady claim included multiple rounds of petitions for review to the United States Supreme Court.\footnote{McDowell v. Dixon, 858 F.2d 945, 947 (4th Cir. 1988).} In 1990, after eleven years of court proceedings,\footnote{Dixon v. McDowell, 489 U.S. 1033 (1989); McDowell v. North Carolina, 451 U.S. 1012 (1981); McDowell v. North Carolina, 450 U.S. 1025 (1981).} McDowell

\begin{itemize}
\item \footnote{Mosteller, supra note 5, at 262-64.}
\item \footnote{Email communications with Staples Hughes, N.C. Appellate Defender; Malcolm “Tye” Hunter, Dir., Ctr. for Death Penalty Litig., Durham, N.C.; and Robert Mosteller (on file with the author). I served as an Assistant Appellate Defender under the leadership of Hunter and Hughes, respectively, between 1998 and 2005.}
\item \footnote{Id.}
\item \footnote{State v. McDowell, 271 S.E.2d 286, 289 (N.C. 1981).}
\item \footnote{Id.}
\item \footnote{McDowell v. Dixon, 858 F.2d 945, 947 (4th Cir. 1988).}
\item \footnote{Upon discovering the Brady information, postconviction counsel filed a motion seeking a new trial. After a hearing, the trial judge granted the motion. State v. McDowell, 310 S.E.2d 301, 303-05 (N.C. 1984). The state Supreme Court reversed. Id. at 303. On federal habeas, the United States Court of Appeals for the Fourth Circuit ordered relief on that claim. Dixon, 858 F.2d at 951. Although McDowell’s case appears to be missing from a 2006 assessment of Fourth Circuit habeas decisions, he is among a small minority of petitioners to prevail in a Circuit whose rates of habeas relief are “significantly lower than in any other circuit.” Sheri Lynn Johnson, Wishing Petitioners to Death: Factual Misrepresentations in Fourth Circuit Capital Cases, 91 CORNELL L. REV. 1105, 1150-51 & nn.373-79 (2006).}
\end{itemize}
received a new trial and a conviction on the lesser, noncapital charge of second-degree murder.356

Thus, full open file discovery reform began as a political compromise inspired by protracted litigation over due process disclosure duties. A coalition (or at least a collection) of odd bedfellows designed full open file reform in part to reduce the wasted resources and uncertainty created in cases like Robert McDowell’s by nondisclosure of Brady evidence to the defense. The new open file statute “changed the landscape” of capital postconviction litigation.357 As discussed below, the state courts rejected the few challenges that prosecutors raised in the early phases of implementation. Instead, the legislature addressed prosecutors’ concerns by amending the statutes. And the general trajectory of the amendments has been toward expanding and strengthening open file rights and duties.

C. Full Open File Reform in Action: Case Law, Legislative History, and a View from the Trenches

The state Supreme Court issued the first significant interpretation of the 1996 full open file discovery statute within two years of enactment. The state had sought to restrict the law’s scope by labeling the bulk of the prosecution’s file “work product.”358 The Court applied the law’s plain language and rejected the state’s argument.359 Then compliance with the 1996 statute revealed Brady violations in a series of capital cases. The state courts vacated judgments in ten of those cases.360 That litigation focused attention on noncompliance with Brady discovery duties and the resulting burdens on courts, victims’

357 Mosteller, supra note 5, at 263.
358 Id. at 263 & n.17 (citing State v. Bates, 497 S.E.2d 276, 280 (N.C. 1998)).
359 Bates, 497 S.E.2d at 280-82.
360 Mosteller, supra note 5, at 261 & n.11 (citing cases). Brady claims led to new trials in these cases only after significant expenditure of time and resources on direct appeal, during which the office of the state Attorney General represented the prosecution. See, e.g., State v. Gell, 524 S.E.2d 276 (N.C. 2000) (new trial awarded on post-conviction Motion for Appropriate Relief, Bertie County Superior Court Docket No. 95 CRS 1884, Dec. 9, 2002); State v. Hamilton, 519 S.E.2d 514 (N.C. 1999) (same, Richmond County Superior Court Docket No. 95 CRS 1670, Apr. 23, 2003); State v. Hoffman, 505 S.E.2d 80 (N.C. 1998) (same, Union County Superior Court Docket No. 95 CRS 15695, Apr. 30, 2004); State v. Bishop, 472 S.E.2d 842 (N.C. 1996) (same, Guilford County Superior Court Docket Nos. 93 CRS 20410-20423, Jan. 10, 2000); State v. Womble, 473 S.E.2d 291 (N.C. 1996) (same, Columbus County Superior Court Docket Nos. 93 CRS 1992-1993).
families, taxpayers, defendants, and public confidence in the criminal justice system.

By this point, about half of the state prosecutors had implemented some form of open file discovery. Subsequent media coverage and public pressure led to the first major expansion of the 1996 full open file statute in 2004. That amendment created the broadest criminal discovery rights and duties in the nation. As discussed below, subsequent amendments in 2007 and 2011 continued the trajectory toward increasingly expansive and enforceable discovery duties.

The 2004 statute extended discovery duties beyond the postconviction phase, requiring the prosecution in all felony cases to provide the defense, before trial, with “the complete files of all law enforcement and prosecutorial agencies involved in the investigation . . . or prosecution” of the case. File was defined expansively as well, to include all statements by defendants, codefendants, and witnesses, all investigative notes, all test or examination results, and “any other matter or evidence obtained during the investigation” of the case. Crucially, the statute mandated that “[o]ral statements shall be in written or recorded form” and shielded prosecutors’ interview notes from disclosure as work product only “to the extent they contain the opinions, theories, strategies or conclusions” of the prosecutor or other legal staff. Another provision allowed both prosecution and defense to move ex parte for protective orders to prevent disclosures that create “a substantial risk to any person of physical harm, intimidation, bribery, economic reprisals, or unnecessary annoyance or embarrassment.”

In April 2007, the state’s intermediate court of appeals interpreted the crucial provision requiring recordation and disclosure of all oral statements. As was the case with the 1996

362 Mosteller, supra note 5, at 264-76. Discontent over the perceived mildness of sanctions for nondisclosure also led to expanded duties under the governing disciplinary rules. Id.
363 Id. at 274-76.
365 Id.
366 Id.
367 Id. § 15A-904(a).
368 Id. § 15A-908(a). The new provision for ex parte protective orders required notice to opposing counsel of the order's existence, but “without disclosure of the subject matter.” Id. A subsequent amendment requires any “affidavits or statements” supporting such an ex parte order to be sealed for appellate review. N.C. GEN. STAT. § 15A-908(b) (2012).
discovery statute, the Court again held that the legislature meant what it said: “The plain, unambiguous meaning of this requirement is that ‘statements’ need not be signed or adopted before being subject to discovery.” The panel majority also rejected the state’s objection that the requirement to reveal all recorded witness statements trenched too far upon work product privileges. The Court held that the amendment adequately protects prosecutors’ “mental processes while allowing the defendants access to factual information collected by the state.”

Like the state courts, the legislature also rejected an attempt by prosecutors to expand the work-product privilege in the discovery reform statutes. But prosecutors did obtain amendments addressing other concerns. At the prosecutors’ request, protections were added to shield personal identifying information and, more specifically, to protect identities of confidential informants. The legislature also eased requirements for recording witness statements to prevent redundancies.

Other amendments responded to a DNA laboratory’s failure to reveal exculpatory information in the Duke lacrosse

---

370 Id. at 525 n.2 (quoting John Rubin, N.C. Inst. of Gov’t, Administration of Justice, BULL. 2004/06, at 8); State v. Hardy, 235 S.E.2d 828, 841 (N.C. 1977) (“Only roughly and broadly speaking can a statement of a witness that is reduced verbatim to a writing or a recording by an attorney be considered work product, if at all . . . . Such a statement is not work product in the same sense that an attorney’s impressions, opinions, and conclusions or his legal theories and strategies are work product.”).
372 § 15A-904(a)(1)-(2), 2007 N.C. Sess. Laws 377 § 2. These protections were later expanded to include specific categories of confidential informants and victim impact statements. N.C. GEN. STAT. § 15A-904(a3)-(a4) (2012).
373 § 15A-903(a)(1), 2007 N.C. Sess. Laws 377 § 1. The amendment eliminated the recordation requirement when prosecutors interview a witness without a third party present unless the witness says something “significantly new or different” from prior recorded statements. Id. The Shannon dissent had noted the inefficiencies resulting from the previous double-recordation requirement. 642 S.E.2d at 524, 526 (McCullough, J., dissenting in part). After obtaining agreement on the 2007 amendments, the state dropped its appeal of the Shannon majority’s decision. State v. Shannon, 654 S.E.2d 246 (N.C. 2007).
cases and to new concerns raised by both prosecutors and defense lawyers. Prosecutors protested that unwarranted disciplinary charges were being filed against them when other agencies hindered compliance by failing to give prosecutors the information they needed to satisfy their discovery duties. In response, the legislature redefined the term *prosecutorial agencies* to accommodate prosecutors’ limited ability to sanction nondisclosure by investigators. But the amendments also expanded the statutes’ scope. *Agencies* was redefined to include “any public or private entity that obtains information on behalf of a law enforcement agency or prosecutor’s office in connection with the investigation... or prosecution” of a case. The legislature also clarified that all testing data must be disclosed, “including, but not limited to, preliminary tests or screening results and bench notes.”

In addition, these recent amendments tightened the mandate for disclosure by law enforcement or other investigative agencies of the complete investigative file to the prosecution. The legislature added criminal penalties for willful noncompliance with disclosure duties by investigative agencies. Finally, the revised statutes require that a prosecutor’s “reasonably diligent inquiry” to comply with disclosure duties triggers a presumption that he or she acted in good faith. The Governor signed these amendments into law on June 23, 2011.

The well-known Duke lacrosse case demonstrates that full open file discovery is a prerequisite—a necessary, if not a sufficient, condition—for preventing *Brady* violations as well as the wasteful litigation and wrongful prosecutions and

---

374 See Mosteller, supra note 5, at 285-306.
379 N.C. GEN. STAT. § 15A-903(d).
380 Id. § 15A-910(c)-(d).
381 H.B. 408 Bill Tracking Report, supra note 375.
Additional evidence of full open file discovery’s success includes the rarity of court challenges or other significant litigation arising from the statutes in the reported cases; the judiciary’s plain-language resolution of those claims; and the legislature’s accommodation of prosecutors’ concerns as they arise while maintaining and strengthening full open file rights and duties.

Observations from those involved with training and implementation indicate that after initial resistance, primarily from some of the more experienced prosecutors and law enforcement officers, full open file reform is finding increasingly broad acceptance. Under the statutes, responsibility for identifying materially beneficial information is where it belongs—with defense counsel. Full open file discovery appears to be increasing the speed and fairness of plea bargaining. An initial surge in pretrial protective orders subsided with amendments shielding victims and witnesses from inappropriate exposure or interference. Some jurisdictions have seen increased pretrial discovery hearings, mainly in high-level cases, as the parties document good-faith efforts to comply with their statutory duties. Issues regarding appropriate sanctions for discovery violations by both sides are being resolved through the normal course of trial rulings and appellate review.

On the other hand, logistical problems with discovery production have been significant, particularly in major cases with deep investigative histories. Some of these concerns are being addressed through training on best practices and through development of an electronic compliance program. This Discovery Automated System, when fully operational, will be used by investigative agencies, prosecutors, and defenders, respectively, to record, organize, and receive information. Every action in a case will be recorded, and date-stamped. Such

---

382 Mosteller, supra note 5, at 276-309.
383 Telephone Interviews with Brad Bannon and Professors Jim Drennan, Jeff Welty, and Jessica Smith, supra note 375; Telephone Interview with Kimberly Overton, supra note 361.
384 See sources cited supra note 383.
385 See sources cited supra note 383.
386 See sources cited supra note 383.
388 See sources cited supra note 383.
recording will include any redactions or edits made to any of the information contained in the file.  

Certainly there remain key sticking points for successful implementation of full open file reform. Thousands of new and experienced investigators must be trained. These front-line justice system personnel range from law enforcement officers to employees in the state Department of Social Services. Nor will it suffice merely to train these key stakeholders on the scope and meaning of full open file rights and duties. Investigators also must have adequate resources to comply with recording and reporting requirements. Finally, as Daniel Medwed, R. Michael Cassidy, and other scholars have noted, full open file discovery is not a cure-all. The recent removal of Durham, North Carolina District Attorney Tracy Cline from office was due in part to discovery violations that underscore the recalcitrance of some agency cultures, and of corresponding enforceability problems, related to prosecutors’ discovery duties. Nevertheless, when compliance can be enforced through criminal penalties, and when robust opportunities for defense investigation and litigation at the trial, appellate, and postconviction stages create a meaningful opportunity to prevent or promptly detect and correct discovery violations, the full open file model is the best option for improving efficiency, reliability, and fairness in criminal adjudications.

D. Research and Reform Opportunities for the Future

Although much work remains to be done to improve compliance with criminal discovery obligations, the readily available empirical evidence highlights significant steps toward successful implementation of full open file discovery as a reform model. The history of full open file reform also demonstrates the

---

389 See sources cited supra note 383.
390 Cassidy, supra note 37, at 1477; Medwed, supra note 98, at 1557-66.
continued vitality of litigation and legislation as necessary complements to internal agency reform in the context of discretionary prosecutorial decision making. Additional research should test for any influence of full open file discovery reform on plea and conviction rates and pretrial or postconviction litigation over discovery issues. Further research should also identify the conditions that foster such reform in some jurisdictions and impede it in others. With respect to North Carolina’s expansion of full open file discovery from capital postconviction cases to all felonies, Professor Mosteller summarizes the accomplishment in half-a-dozen words: “The adversaries hammered out a deal.”

That statement raises interesting questions. What conditions enabled the defense function, whose efficacy is often strongly discounted, to come to the negotiating table as a meaningful adversary to the prosecution? How might those conditions be duplicated in other jurisdictions? To what extent did the adversaries channel the interests of the disproportionately low income and minority individuals who bear the brunt of crime and criminal proceedings, but who, according to popular wisdom, are excluded from meaningful participation in the development of criminal justice policy? What conditions enable those same individuals to ask their own policy questions, build their own coalitions and advocate for their own solutions to what are too often perceived as “criminal injustice systems”?

Answering such questions could help move justice systems more quickly toward greater fairness and finality in criminal adjudications. With respect to the most recent set of full open file amendments, key negotiators for the prosecution and the defense concur that two skills were essential to successful resolution of the issues. First, the negotiators were able to defuse emotions inherent in the highly adversarial prosecutor-defender relationship. Second, they were able to listen closely to opposing views to detect and address the core concerns being brought to the table. Combining those

392 Mosteller, supra note 5, at 272 & n.76.
393 See generally CONSTITUTION PROJECT, supra note 113.
394 See supra note 343 and accompanying text.
396 Telephone Interviews with Brad Bannon and Professors Jim Drennan, Jeff Welty, and Jessica Smith, supra note 375; Telephone Interview with Kimberly Overton, supra note 361.
397 Id.
capacities allowed negotiators to tailor solutions to the specific problems at issue, and to build trust essential to the resolution of future problems. As I have discussed in a different doctrinal context, such capacities are central to the exercise of moral imagination within a discourse model of political ethics. Ideally, they will be brought to bear at the national level, through exploration of the full open file statutes as a model for a uniform criminal discovery code encompassing misdemeanors as well as felonies.

In the absence of such conditions, other “carrot-and-stick” options might be considered. The “stick” component could emulate federal legislation adopted after the Rampart police misconduct scandal exploded in Los Angeles. In response, Congress enacted 42 U.S.C. § 14141, which opened the door for Department of Justice oversight of local police reform efforts initiated through litigation. Ironically, it was in the context of public outrage over the Rampart scandal that Richard Ceballos felt bound to report evidence of officer perjury in the Cuskey and Longoria cases. Similar outrage over injustices and inefficiencies caused by Brady’s weak enforceability might lead to comparable legislation, oversight, and reform in the arena of prosecutorial disclosure duties. Reform advocates also might seek to tie the receipt of federal and other grant funding to the achievement of benchmarks in progress toward full open file reform. The “carrot” approach to reform has some track record of success in the context of policing, and might translate well in bringing appropriate levels of transparency and accountability to discretionary prosecutorial decision making.

---

398 Id.
400 See supra note 30.
402 See supra notes 225-27 and accompanying text.
403 MALCOLM M. FEELEY & AUSTIN D. SARAT, THE POLICY DILEMMA: FEDERAL CRIME POLICY AND THE LAW ENFORCEMENT ASSISTANCE ADMINISTRATION, 1968-78, at 4 (1980) (noting that legislation that created LEAA adopted a block grant approach, in which fighting crime would remain a state and local function, and federal government’s primary role would be to provide revenues and ideas allowing states to develop programs for their own use); NANCY E. MARION, A HISTORY OF FEDERAL CRIME CONTROL INITIATIVES, 1960-93, at 56-58 (1994) (noting LEAA’s block grant design was supported by Republicans and Southern Democrats in Congress, who felt that federal government should not involve itself directly in local police efforts); Paul Hoffman, The Feds, Lies, and Videotape: The Need for an Effective Federal Role in Controlling Police Abuse in Urban America, 66 S. CAL. L. REV. 1453, 1530-31 nn.296-97 (1993) (urging conditional grants of federal funds to curb police misconduct).
CONCLUSION

The implementation and expansion of full open file discovery provides two object lessons. First, the model’s successes to date offer lessons for jurisdictions seeking greater fairness, finality, and efficiency in criminal case outcomes. Second, full open file reform demonstrates the continued vitality of law and politics as effective and quintessentially democratic tools for opening the black boxes that nest throughout the nation’s criminal justice systems. Connick and Garcetti can introduce an important new chapter in this reform process. The cases simultaneously underscore the weak enforceability of prosecutors’ due process disclosure duties and highlight the benefits of the full open file discovery model.

The core question animating this work is a search for sustainable production of the conditions that allow jurisdictions to pursue, through the traditional clash of law and politics as necessary complements to internal agency reform, significant “smart on crime” improvements such as full open file discovery. For example, Miller and Wright correctly stress the critical role of internal data gathering and assessment in identifying and correcting race effects on prosecutorial decision making. Yet it was zealous litigation and aggressive, grassroots-to-grasstops political advocacy that led to enactment of the nation’s first Racial Justice Act, allowing death row inmates to challenge death sentences based on statistical evidence of unconscious racial bias in charging, sentencing, and jury selection.

Litigation and policy advocacy also has motivated key indigent defense reforms, including the creation of politically independent oversight bodies with the authority to promulgate and enforce standards for attorney qualification, training, and

404 Miller & Wright, supra note 29, at 195.
performance. In a like vein, crises in state and local budgets have led legislatures to accompany new criminal justice statutes with fiscal note and racial impact assessment requirements. Such benchmarks can help check the massive and virtually unregulated investment of increasingly scarce tax dollars in what too often becomes a rapidly-spinning set of revolving doors into and out of local, state, and national criminal justice systems.

The complementary nature of bureaucracy, law, and politics warrants more scholarly analysis at these and other pivot points in the discretionary decision making that drives criminal justice systems. Legal scholars can enrich the analysis through interdisciplinary cooperation with specialists in criminology, sociology, political science, public health, and social work. Inter-institutional partnerships between the legal academy and government, nonprofit agencies, and foundations may yield more effective, sustainable system improvements.

Such engaged scholarship may also offer an antidote for despair over the possibility of truly democratic criminal justice reform. This is the political black box that can tempt scholars to privilege internal administrative reform over law and politics in the quest for criminal justice reform. Engaged scholarship may be able to empower the disproportionately low-income and minority members of our communities—those whose lives most often intersect with criminal justice systems, but who seldom have an effective voice in shaping policy and procedure—to pose their own research questions, formulate their own reform proposals, and create their own policy advocacy coalitions. Their voices are crucial to sustaining the quest for greater transparency and accountability in criminal justice decision making.


408 See generally Alexander, supra note 343.
The Class Differential in Privacy Law

Michele Estrin Gilman†

INTRODUCTION

Many Americans are willing to divulge personal information and even sacrifice civil liberties for the benefits of a wired world. They will turn over their spending habits for the convenience of shopping on-line, submit to body scanners and suitcase searches to travel by air, and tolerate Facebook selling their personal information to third parties in order to network with friends.1 These sorts of surveillance bargains are rarely struck by the poor. Low-income Americans travel more often by bus than plane, they lack money to shop at Amazon.com, and they are less likely to have a computer that makes social networking possible in the first place.2 This digital and economic divide does not mean, however, that the poor are insulated from privacy intrusions. On the contrary, they endure a barrage of

† Professor of Law and Director, Civil Advocacy Clinic, University of Baltimore School of Law. B.A. 1990, Duke University; J.D. 1993, University of Michigan Law School. For their feedback on this article, I would like to thank Matthew Fraidin, Leigh Goodmark, Dan Hatcher, Rebecca Hulse, Margaret Johnson, Rob Rubinson and the faculty participants at the AALS Women Rethinking Equality panel on Gender and Economic Equality, and the Poverty and Economic Mobility symposium at the Washington College of Law. The University of Baltimore School of Law provided support for this article.

1 See Gary T. Marx, Soft Surveillance: The Growth of Mandatory Volunteerism in Collecting Personal Information—“Hey Buddy Can You Spare a DNA,” 10 Lex Electronica 2 (Winter 2006), http://www.lex-electronica.org/articles/v10-3/marx.pdf (listing various bargains individuals strike between privacy and surveillance). “Although people acknowledge the importance of privacy, most value other things even more.” Jeff Sovern, Opting in, Opting out or No Options at All: The Fight for Control of Personal Information, 74 Wash. L. Rev. 1033, 1058 (1999). For instance, while many people will claim to oppose privacy intrusions, they will willingly turn over Social Security numbers and other personal information to telemarketers. Andrew Askland, What, Me Worry? The Multi-Front Assault on Privacy, 25 St. Louis U. Pub. L. Rev. 33, 34 (2006); see also James P. Nehf, Recognizing the Societal Value in Information Privacy, 78 Wash. L. Rev. 1, 4 (2003) (people support information privacy, but “understand that [their] interests in privacy must be balanced against other interests”).

2 See Lindsay Greer, Questioning Digital Citizenship: The Answer to Economic and Political Inequality?, 11 N.Y.U. J. LEGIS. & PUB. POLY 651, 655-57 (2008) (reviewing Karen Mossberger et al., Digital Citizenship: The Internet, Society, and Participation (2008)) (explaining that historically disadvantaged groups, including low-income Americans, have less access to the Internet and lower rates of use).
information-collection practices that are far more invasive and degrading than those experienced by their wealthier neighbors. The law reinforces this class differential in privacy.

Consider the case of Rocio Sanchez.\(^3\) In June 2000, Ms. Sanchez, after separating from her husband, applied for welfare benefits and food stamps at a San Diego County welfare office to support her infant daughter.\(^4\) One month later, an investigator from the Public Assistance Fraud Division of the San Diego District Attorney’s Office made an unannounced visit to her home pursuant to a county policy called Project 100%, which required home visits of all welfare applicants who were not suspected of fraud or ineligibility.\(^5\) The investigator asked Ms. Sanchez a series of questions about her husband and his whereabouts, when she had last talked with or seen him, and the reasons for their separation.\(^6\) He then searched the home, including her bedroom closet, and left to question her neighbors.\(^7\)

Ms. Sanchez encountered the investigator a few days later when he arrived at her former residence searching for her husband.\(^8\) She was there alone, cleaning the residence so that she could recover the rental security deposit.\(^9\) In her presence, the investigator proceeded to search the bathroom cabinets, the bedroom, and the dresser drawers—all of which were empty.\(^10\) Again, he questioned Ms. Sanchez about her husband, including asking why she was still speaking to her sister-in-law if she was in fact separated.\(^11\) He demanded that she pull out papers from her husband’s trash can that might lend clues to his location, remarking that it was “funny” that she had never filed a domestic violence complaint.\(^12\) Two months later, the county approved her application for benefits.\(^13\) Nevertheless, Ms. Sanchez was upset by these interrogations, particularly the accusatory tone taken by the investigator,\(^14\) and she became

\(^3\) These facts are taken from the First Amended Complaint in Sanchez v. County of San Diego. First Amended Complaint for Declaratory & Injunctive Relief, Sanchez v. Cnty. of San Diego, No. 00-CV-1467JM(JFS), 2003 WL 25655642 (S.D. Cal. Mar. 10, 2003), aff’d, 464 F.3d 916 (9th Cir. 2006).
\(^4\) Id. ¶ 12.
\(^5\) Id. ¶¶ 12-13.
\(^6\) Id. ¶ 14.
\(^7\) Id.
\(^8\) Id. ¶ 15.
\(^9\) Id.
\(^10\) Id.
\(^11\) Id. ¶ 16.
\(^12\) Id.
\(^13\) Id. ¶ 17.
\(^14\) Id. ¶ 16.
a plaintiff in a class action lawsuit challenging the constitutionality of Project 100%.

Ultimately, in *Sanchez v. San Diego*, the Ninth Circuit relied on *Wyman v. James*, a Supreme Court opinion from 1973, and upheld the home visits against a Fourth Amendment challenge, reasoning that “a person’s relationship with the state can reduce that person’s expectation of privacy, even within the sanctity of the home.” In a bitter dissent from a denial of a petition for rehearing en banc, seven Ninth Circuit judges called the case “nothing less than an attack on the poor.” As the dissenters stated, most government benefits do not flow to the poor, “yet this is the group we require to sacrifice their dignity and their right to privacy.” By contrast, “[t]he government does not search through the closets and medicine cabinets of farmers receiving subsidies. They do not dig through the laundry baskets and garbage pails of real estate developers or radio broadcasters.” As the dissenters concluded, “This situation is shameful.”

Welfare administration is highly devolved in that states and localities have great discretion in how they structure their welfare programs. So in another jurisdiction, Ms. Sanchez might have been subjected to drug tests or finger imaging or unsolicited family-planning advice, such as pressure to implant a Norplant birth control device. Throughout the country, poor women such as Ms. Sanchez face constant surveillance as they must comply with extreme verification requirements to establish eligibility for welfare benefits, travel to scattered offices to procure needed approvals, reappear in person at welfare offices at regular intervals to prove their ongoing eligibility and answer intrusive questions about their child

15 *Sanchez v. Cnty. of San Diego*, 464 F.3d 916, 927 (9th Cir. 2006).
16 *Sanchez v. Cnty. of San Diego*, 483 F.3d 965, 969 (9th Cir. 2007) (Pregerson, J., dissenting from the denial of rehearing en banc).
17 *Id.*
18 *Id.*
19 *Id.*
rearing and intimate relationships. Further, under federal law, all jurisdictions engage in extensive data sharing with other government and private offices to ferret out fraudulent public assistance applications.

The poor are subjected to privacy intrusions not only by governments but also by private parties, such as employers. Although a stated goal of welfare is to move welfare recipients into the workforce, even if Ms. Sanchez obtained a low-wage job, surveillance of her life and personal choices would likely continue. In the white collar workforce, employers regularly monitor e-mail, Internet, and phone communications of their employees, raising the specter that information could “fall[] into the wrong hands or . . . [be] used for a purpose we did not envision when we disclosed it.” Nevertheless, most white collar workplace monitoring is invisible and easy to ignore, which may in part explain the lack of public outrage or legislative protections against workplace surveillance. By contrast, low-wage workers are concentrated in service industries. They are more subject to visible—sometimes humiliating—surveillance tactics such as psychological testing, regular drug screening, and overt videotape monitoring.

Criminal justice scholars have described a class differential in privacy under the Fourth Amendment, which protects reasonable expectations of privacy from warrantless government searches and seizures. People who live in crowded, urban neighborhoods and who cannot afford “a freestanding home, fences, [and] lawns,” have a lowered expectation of privacy and are thus more likely to suffer

---

23 Id. at 672-73.
25 JOEL F. HANDLER & YEHESKEL HASENFELD, BLAME WELFARE, IGNORE POVERTY AND INEQUALITY 191-99 (2007) (describing how TANF welfare departments have adapted to the work-first mandate).
26 See Nehf, supra note 1, at 27.
29 See infra notes 219-24 and accompanying text.
30 U.S. CONST. amend. IV.
31 See, e.g., Christopher Slobogin, The Poverty Exception to the Fourth Amendment, 55 Fla. L. Rev. 391, 401-05 (2003).
warrantless searches by government agents. These realities of geography also mean that poor families are far more likely to become entangled in intrusive child welfare and domestic violence investigations. In addition, the homeless have virtually no privacy at all; courts have held that their meager shelters are not entitled to protection from government searches. At the same time, the federal government mandates computerized tracking of the homeless, requiring them to divulge personal data when they seek social services. Meanwhile, more Americans with financial means are moving into gated and private communities, thus buying themselves privacy. As the experience of Ms. Sanchez reveals, this class differential extends beyond the criminal justice context into every corner of daily life.

Low-income Americans are a diverse group living individualized lives; they are “indescribably varied and multifaceted.” Despite this diversity, they share the reality of

32 Id. at 401.
34 See, e.g., D’Aguanno v. Gallagher, 50 F.3d 877, 880 (11th Cir. 1995) (people who lived in a homeless campsite on private property did not have a reasonable interest in privacy); United States v. Ruckman, 806 F.2d 1471, 1473 (10th Cir. 1986) (no reasonable expectation of privacy in a cave located on public lands); Amezquita v. Colon, 518 F.2d 8 (1st Cir. 1975) (squatters on public land had no reasonable expectation to privacy); People v. Thomas, 45 Cal. Rptr. 2d 610, 613 (Cal. Ct. App. 1995) (no reasonable expectation of privacy in a cardboard box used as a residence); Whiting v. State, 885 A.2d 785, 799-801 (Md. 2005) (squatter’s expectation of privacy was not reasonable); Commonwealth v. Cameron, 561 A.2d 783, 786 (Pa. Super. Ct. 1989) (squatter in abandoned rowhouse did not have a reasonable expectation of privacy). Homeless persons have had better success with Fourth Amendment challenges to the searches of their personal property. Pottinger v. City of Miami, 810 F. Supp. 1551, 1570-71 (S.D. Fla. 1992) (homeless persons had reasonable expectation of privacy in their personal belongings); State v. Mooney, 588 A.2d 145 (Conn. 1991) (defendant had a reasonable expectation of privacy in a duffel bag and box located in the area under a highway bridge where he was living; court did not rule on whether he had a reasonable privacy interest in his “home” itself). The law is contradictory with regard to homeless shelters. See Steven R. Morrison, The Fourth Amendment’s Applicability to Homeless Shelters, 32 HAMLIN L. REV. 319, 333 (2009).
35 See infra notes 52-63 and accompanying text.
36 See George C. Chapman Phillips, Boundaries of Exclusion, 72 Mo. L. REV. 1287, 1302-03 (2007) (“In 2001 there were 7,058,427 households living year-round in gated communities in the United States” and the law upholds these residents’ expectations of privacy and rights to exclude others).
37 JOHN GILLIOM, OVERSEERS OF THE POOR: SURVEILLANCE, RESISTANCE, AND THE LIMITS OF PRIVACY 20-21 (2001) (describing demographic, political, physical, and regional variations among the poor). On the importance of seeing and understanding the poor as individuals with their own narratives, see Frank Munger, Identity as a Weapon in the Moral Politics of Work and Poverty, in LABORING BELOW THE LINE 1, 20
experiencing privacy quite differently on a daily basis than do middle and upper class Americans. Given that at least 15 percent of the population—or one out of seven Americans—currently lives below the poverty line, this differential demands attention and discussion. To be sure, sophisticated surveillance technology combined with limited legal protections should be a serious concern for all Americans. Still, for most Americans, these privacy intrusions are felt as a vague sense of unease over being watched along with concerns about identity theft (an old-fashioned crime with a new twist). By contrast, low-income Americans suffer tangible harms from government and corporate surveillance that go beyond discomfort. The privacy intrusions they face are often overt, and the harms are concrete.

The American privacy law regime has developed to reflect middle-class concerns, and as such is focused on preventing the misuse of information after it is collected. By contrast, low-income people tend to face stigmatization at the time information is gathered. The poor interact with the government and low-wage employers in ways that are ongoing and interpersonal, and, as a result, a “right to be left alone” does not protect their interests in dignity and autonomy. This article argues that poor Americans experience privacy differently than those with economic resources and that the law reinforces this differential. This differential has costs not only for the poor, but for all citizens. Our privacy laws and policies should reflect equality norms to ensure that poor Americans are not unfairly subjected to humiliating surveillance tactics.

The class differential in privacy law results from complex interactions between class, race, and gender. Because poor Americans are disproportionately minority and female, it is impossible to talk about class without taking into account how subordination is linked to race and gender. Critical legal theorists have shown that anti-poverty public policies are built on racist, gendered notions about the people who need relief. In

(Frank Munger ed., 2002). Munger writes, “If we understand [poor persons] not as sinners or saints but as constituents of the mainstream like ourselves, we will be willing to allow them the same latitude to fail or succeed that we grant insiders within our own communities.” Id. at 15.

38 See Sabrina Tavernise, Poverty Rate Soars to Highest Level Since 1993, N.Y. TIMES, Sept. 14, 2011, at A1 (stating that the Census Bureau reported a 15.1 percent poverty rate for 2010).

39 This foundational concept of privacy was articulated by Samuel D. Warren & Louis D. Brandeis in The Right to Privacy, 4 HARV. L. REV. 193, 195 (1890); see also Olmstead v. United States, 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting).

40 See JOHN ICELAND, POVERTY IN AMERICA 81, 88 (2d ed. 2006).
turn, these public policies impact all poor people—even those who are not minorities or female. The result is that the poor as a group suffer extreme privacy violations, which in turn pose a barrier to self-sufficiency and democratic participation.

Part I of the article describes how low-income Americans experience privacy in the welfare system and low-wage workforce. Women moving along the welfare-to-work continuum face a bevy of humiliating surveillance tactics, and, as Part I explains, the physical and mental health consequences of unfair surveillance undermine welfare’s statutory goal of self-sufficiency. Part II of the article analyzes how the law privileges middle-class privacy interests but fails to protect the poor from privacy intrusions. As a constitutional matter, courts often hold that the poor do not have reasonable expectations of privacy entitled to protection. As a statutory matter, laws primarily protect against the misuse of data, rather than its collection, and thus reflect middle-class privacy concerns. As a common law matter, the law focuses on reputational injuries; it has not expanded to prohibit demeaning surveillance practices that target the poor. Accordingly, Part III explores robust conceptions of privacy that focus on values safeguarded by privacy, such as dignity, respect, and trust. This part views privacy as a means rather than an end. It concludes that class equality needs to be a central concern of privacy law and suggests this can be achieved, in part, by ensuring that data collection practices are fair and undertaken with respect for their subjects.

I. SURVEILLANCE OF THE POOR

Privacy is an amorphous concept, despite the considerable efforts of theorists to pin it down. Scholars have variously described privacy as the right to be let alone; the

---

41 See infra text accompanying notes 300-03 (discussing Dorothy Brown’s theory that racist ideas underpinning Earned Income Tax Credit (EITC) enforcement impact poor white Americans as well); see also ICELAND, supra note 40, at 38-39 (discussing research showing a misperception that that most poor people are black and how that perception has political consequences); MARTIN GILENS, WHY AMERICANS HATE WELFARE: RACE, MEDIA, AND THE POLITICS OF ANTIPOVERTY POLICY 3-4, 60-80, 154-73, 204-06 (1999) (explaining how negative media coverage of poor black Americans has perpetuated stereotypes of the black poor as undeserving and generated public opposition to welfare).

42 From the perspective of welfare recipients, economic self-sufficiency has multiple dimensions, including “the exercise of personal power and freedom.” Elizabeth A. Gowdy & Sue Pearlmutt, Economic Self-Sufficiency: It’s Not Just Money, 8 AFFILIA 368, 383 (Winter 1993). As discussed infra, surveillance undermines autonomy.

43 Warren & Brandeis, supra note 39, at 193.
ability to control personal information and access to the self; a precondition to intimate relationships and the development of community; and an essential component of human dignity, autonomy, personhood, and self-determination. Clearly, privacy is connected with multiple values, and thus a lack of privacy can impact the self as well as interpersonal relationships. For the poor, surveillance by the government and employers implicates multiple values. Part A describes the nature of privacy invasions suffered by the poor in the areas of welfare receipt and the low-wage workplace. These focus areas are not comprehensive, but they illustrate how large societal institutions encroach upon the lives, homes, and bodies of poor individuals. Part B then examines emerging psychological research to assess the harms caused by privacy invasions. These harms amount to far more than hurt feelings; the poor suffer physical and mental health injuries associated with extreme stress.

A. Privacy in Welfare and Low-Wage Employment

Poor people regularly experience privacy deprivations related to their personal information, bodies and homes, and decision making. This surveillance serves to control and limit the autonomy of poor people, and has strong historical roots. “Politically, the purposes of surveying the poor have largely stayed constant for three centuries: containment of alleged social contagion, evaluation of moral suitability for inclusion in public life and its benefits, and suppression of working people’s resistance and collective power.” Technology has enhanced the ability of the state and private corporations to achieve these ends. Although participation in government entitlement programs and particular employment is technically voluntary, the reality of life and the marketplace is such that the poor

44 ALAN WESTIN, PRIVACY AND FREEDOM 7 (1967).
48 The three broad categories of privacy interests are information, spatial, and decisional. See Jerry Kang, Information Privacy in Cyberspace Transactions, 50 STAN. L. REV. 1193, 1202-03 (1998).
49 Virginia Eubanks, Technologies of Citizenship: Surveillance and Political Learning in the Welfare System, in SURVEILLANCE AND SECURITY: TECHNOLOGICAL POLITICS AND POWER IN EVERYDAY LIFE 89, 90 (T. Monahan ed., 2006) (emphasis omitted); see also Gilliom, supra note 37, at 19 (“From today’s computerized information systems . . . and back to the surveying and badging of the poor in sixteenth-century Europe, governments have closely examined those who seek assistance.”).
have little choice but to subject themselves to these privacy invasions. Notably, the surveillance imposed on the poor is usually overt; indeed, part of the purpose is to create stigma. By contrast, other Americans are increasingly subject to “soft” surveillance, which involves less invasive techniques, hidden technologies, and implied consent. These methods of soft surveillance raise deep concerns about civil liberties and lack of choice, but they are profoundly different in character and effect than the unceaseful, coercive surveillance tactics used with low-income populations.

1. Welfare

As discussed in the Introduction, Ms. Sanchez, a welfare applicant, faced informational, spatial, and decisional privacy intrusions by the state. She had to provide welfare officials with detailed information about her family, submit to home searches, and justify the state of her marriage. Her experience is emblematic. A typical applicant for welfare, called Temporary Assistance for Needy Families (TANF), must undergo a multistage, multiday application process consisting of screening interviews, application interviews, group orientations, and employability assessments. She must answer questions

---

50 See Marx, supra note 1, at 1-3. For instance, Marx describes police asking community members to provide voluntary DNA samples by mouth swab. Id. at 1. Marx also describes other “disingenuous communication that seeks to create the impression that one is volunteering when that isn’t the case,” such as building signs stating that one agrees to a search upon entering the premises and consumer reward programs that rely on giving up personal information. Id. at 2.

51 Id. at 5-6. These new tactics are driven by “the mass media’s encouragement of fear and perceptions of crises,” “the seductiveness of consumption,” and “the development of inexpensive, less invasive broad searching tools.” Id. at 6 (footnote omitted).

52 For the similarly intrusive questioning poor pregnant women are subjected to as a condition of receiving Medicaid benefits, see infra text accompanying notes 358-61, discussing Khiara M. Bridges, Privacy Rights and Public Families, 34 HARV. J.L. & GENDER 113, 114-16 (2011).


Today, a person who wants to apply for public assistance in L.A. County must visit an Eligibility Office. In these prisonlike structures, visitors pass through
ranging from her resources and sustenance needs to her psychological well-being.54 Her own word is not enough; she must also provide independent verification of her answers to many of these questions, either through her own documentation or through information gathered from third parties,55 and in some cases, caseworkers conduct investigations themselves.56 As part of TANF, an applicant must also comply with child support enforcement efforts by providing detailed paternity information about her children.57

Once a welfare mother turns over personal data, this information is electronically shared and compared with numerous federal and state databases, as well as commercial databases, to verify eligibility and to ferret out duplicate or otherwise fraudulent applicants.58 At any time, law enforcement officials can demand that welfare and housing officials turn over personal information about benefits recipients, even when the recipient is not herself suspected of any crime.59 By contrast, state officials cannot conduct similar fishing expeditions into the bank accounts of those individuals with the means to maintain savings.60

Many jurisdictions distribute benefits electronically, which allows them to monitor when and how funds are spent.61 While affluent Americans submit to market-research surveillance designed to cater to their purchasing preferences, the electronic systems that monitor the poor “facilitate the invasive scrutiny of their purchases and discipline of their

---

metal detectors and past armed security guards on their way to the clerk who is cloistered behind a Plexiglas window. There they must wait for hours in a crowded waiting room before being seen by an Eligibility Worker.


54 See STUDY OF THE TANF APPLICATION PROCESS, supra note 53, at 3-7.
55 See MEYERS & LURIE, supra note 53, at 27; HOLCOMB ET AL., supra note 53, at 3-16.
56 See Mulzer, supra note 22, at 676.
60 Id. at 669.
behavior.”62 As a result, these systems limit “clients’ autonomy, opportunity, and mobility: their ability to meet their needs in their own ways.”63

Physical privacy is not only invaded by home searches, but public benefits recipients may also be fingerprinted and photographed, usually through biometric imaging.64 Moreover, as part of child support enforcement, TANF recipients must agree to DNA testing for themselves and their children if paternity is contested.65

Further, TANF permits states to invade the decisional privacy of welfare mothers in order to control their behavior to align with middle-class norms.66 While the law governing decisional privacy is outside the scope of this article (it involves reproductive rights and family autonomy), it overlaps with data collection because the state transmits its behavioral expectations to poor mothers during the application and certification stages of public benefits programs.67 The most controversial of these sexual regulation policies is the imposition of family caps.68 Typically, family caps provide no cash benefit increases for any children conceived while the mother is on welfare.69 Several jurisdictions have also pushed “Norplant” bonuses, which cover the cost of

62 Torin Monahan, Surveillance and Inequality, 5 SURVEILLANCE & SOC’Y 217 (2008).
63 Eubanks, supra note 49, at 90-91.
64 See Nina Bernstein, Experts Doubt New York Plan to Fingerprint for Medicaid, N.Y. TIMES, Aug. 30, 2000, at B1 (listing states that fingerprint welfare recipients); see also HOLCOMB ET AL., supra note 53, at 3-1 to 3-25 (explaining that Dallas, TX and New York, NY use fingerprinting and photographing to screen public benefit recipients); Murray, supra note 21, at 39-40.
66 See HANDLER & HASENFELD, supra note 25, at 282-315 (describing and critiquing the “family values” provisions of TANF).
68 See Anna Marie Smith, The Sexual Regulation Dimension of Contemporary Welfare Law: A Fifty State Overview, 8 MICH. J. GENDER & L. 121, 173-77 (2002). Slightly less than half the states have adopted a family cap. Id. at 174.
69 See Rebekah J. Smith, Family Caps in Welfare Reform: Their Coercive Effects and Damaging Consequences, 29 HARV. J.L. & GENDER 151, 165-67 (2006). In states with the family cap, about nine percent of the caseload has been impacted by the family cap policies, resulting in about twenty percent less in cash assistance per family. Id. at 170-71.
implanted, long-term contraceptive devices for welfare mothers, sometimes with an additional cash award.\textsuperscript{70} In addition, many states bestow upon welfare mothers unsolicited family-planning advice in the form of counseling sessions, family-planning classes, pamphlets, and encouragement to give their children up for adoption.\textsuperscript{71} In short, poor women seeking public benefits face limitations across multiple dimensions of privacy.

2. Low-Wage Workplace

Low-wage workers—currently one-third of the workforce—are workers whose wages are so low that full-time work does not push them over the poverty line.\textsuperscript{72} Immigrants, single mothers, and African-Americans are disproportionately represented among low-wage workers.\textsuperscript{73} They are concentrated in the service sector, often working in retail, as custodians, as care providers for children and the disabled, and as security guards.\textsuperscript{74} The service sector uses the widest range of surveillance techniques. With modern technology, employers can log computer keystrokes, listen to telephone calls, review e-mails and Internet usage, conduct drug tests, employ mystery shoppers, watch closed-circuit television, track employee movements through GPS or radio frequency devices, and require psychometric tests.\textsuperscript{75} Yet the uses of the data generated by these tactics “are not made clear to employees, policies outlining their use are not in place, and information practices are not subject to any third-party audits or checks.”\textsuperscript{76}

While there are ample studies and recommendations about employer surveillance in the white collar workforce, the low-wage workforce remains mostly ignored by privacy scholars. Due to the lack of study, much evidence about surveillance practices in the low-wage workforce is anecdotal. In the best-selling book \textit{Nickled and Dimed}, author Barbara Ehrenreich went undercover in a series of low-wage jobs such

\textsuperscript{70} See Smith, supra note 69, at 168-69; see also Bridgewater, supra note 21, at 404-05 (arguing that the state's coercive use of Norplant to hinder the reproductive rights of African-American women violates the Thirteenth Amendment).

\textsuperscript{71} See Smith, supra note 68, at 169, 177-81.


\textsuperscript{74} LAWRENCE MISHEL ET AL., THE STATE OF WORKING AMERICA 329 (2009).


\textsuperscript{76} Id. at 91.
as a diner waitress, nursing home attendant, cleaning woman, and Wal-Mart salesperson.77 Throughout the book, she explains how each job was physically and mentally demanding as well as financially draining, given the costs of housing, food, and lengthy commutes.

The jobs were also privacy-stripping. Of all the workforce indignities, Ehrenreich was most surprised and offended at the loss of self-respect engendered in low-wage jobs.78 For instance, as a waitress, Ehrenreich “was warned that my purse could be searched by management at any time.”79 She remarks that “[d]rug testing is another routine indignity. . . . In some testing protocols, the employee has to strip to her underwear and pee into a cup in the presence of an aide or technician.”80 Ehrenreich also found pre-employment personality tests demeaning; they include “questions about your ‘moods of self-pity,’ whether you are a loner or believe you are usually misunderstood.”81 Reflecting on these intrusions, she states, “It is unsettling, at the very least, to give a stranger access to things, like your self-doubts and your urine, that are otherwise shared only in medical or therapeutic situations.”82

Ehrenreich’s experience appears widespread in the low-wage workforce. For instance, a study of Latina nannies found that they escaped the isolation of their jobs by congregating with other nannies in public parks.83 Thus, when their employers came to the park to check on their children, the nannies felt that it signaled a lack of trust.84 Whereas the nannies wanted autonomy to do their jobs, these visits transformed the park from “the nannies’ space into another site of surveillance” similar to the employers’ homes.85 For these workers, a public space was a private refuge, which demonstrates the context-specific nature of privacy and how it promotes values of dignity and autonomy. In other words, domestic workers sometimes have too much “bad” privacy (they work isolated in someone else’s home) and not enough “good” privacy (ability to work and live

77 BARBARA EHRENREICH, NICKLED AND DIMED: ON (NOT) GETTING BY IN AMERICA (2001).
78 Id. at 208.
79 Id.
80 Id. at 209.
81 Id.
82 Id.
84 Id. at 289.
85 Id. at 290.
with some measure of respect and independence). For domestic workers, privacy can be hard to come by, even as they are expected to respect the privacy of their employers by discreetly staying out of sight. Moreover, their working conditions are negotiated in the privacy of the employer’s home, where the bargaining differential is pronounced and where workers are susceptible to abuse.

Migrant farmworkers also face an alarming lack of privacy, as they live in employer-provided housing. David Shipler describes a North Carolina barracks used to house migrant workers during sweet potato season, in which up to fourteen men live in a room that measures twelve by fifteen feet with no opportunity for solitude.

These sorts of class differentials are found throughout the low-wage workforce. For most white collar workers, drug testing is limited to situations that implicate public health or safety. By contrast, drug testing is regularly part of pre-employment screening for low-wage jobs. Routine drug testing and location tracking via satellite positioning devices are also most prevalent in jobs with the lowest status and salaries. Likewise, “[t]he majority of employees being electronically monitored are women in low-paying clerical positions.”

Similarly, pre-employment psychological tests—commonly called

---

86 See Joy M. Zarembka, America’s Dirty Work: Migrant Maids and Modern Day Slavery, in GLOBAL WOMAN: NANNIES, MAIDS, AND SEX WORKERS IN THE NEW ECONOMY 142, 142 (Barbara Ehrenreich & Arlie Russell Hochschild eds., 2002) (describing the isolation and dangers faced by domestic workers). This is true for home health care attendants as well, who are supposed to be “invisible” to foster the independence of those for whom they care. See Lynn Mae Rivas, Invisible Labors: Caring for the Independent Person, in GLOBAL WOMAN, supra, at 70, 72-74.

87 See MARY ROMERO, MAID IN THE U.S.A. 147 (2002). Speaking about live-in maids, Romero writes: “The combination of not having a bedroom and not having access to the rest of the house for resting or leisure activity continually affirms the worker’s inferior status in the employer’s home.” Id.

88 Id. at 129.


90 Id. at 98 (“its configuration could have had no purpose other than to house workers—and to deprive them of their dignity”).


92 Id. at 66.

93 KENNETH D. TUNNELL, PESSING ON DEMAND: WORKPLACE DRUG TESTING AND THE RISE OF THE DETOX INDUSTRY 24 (2004) (“The body of evidence clearly shows that social class and income are inversely related to drug testing; working-class members with the lowest incomes are those most likely to be subjected to drug testing.”).

honesty tests—are concentrated in the “retail, fast food, banking, and other service industries,” even though “their accuracy and predictive value are doubted.”

Many low-wage employers use multiple methods to control their workers. For instance, a study of workers in the fast food and grocery industries found extensive forms of surveillance, ranging from rows upon rows of hanging video cameras to drug tests to honesty tests, in which applicants were “asked about illegal behavior, such as whether they have taken drugs or stolen anything in the past, or about the behavior of their friends and acquaintances, such as whether they know anyone who takes drugs or steals,” as well as questions about how they would react in various scenarios.

At the same time, just because “people give up information in exchange for jobs and other valued outcomes should not be construed as meaning that doing so is voluntary.” Rather, in today’s workplace, an employer has almost no restraints on the forms and extent of surveillance it chooses to adopt, and in today’s tight labor market, employees have little choice but to submit.

B. Harms and Justifications

Poor individuals, families, and communities suffer tangible harms—psychological, material, and physical—as a result of the class differential in privacy policies and law. These harms are disproportionate to the justifications underlying privacy intrusions, and yet they remain invisible to many Americans. One privacy scholar has criticized other privacy theorists for failing to show how “privacy violations can negatively impact the lives of living, breathing human beings beyond simply provoking feelings of unease.” She argues that if we do not identify the harms of privacy invasions, privacy will continue to deteriorate. Part of the apathy may be that privacy theorists have neglected the poor. Indeed, one privacy

95 Sharona Hoffman, Preplacement Examinations and Job-Relatedness: How to Enhance Privacy and Diminish Discrimination in the Workplace, 49 U. KAN. L. REV. 517, 540 (2001). There are almost no legal protections against these tests. Id. at 541.
scholar admits that “privacy is seldom a matter of life and death;” rather, “[m]ost of the injuries caused by the misuse of data in modern society are not particularly embarrassing or emotionally disturbing.”99 This is not true for the poor.

For the poor, the injuries are concrete. Recent studies within the public health field show that when the state treats marginalized people with a lack of dignity, the results can include “loss of respect, loss of self worth, ego, sense of self, and soul, loss of status, social standing, and moral standing, loss of confidence and determination.”100 There are long-term effects as well, such as “social isolation or marginalization, a reluctance to seek help or access resources, passivity or ‘learned helplessness,’ a ‘small’ life of constrained choices, chronically poor physical and mental health, and a cycle of victimization and abuse, in which the violated individual turns to violating others.”101 These are psychological attributes that undermine the odds that poor families will become self-sufficient,102 which is the goal of the welfare system and, indeed, our liberal society.

1. Welfare

The welfare system of surveillance causes recipients to suffer psychological injuries including stress, fear, and feelings of degradation.103 While procedures such as drug tests and finger imaging may have instrumental purposes, they also send a degrading message to and about welfare recipients.104 This is because these procedures have symbolic meaning within our cultural traditions—“drug testing challenges traditions that urination is a private affair; and finger-imaging conjures up an image of criminality.”105 These procedures “freeze a moment in time while ignoring the ongoing context of inequality and structural violence.”106 Not only is the subject’s dignity degraded by these procedures, but society also receives a message that “reinforce[s] negative stereotypes of the

99 Nehf, supra note 1, at 30.
101 Id.
102 See generally Joaquina Palomar Lever et al., Poverty, Psychological Resources and Subjective Well-Being, 73 SOC. INDICATORS RES. 375 (2005) (describing how certain psychological attributes help people cope with poverty, while others are harmful).
103 See Gilliom, supra note 37, at 66-67, 78 (summarizing interviews with welfare recipients in Appalachia in the early 1990s).
104 Murray, supra note 21, at 40.
105 Id.
106 Campbell, supra note 91, at 72.
targets.”

In turn, these stereotypes drive punitive laws directed at the poor.

Not surprisingly, the privacy deprivations and humiliations associated with welfare discourage many needy women from seeking assistance. Without state assistance, these nonentrants to the TANF system and their children often lack adequate resources for food, shelter, and other basic needs—even if they are working. Studies have shown that nonentrants struggle to make ends meet by juggling a shifting array of nonpublic resources and that this hardship negatively impacts their health. In short, accepting welfare can subject one to humiliation, but refusing it can result in hunger. This “choice” hardly promotes autonomy or dignity.

Further, mandatory child support cooperation policies can result in the unintentional perpetuation of domestic violence. Battered women are overrepresented in the TANF population. To reduce the dangers of exacerbating domestic violence through reporting requirements, TANF attempts to protect victims by allowing states to grant these victims an exemption from the cooperation requirement. Yet many eligible women are not claiming the exemption for a variety of reasons, including the public setting of the welfare office, fear that child welfare authorities may take their children, stringent requirements for independent corroboration, and feelings of humiliation and embarrassment. As a result, the paternity

---

107 Murray, supra note 21, at 42.
108 See ROBERT MOFFITT ET AL., DEP’T OF HEALTH & HUMAN SERVS., A STUDY OF TANF NON-ENTRANTS: FINAL REPORT TO THE OFFICE OF THE ASSISTANT SECRETARY FOR PLANNING AND EVALUATION 2, 14 (2003) (new welfare reform policies discourage participation). “[Many] non-entrants in our study felt that applying for TANF was an unpleasant, time-consuming experience that resulted in little financial benefit . . . . Many felt the application process to be overly intrusive.” Id. at 20 (Part B).
109 Id. at 45. One study found that:

mothers jeopardized their own health and well-being when trying to provide for their families by taking on second, third, and fourth jobs, working odd hours, or commuting long distances via public transportation. . . . [Moreover, i]n order to acquire and maintain affordable housing, many families were forced to live in unsafe neighborhoods. . . . [And finally, m]others with young children consistently had trouble securing stable care for [their children.]

Id.

111 See Smith, supra note 68, at 153-54 (although batterers come from all social classes, TANF clients are especially vulnerable because they have fewer economic supports).
113 See Smith, supra note 68, at 165-66; Notar & Turetsky, supra note 110, at 672-76.
Disclosure required by the child support system poses a substantial risk to domestic violence victims for very little benefit. After all, these mothers do not receive any of the child support checks that are collected; rather, the state keeps the money to repay itself for the costs of welfare.114 Notably, TANF recipients lack the decisional autonomy of nonpoor single mothers, who are “not forced to identify, marry, live with, seek support from, or interact with the biological father.”115

Fraud prevention is the usual justification underlying welfare surveillance.116 For welfare mothers, this translates into home visits, fingerprinting, and elaborate third-party verification schemes. However, there is scant data on welfare fraud, even though states presume it is widespread.117 To be sure, there are applicants who do not report all sources of income. Because it is impossible to survive on welfare benefits (the average monthly benefit for a family of three is $363),118 some welfare applicants accept support from family members or earn additional income from jobs such as babysitting or cutting hair.119 Welfare mothers are in a bind—they must earn unreported income to provide for their children, but this conduct is considered “fraud.” Thus, in some cases, the state’s low welfare stipends and rigid earning limits create the very fraud that the state seeks to eliminate.

Further, studies show that women convicted of welfare fraud sometimes fail to report income for circumstances out of their control, such as when their partners hide their income or force them to keep it secret.120 In addition, many cases of fraud are unintentional and occur when welfare applicants either do not understand the complex income and resource reporting rules of

---

115 See Smith, supra note 68, at 140. Non-welfare families can use child support enforcement services, but they can withdraw on a voluntary basis. See Notar & Turetsky, supra note 110, at 671.
116 See Gustafson, supra note 59, at 658-61, 674-81, 683-88 (describing how fraud concerns have spurred the criminalization of welfare).
117 Murray, supra note 21, at 50 (“There is little systematic information on the form of welfare fraud known as double-dipping.”).
118 DEPT OF HEALTH & HUMAN SERVS., ADMINISTRATION ON CHILDREN AND FAMILIES: SEVENTH ANNUAL REPORT TO CONGRESS 75 (2006).
119 See Gilliom, supra note 37, at 67, 100; see also Kathryn Edin & Laura Lein, Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work 168 (1997).
120 Richelle S. Swan et al., The Untold Story of Welfare Fraud, 35 J. SOC. & SOC. WELFARE 133, 140 (2008).
TANF, or are misinformed by their caseworkers. The jobs obtained by welfare mothers tend to be unstable with fluctuating schedules and incomes, which can also lead to reporting problems when anticipated and actual income differ. “In short, the U.S. system both produces and punishes lawbreakers.” Better clarity in welfare rules and improved training of caseworkers could go far in fixing the sorts of errors that get mislabeled as intentional fraud.

Still, fraud is grossly overstated in welfare programs, making the draconian methods for rooting out fraud unreasonably invasive. For instance, New York City began fingerprinting welfare applicants in 1995 in an effort to root out imagined fraud. However, out of 148,000 recipients, the city found only forty-three cases of double dipping. Even purveyors of electronic fraud detection systems have admitted that fraud is extremely rare. Moreover, studies suggest welfare fraud is no more rampant in welfare than in other government programs, which are not subject to the same withering scrutiny. “The government takes a far greater risk on graduate student loans, for example, than on any welfare recipient.” Yet graduate students do not have their homes searched and are not fingerprinted.

Similarly, the annual cost of tax fraud, including underreporting and offshore tax shelters, is immense, but nevertheless is not considered “morally indecent.” Because the government can root out welfare double-dipping by computerized

---

121 Id. at 140, 143. In turn, convictions for fraud leave these women with serious collateral consequences, such as inability to pass screenings for housings, credit, or employment. Id.
122 Id. at 140.
123 Gustafson, supra note 59, at 681.
125 Id. at 140.
126 See, e.g., Joshua Dean, Texas Nears Rollout of Fingerprint System, FED. COMPUTER WEEK (Aug. 5, 1999), http://fcw.com/articles/1999/08/05/texas-nears-rollout-of-fingerprint-system.aspx (official from private contractor states that out of 700,000 people fingerprinted for public benefits, twelve cases were referred for further investigation).
127 See Mulzer, supra note 56, at 688-89 (“[F]ear of fraud has always played a larger role in the administration of public benefits programs than it realistically should have.”), Julilly Kohler-Hausman, “The Crime of Survival”: Fraud Prosecutions, Community Surveillance, and the Original “Welfare Queen,” 41 J. SOC. HIST. 329, 343 (2007) (“[M]uch of what became defined as fraud were simply attempts to supplement welfare grants with additional income from low wage work or living with another wage earner.”).
129 Donald Crump, Criminals Don’t Pay: Using Tax Fraud to Prohibit Organized Crime, 9 HOUS. BUS. & TAX L.J. 386, 397 (2009); see also Eric A. Posner, Law and Social Norms: The Case of Tax Compliance, 86 VA. L. REV 1781, 1783-84 (2000) (stating that “the audit rate is currently under 2%, and of those audited only a small fraction (4.1% in 1995) are penalized”).
matching of welfare applicants against social security numbers, more intrusive measures such as fingerprinting, photographing, and home visits serve only to stigmatize recipients. In addition, welfare surveillance has societal consequences, because it reduces democratic participation by welfare recipients. Obviously, a government program must ensure that the proper persons are receiving the appropriate levels of benefits. Further, welfare caseworkers cannot link welfare recipients to available social services without information about their needs. However, the level of information required from TANF applicants goes far beyond what is necessary to meet these goals and is often gathered through demeaning techniques.

2. The Low-Wage Workplace

The privacy losses suffered by low-income employees can cause humiliation, shame, and stigma. After working a series of low-wage jobs, Barbara Ehrenreich concluded: “If you are treated as an untrustworthy person—a potential slacker, drug addict, or thief—you may begin to feel less trustworthy yourself.” Psychological research confirms that “workplace humiliation can itself be as devastating as the physical or economic harms that are legally actionable in employment and other settings.” Moreover, “women, minorities, and some ‘outsider’ groups” suffer disproportionate levels of humiliation.

In the employment context, surveillance serves many purposes. Employers use monitoring to deter theft, protect proprietary information, guard against lawsuits, discourage improper conduct, and monitor work performance. These are legitimate objectives. At the same time, social scientists who study the workplace generally conclude that employer surveillance tactics are overly broad to accomplish these goals, with damaging effects on employees’ stress levels. For

---

130 Gustafson, supra note 59, at 677 n.153.
131 See infra notes 373-82 and accompanying text.
132 EHRENREICH, supra note 77, at 210.
134 Id.
instance, a prominent two-year study of telecommunications employees who worked in directory assistance, as service representatives, and in clerical jobs linked employee monitoring to headaches, backaches, wrist pains, greater fatigue, a 12 percent increase in depression, and a 15 percent increase in extreme anxiety.\(^{137}\) Other studies confirm that a variety of health problems can flow from employer surveillance “such as stress, high tension, headaches, extreme anxiety, depression, anger, severe fatigue, and musculoskeletal problems.”\(^{138}\) Moreover, the consequences of stress are magnified in jobs where employees lack control over their privacy or a voice in establishing monitoring procedures.\(^{139}\)

Employers also pay a cost. At the outset of the employment relationship, invasive application procedures can limit the pool of eligible applicants, as they perceive a lack of trust from potential employers.\(^{140}\) This distrust may disproportionately impact the disabled as well as racial and ethnic minorities who “may fear that they will be stigmatized unfairly by the information revealed by various selection procedures.”\(^{141}\)

The corporate bottom line can also suffer from decreased employee productivity and creativity, low morale, diminished trust, and high turnover caused by intrusive monitoring.\(^{142}\) Extreme surveillance can increase employee resistance.\(^{143}\) For instance, a study of call centers found that workers


\(^{139}\) See Stone-Romero et al., supra note 97, at 346. By contrast, “monitored participants who were given the opportunity to voice their opinions about the design and implementation of monitoring systems had higher perceptions of procedural justice.” Laurel A. McNall & Sylvia G. Roch, A Social Exchange Model of Employee Reactions to Electronic Performance Monitoring, 22 HUMAN PERFORMANCE 204, 205 (2009); see also A.F. Westin, Two Key Factors That Belong in a Macroergonomic Analysis of Electronic Monitoring: Employee Perceptions of Fairness and the Climate of Organizational Trust or Distrust, 23 APPLIED ERGONOMICS 35, 35-42 (1992). Likewise, monitoring can be beneficial when it results in productive feedback to employees. See David Holman, Phoning in Sick? An Overview of Employee Stress in Call Centres, 24 LEADERSHIP & ORG. DEV. J. 123, 128 (2003).

\(^{140}\) Stone-Romero et al., supra note 97, at 351, 364. A study found that employees consider the most invasive procedures to be lie detector tests, drug tests, medical examinations, background checks, and honesty tests. Id. at 363.

\(^{141}\) Id. at 364.

\(^{142}\) See D’Urso, supra note 138, at 287; Ball, supra note 75, at 93.

\(^{143}\) See Ball, supra note 75, at 94 (citing George Callahan & Paul Thompson, We Recruit Attitude: The Selection and Shaping of Routine Call Centre Labour, 39 J. MGMT. STUD. 233 (2002); Stephen Frankel et al., Beyond Bureaucracy? Work Organization in Call Centres, 9 INT’L J. HUM. RESOURCE MGMT. 957 (1998)).
circumvented surveillance by pretending to talk on the phone, leaving call lines open without a customer on the line, and misleading customers, among other tactics.\textsuperscript{144}

Studies also show that privacy intrusions can inhibit organizational citizenship, defined as “discretionary behavior that promotes effective organizational functioning but is not formally recognized by reward systems.”\textsuperscript{145} Surveillance can also lessen communication within an organization because there is less need for managers to interact with surveilled employees.\textsuperscript{146} In turn, less communication “may lower productivity, limit the development of important informal organizational networks, and prevent employees from exchanging key job-related information.”\textsuperscript{147}

By contrast, “information privacy [is] directly associated with psychological empowerment,” as well as “a greater willingness [on the part of employees] to engage in behaviors that help the organization.”\textsuperscript{148} For instance, Federal Express implemented a successful monitoring program of call center employees by including their input into setting work standards, promoting trust, and providing a comfortable work environment.\textsuperscript{149} In sum, both workers and employers pay hidden costs as a result of unfair and intrusive employee monitoring.

II. PRIVACY, POVERTY, AND THE LAW

This part surveys the legal system’s regulation of privacy for low-income people. It concludes that privacy law does not protect the poor. As a constitutional matter, the Fourth Amendment’s reasonableness standard provides scant protection because our society has long deemed it reasonable to intrude upon the lives of the poor. As a statutory matter, our laws focus on ensuring against the misuse of data, which is a middle-class priority, rather than data collection, which tends

\textsuperscript{144} Id. Obviously, these forms of resistance are not good for a corporate bottom-line. Yet resistance is an inevitable response to surveillance systems; as Gary Marx writes, “[S]urveillance targets often have space to maneuver and can use counter-technologies. . . . Humans are wonderfully inventive at finding ways to beat control systems and to avoid observation.” Gary T. Marx, \textit{A Tack in the Shoe: Neutralizing and Resisting the New Surveillance}, 59 J. SOC. ISSUES 369, 372 (2003).

\textsuperscript{145} Bradley J. Alge et al., \textit{Information Privacy in Organizations: Empowering Creative and Extrarole Performance}, 91 J. APPL. PSYCHOL. 221, 221, 223 (2006); see also Myria Watkins Allen et al., \textit{Workplace Surveillance and Managing Privacy Boundaries}, 21 MGMT. COMM. Q. 172, 192 (2007) (“High levels of surveillance can damage trust, leading to a less efficient workforce . . . and other costly consequences for organizations.”).

\textsuperscript{146} See Allen et al., \textit{supra} note 145, at 193.

\textsuperscript{147} Id.

\textsuperscript{148} Alge et al., \textit{supra} note 145, at 228.

\textsuperscript{149} Westin, \textit{supra} note 136, at 300.
to stigmatize the poor. As a common law matter, the law is geared toward elite concerns about reputation rather than the humiliation that surveillance causes to low-income Americans.

At bottom, the core principle of privacy law—the “right to be left alone”—ill-fits the needs of low-income Americans. Their vulnerable economic status leaves these citizens dependent on government assistance, which inevitably entails an ongoing relationship between the citizen and the state. Likewise, an employment relationship is necessarily continuing and interactive. Yet privacy is like an on/off switch; you either have it or you don’t. Privacy law does not account for intertwined relationships between citizens and larger institutional actors.

A. Constitutional Privacy Rights

The Fourth Amendment to the Constitution protects citizens from unreasonable searches by the state. In addition, the Supreme Court has recognized a constitutional right to information privacy, which protects certain personal information from government disclosure. However, these constitutional privacy protections are applied differently to the poor than to their wealthier counterparts. As this part explains, this differential arises from an ingrained bias against the poor.

1. Fourth Amendment

Fourth Amendment privacy hinges upon reasonableness. In assessing government searches, courts balance the individual’s reasonable expectation of privacy against the government’s rationale for the intrusion. This test is widely criticized as malleable and overly favorable to the government. Moreover, encroaching technology has put personal information in the public square. The resulting dilemma is that when everyday expectations of privacy diminish, it becomes less reasonable to expect the government to respect individual privacy.

150 U.S. CONST. amend. IV.
151 See infra text accompanying notes 199-207.
153 See, e.g., John D. Castiglione, Human Dignity Under the Fourth Amendment, 2008 WIS. L. REV. 655, 656 (calling the Supreme Court’s reasonableness standard “just about the most unhelpful guidepost one could have concocted”).
Nevertheless, the Supreme Court has held steadfast to the principle that people have a reasonable expectation of privacy in their home. Based on the deep-rooted Anglo-American maxim that “a man’s home is his castle,” the Court has drawn “a firm line at the entrance to the house,” stating that “physical entry of the home is the chief evil against which the wording of the Fourth Amendment is directed.” The Court’s “housing exceptionalism” is consistent with empirical research showing that people rate searches of bedrooms and home interiors as highly intrusive. The Court’s property-based conception of privacy therefore favors property owners. By contrast, the Supreme Court has held that a person who seeks government assistance gives up her rights to privacy, even in her home. The Court thus “define[s] privacy in a way that makes people who are less well-off more likely to experience warrantless, suspicionless government intrusions.”

a. Home Visits

The primary case demonstrating this discrepancy is the 1971 case of Wyman v. James, in which the Court upheld home visits by welfare officials, reasoning that the visits were not searches covered by the Fourth Amendment because they were consensual. Of course, one can question whether someone who is hungry and who would otherwise be homeless without public benefits can truly consent in a voluntary manner. Nevertheless, the Court applied a rational basis standard, ruling that even if the home visits were searches, they were reasonable given the state’s interest in deterring fraud, the

155 See Kyllo v. United States, 533 U.S. 27, 31 (2001) (“At the very core of the Fourth Amendment stands the right of a man to retreat into his own home and there be free from unreasonable governmental intrusion.”) (citations omitted)).
156 Jonathan L. Hafetz, “A Man’s Home Is His Castle?”, Reflections on the Home, the Family, and Privacy During the Late Nineteenth and Early Twentieth Centuries, 8 WM. & MARY J. WOMEN & L. 175, 180, 198-99 (2002).
162 Wyman v. James, 400 U.S. 309, 326 (1971); see also infra text accompanying notes 164-72.
163 Slobogin, supra note 161, at 400.
164 400 U.S. at 317-18.
need to protect the children of welfare mothers, the
rehabilitative purpose of the searches, and the lack of criminal
consequences that flowed from the searches. In finding that
the privacy deprivations were negligible, the Wyman Court
disregarded affidavits from twelve aid recipients alleging that
the unannounced visits were not only embarrassing when
guests were in the home, but also when personal questions
were asked in front of their children. In silencing the voices
of poor women, the Court ignored the social context in which
these women live and mistakenly equated forced consent with
free choice. Moreover, the Court’s disregard of their voices is
inconsistent with psychological and sociological research
showing that people value home privacy because it protects
interpersonal relationships.

The Wyman Court further intimated, based on Ms.
James’s social services case file (and not evidence adduced at
trial), that Ms. James’s son had been physically abused and
bitten by rats, concluding that “[t]he picture is a sad and
unhappy one.” The Court’s clear assumption was that poor,
single women are terrible mothers who warrant suspicion and
distrust. Throughout the opinion, the Court also expressed its
distaste for Ms. James and how she ran her household. The
Court disliked her “attitude,” “evasiveness,” and “belligerency”—
all of which arose from her resistance to the state and her
entirely reasonable belief that the state could verify her
eligibility through personal interviews and documents. Her
request was simply to be treated the same as other
beneficiaries of governmental largesse. As Justice Douglas
remarked in dissent, “No such sums are spent policing the
government subsidies granted to farmers, airlines, steamship
companies, and junk mail dealers, to name but a few.”
Because the poor are not a protected class under the
Fourteenth Amendment, equality law arguments based on
class will likely be insufficient. Legislation that discriminates

\[\text{Id.}\] pro
\[\text{Id.}\] at 318-24.
\[\text{Id.}\] at 320 n.8.
\[\text{Stern, supra note 159, at 940. Search activity in the home can “disrupt
domestic life, engender interpersonal conflict, reveal personal information that is
private to and constitutive of relationships, and chill socialization and intimacy.” Id.}\]
\[\text{Id.}\] 400 U.S. at 322 n.9.
\[\text{Id.}\]
\[\text{Id.}\]
\[\text{Id.}\] at 332 (Douglas, J., dissenting) (quoting J. Skelly Wright, Poverty,
Minorities, and Respect for Law, 425 DUKE L.J. 425, 437-38 (1970)).
\[\text{See Julie A. Nice, No Scrutiny Whatsoever: Deconstitutionalization of Poverty
Law, Dual Rules of Law & Dialogic Default, 55 FORDHAM URB. L.J. 629, 630 (2008).}\]
against or burdens the poor is reviewed under a lenient rational basis standard.\(^{173}\)

In 2006, in *Sanchez v. San Diego*, the Ninth Circuit reaffirmed the current validity of *Wyman*, when it ruled that Project 100\% (discussed in the Introduction) was constitutional.\(^{174}\) The court expressly lumped welfare mothers with criminals on probation to conclude that neither group has a reasonable expectation to privacy.\(^{175}\) In holding that *Wyman* was governing precedent, the *Sanchez* Court refused to recognize differences between the *Wyman* home visits and those of San Diego’s Project 100\%. Key to the *Wyman* holding was the Court’s view that the social worker visits at issue were “rehabilitative.”\(^{176}\) By contrast, Project 100\% is “expressly investigatory in nature, with no rehabilitative or service component,” and is carried out by law enforcement fraud investigators.\(^{177}\)

Moreover, the *Sanchez* Court disregarded thirty years of post-*Wyman* jurisprudence, which has significantly limited suspicionless, administrative searches.\(^{178}\) Under current law, a warrant and probable cause are not required for administrative searches that are driven by “special needs, beyond the normal need for law enforcement.”\(^{179}\) Yet this special needs doctrine applies only where public safety is at issue, such as in drug testing of railroad employees and federal customs agents, or where necessary to protect the health and safety of public school students under a loco parentis theory.\(^{180}\) Nevertheless, the *Sanchez* Court wedged welfare home visits into the special needs category, even though welfare is not an issue of public safety.

Despite the evolution of thirty years of Fourth Amendment law, courts in other states have also upheld TANF home visits, leading Jordan Budd to conclude that “the law actually matters little; the poor, presumptively different, inhabit their own constitutional universe.”\(^{181}\) The adherence to *Wyman* is all the more indefensible given how welfare has changed since

---


\(^{174}\) 464 F.3d 916, 916 (9th Cir. 2006), *reh’g en banc denied*, 483 F.3d 965 (2007).

\(^{175}\) Id. at 927.

\(^{176}\) *Wyman*, 400 U.S. at 319-20.


\(^{178}\) Id.


\(^{180}\) Budd, *supra* note 177, at 398.

\(^{181}\) Id. at 403-04.
the 1970s, when it was a means-tested program that came under attack for encouraging welfare dependency.\textsuperscript{182} Since 1996, welfare recipients have not only been subject to a five-year lifetime limit on receipt of benefits, but they also must work as a condition of receiving aid.\textsuperscript{183} They are fulfilling their part of this new social contract, but the terms still include humiliation and stigma.

\textit{b. Drug Testing}

Welfare recipients have fared somewhat better in challenging state-mandated drug testing, which is expressly authorized in TANF.\textsuperscript{184} However, their victory here is tenuous and may be short-lived. In 2000, the district court in \textit{Marchwinski v. Howard} struck down a Michigan law authorizing suspicionless drug testing of TANF applicants.\textsuperscript{185} The court stated that although the state’s professed need to address substance abuse as a barrier to employment was “laudable and understandable,” it was not a public safety issue and thus, did not justify dispensing with the ordinary Fourth Amendment requirement of individualized suspicion.\textsuperscript{186} The court rejected the state’s argument that a “special need” arose from its interest in protecting children from drug abusing parents, explaining that TANF is not directed at child abuse or neglect.\textsuperscript{187} Thus, the TANF program “cannot be used to regulate the parents in a manner that erodes their privacy rights in order to further goals that are unrelated to the welfare program.”\textsuperscript{188} In so holding, the district court refused to allow governmental assistance to become an unlimited tool for social control.

By contrast, the initial Sixth Circuit panel concluded on appeal that welfare mothers have a diminished expectation of privacy because “welfare assistance is a very heavily regulated

\textsuperscript{182} \textit{See} WALTER I. TRATTNER, \textit{FROM POOR LAW TO WELFARE STATE: A HISTORY OF SOCIAL WELFARE IN AMERICA} 362-85 (6th ed. 1999) (describing attacks on AFDC that lead to enactment of TANF).

\textsuperscript{183} 42 U.S.C. § 602(a)(1)(A)(ii) (work requirement); \textit{id.} § 608(a)(7) (five year limit).


\textsuperscript{185} 113 F. Supp. 2d 1134 (E.D. Mich. 2000). The decision was overturned by the Sixth Circuit, 309 F.3d 330 (6th Cir. 2002), but subsequently the en banc court split evenly on the issue, 319 F.3d 258 (6th Cir. 2003). Under Sixth Circuit rules, the split resulted in an affirmance of the district court’s opinion. 60 F. App’x 601, 601 (6th Cir. 2003).

\textsuperscript{186} \textit{Marchwinski}, 113 F. Supp. 2d at 1140.

\textsuperscript{187} \textit{Id.} at 1142.

\textsuperscript{188} \textit{Id.}
area of public life.” In reversing the district court, the Sixth Circuit identified two public safety justifications for conducting a suspicionless search: (1) the need to protect children from abuse by drug-addicted welfare mothers; and (2) the need to protect the public from the crime associated with illicit drug use and trafficking. This reasoning ignores empirical evidence that the use of illicit drugs by welfare recipients is no greater than in the U.S. population at large. The appellate court also found that there was a “special need” to protect the public fisc from abuse. Yet under this expansive reasoning, “the simple receipt of a tax deduction, credit, or subsidy empowers the state to conduct warrantless and suspicionless searches to verify that the beneficiary does not use the funds to buy contraband.” Of course, this reasoning can justify drug testing on all Americans, but the government is unlikely to use such strategies on middle-class Americans.

On en banc review of the court of appeals decision, the Sixth Circuit split evenly, leaving the result of the district court opinion intact—for now. Across the country, state legislatures have expressed a renewed interest in suspicionless drug testing of welfare recipients, and Florida recently implemented drug testing for all welfare applicants. Some congresspersons have even suggested drug testing for recipients of unemployment insurance. As the class of economically stressed Americans grows, so do calls for increased public drug testing programs. The stigma of drug testing is a way to discourage the needy from seeking assistance. It diverts attention

189 Marchwinski, 309 F.3d at 337.
190 Id. at 335-36.
191 Budd, supra note 21, at 776-77.
192 Marchwinski, 309 F.3d at 337.
193 Budd, supra note 21, at 799.
194 Marchwinski v. Howard, 60 F. App’x 601, 601 (6th Cir. 2003).
195 See, e.g., Chris L. Jenkins, Bill Would Require Some to Pass Drug Test to Get Aid, WASH. POST., Feb. 19, 2008, at B5 (discussing proposed bill in Virginia, as well as efforts in Kentucky and Arizona); Budd, supra note 21, at 754 (“[O]ver half of the states have considered legislation linking the receipt of public assistance to mandatory screening for drug use.”).
away from systemic problems underlying the modern economy and towards the private behavior of citizens. It allows the government to wash its hands of need.

2. Informational Privacy Under the Constitution

The Supreme Court has suggested there might be a right to informational privacy under the Fourteenth Amendment’s substantive due process clause, although the contours of this right remain murky.\(^{198}\) The right appears to protect against disclosure of personal information to third parties, rather than its collection, and as a result, it provides scant protection for the poor.

The Court’s most recent articulation of this right occurred in 2011 in *NASA v. Nelson*,\(^ {199}\) which involved twenty-eight employees, including scientists, engineers, and administrators, who worked for the Jet Propulsion Laboratories at the California Institute of Technology pursuant to a contract with the National Aeronautics Space Agency (NASA).\(^ {200}\) In 2007, NASA began requiring that these workers submit to a background investigation, regularly used for federal workers, that asks whether they have used, possessed, manufactured or sold drugs in the past year, and if so, if they received drug counseling or treatment in the past year, and that also asks a wide range of personal references if they have any reason to believe that the employee is unsuited for federal work.\(^ {201}\) Failure to comply with the background investigation results in termination of employment.\(^ {202}\) The employees at issue were all classified as “non-sensitive” employees for security purposes, and thus claimed that the background investigation violated their constitutional right to information privacy.\(^ {203}\) The Ninth

\(^{198}\) *NASA v. Nelson*, 131 S. Ct. 746, 756 (2011) (assuming without deciding that there is a constitutional informational privacy right).

\(^{199}\) *Id.* at 746. Prior to *Nelson*, the only Court precedent on point dated from other thirty years ago. In *Whalen v. Roe*, 429 U.S. 589, 599 (1977), the Court recognized “an individual interest in avoiding disclosure of personal matters.” *Id.* at 599. Still, the Court in *Whalen* held that New York State could maintain a centralized computer file containing the names and addresses of all persons who obtained legal prescriptions for Schedule II drugs, which are drugs that have both legitimate and illegal uses. *Id.* at 591. In *Nixon v. Administrator of General Services*, 433 U.S. 425 (1977), the Court held that despite an information privacy right, President Nixon could not prevent government archivists from reviewing his papers because there were adequate protections against dissemination and the intrusion was limited. *Id.* at 455-65.

\(^{200}\) 131 S. Ct. at 752.

\(^{201}\) *Id.* at 752-53.

\(^{202}\) *Id.* at 752.

\(^{203}\) *Id.* at 752, 754.
Circuit agreed and enjoined the investigations, reasoning that the requested information was not narrowly tailored to the government’s interests in a drug free and secure workplace.\textsuperscript{204}

The Supreme Court reversed. It assumed, without explicitly holding, that there is a right to informational privacy under the Constitution.\textsuperscript{205} The Court concluded, however, that the background investigation did not implicate that right because there were adequate safeguards against public disclosure.\textsuperscript{206} The Court also pointed to the government’s interest in ensuring the security of its facilities, the fact that the employees engage in “important work” on the space program, and the pervasiveness of background checks in private employment.\textsuperscript{207}

In concurrence, Justice Scalia said he would reject a constitutional right to informational privacy because it is unmoored to any constitutional provision, and he mocked the majority’s “sheer multiplicity of unweighted, relevant factors.”\textsuperscript{208} He asked if the outcome would be different if the employees were not engaged in “important work,” but were instead “janitors and maintenance men.”\textsuperscript{209} Of course, the answer is no: low-wage employees have never had reasonable expectations of workplace privacy. Under current law, the informational right to privacy is not implicated by the manner of the government’s collection of personal information.\textsuperscript{210} At most, it protects against public disclosure of that information.\textsuperscript{211} For the poor, public disclosure is a concern, but so is the humiliating procedure by which personal information is gathered. On this, the courts are silent.\textsuperscript{212}

\begin{flushright}
\textsuperscript{204} Id. at 754.
\textsuperscript{205} Id. at 756.
\textsuperscript{206} Id. at 756-57.
\textsuperscript{207} Id. at 757-60.
\textsuperscript{208} Id. at 763, 768 (Scalia, J., concurring).
\textsuperscript{209} Id. at 768.
\textsuperscript{210} Id. at 761 (majority opinion) (the concern is protecting against government disclosure of private information).
\textsuperscript{211} Id. at 755-56.
\textsuperscript{212} A few state constitutions protect privacy, and California has extended privacy protection to private conduct. Interpreting the California Constitution, a California appellate court held that pre-employment personality tests that asked about sexual orientation and religion violated the rights of applicants for security guard positions at Target. See Soroka v. Dayton Hudson Corp., 1 Cal. Rptr. 2d 77 (Cal. Ct. App. 1991). The court held that the questions were not related to job competence. Id. at 86. The case settled before it could be reviewed by the California Supreme Court.
\end{flushright}
B. Privacy Statutes

Unlike many other countries, the United States does not have comprehensive privacy regulation. Instead, our federal and state statutes tackle discrete privacy issues in a piecemeal and reactive fashion. A high-profile, but typical example is the Video Privacy Protection Act of 1988, which forbids video stores from disclosing video rental records. This law was enacted after the Supreme Court confirmation hearings of Judge Robert Bork, when newspapers published a list of videos rented by the judge, thereby causing a public uproar. Rather than a comprehensive privacy law, the United States relies mostly on self-regulation by the entities that gather and maintain personal data and puts the onus on individuals to police their own data disclosures.

Generally, American privacy statutes are guided by Fair Information Practices, which the Department of Health Education and Welfare developed in 1973 in recognition that “people have come to distrust computer-based record-keeping operations.” There are five underlying principles: (1) record-keeping systems should not be secret; (2) people should be able to find out what personal information is contained in records; (3) people should be able to prevent information obtained for one purpose from being used for another; (4) people should be able to correct records about them; and (5) organizations that maintain personal data should ensure the data is reliable and take steps to prevent its misuse. These principles require transparency in data collection and storage, but otherwise do not constrain the methods or manner by which data is collected. As one scholar has summarized, “the Golden Rule of informational privacy [is that] sensitive personal information given for one purpose ought not be used for other purposes

---

214 Schwartz, supra note 213, at 912.
216 Schwartz, supra note 213, at 935-36.
217 See Nehf, supra note 1, at 6-7.
219 Id. at 41-42.
without the express consent of the person to whom the information relates.”

1. The Golden Rule and Information Disclosure

Certainly, adherence to this Golden Rule would benefit all Americans, including the poor. For instance, the federal government oversees the Homeless Management Information System (HMIS), which requires homeless service providers to gather data about the homeless for the stated purpose of having a more accurate count of the homeless and better understanding for meeting their needs. The homeless are asked to reveal general biographical information (such as name, birth date, and social security number), and can also be asked about any physical or developmental disabilities, HIV/AIDS status, mental health, substance abuse, and domestic violence. Each homeless person is given a “unique person identification number.”

The Department of Housing and Urban Development (HUD), which regulates HMIS programs, is aware that homeless individuals might be reticent to turn over personal data due to “not wanting to be tracked, general privacy issues, vanity, embarrassment, paranoia, a desire not to qualify for a particular service, fear of being turned away, or simply not caring enough.” Accordingly, HUD standards regulate the uses and disclosure of personal information by homeless service providers. Disclosures are permitted only under certain circumstances, such as to avert a serious threat to health or safety. Yet there are still concerns that these protections are

---

222 ENHANCING HMIS DATA QUALITY, supra note 221, at 50-51.
223 Id. at 53.
224 Id. at 12.
225 See Homeless Management Information Systems (HMIS), Data and Technical Standards Final Notice, 69 Fed. Reg. 45,888, 45,928 (July 30, 2004). Domestic violence advocates were concerned about HUD requirements that shelters collect data about clients because:

The confidentiality of the data can be breached in various ways. The rules permit disclosures to oral law enforcement requests, which facilitates impostors pretexting the data. The technical standards do not require that data disclosures be logged, which limits the ability to track these impostors.
inadequate. For instance, permissible disclosures include those made in response to oral requests by law enforcement officials for the purpose of identifying or locating a suspect or material witness. As one commentator has noted, “The ease of accessibility to client [personal information] through oral requests threatens to compound the already challenging task of eliciting complete and accurate information from homeless clients,” who are, by virtue of their homelessness, often living in violation of laws that regulate their public conduct. Thus, homeless individuals may decline social services in order to protect themselves from arrest. Accordingly, protections against disclosure of personal information are extremely important for the homeless.

At the same time, even the best protection against disclosures does not ameliorate the impact of data collection. “[W]hether or not a specific individual can be related back to data generated out of that individual, the life of that data will absorb and transform the life of that individual.” This is because the entire homeless population is subject to the decisions that result from the aggregation of the data. In other words, “the data determines what kinds of life are made available by programs targeting the homeless.” For this reason, one commentator states, “Contrary to HUD’s claims, this population does not merely present an accurate picture of homelessness in the U.S., but it rather re-makes homelessness by reconfiguring what needs are allowed to register, and what services can address those needs.”

Yet privacy law focuses resolutely on the individual. Moreover,

---

Insider fraud in law enforcement agencies can also be used to breach security.

*Homeless Management Information Systems and Domestic Violence*, ELEC. PRIVACY INFO. CTR., http://epic.org/privacy/dv/hmis.html (last visited Feb. 1, 2012). HUD is planning guidance that responds to these issues in light of requirements in the Violence Against Women Act that require client consent before data disclosures are made. *Id.*


228 *Id.* (“The population is a living entity injected with biopolitical force that acts back upon that which made it.”).

229 *Id.* (noting that the agencies are using the data to secure funding and HUD approval, rather than to improve services).

230 *Id.* at 248.

231 See J.P. v. DeSanti, 653 F.2d 1080, 1081-82 (6th Cir. 1981) (finding no constitutional violation when juvenile social histories were shared with fifty-five different social, governmental, and religious agencies).
even with adherence to the Golden Rule, poor people would still suffer the stigmatization and humiliation that occur when information is collected because statutes do not generally address this phase of information transfer.

2. Federal and State Privacy Statutes

The Privacy Act of 1974 is the primary statute regulating how federal government agencies manage information about individuals. In 1998, the Act was extended to “computer matching,” which occurs when federal and state agencies compare data about individuals. TANF applicants are subject to extensive computer matching. The Privacy Act requires, among other things, that individuals subject to matching have opportunities to receive notice and to refute adverse information when benefits are denied or terminated. As a result, when an applicant applies for TANF, she should receive notice that the state agency may be obtaining and matching federal records to verify her eligibility information.

The Act’s protections are detailed and elaborate, but offer limited protection for welfare applicants. To begin with, the Privacy Act is focused on protecting information from governmental misuse once it is gathered. It does not focus on the methods or forms of collection, which in the welfare system are demeaning and stigmatizing. Further, the Act’s requirements of notice and consent are generally meaningless, because on welfare applications these provisions usually contain difficult to understand jargon hidden among the reams of information and questions contained in the forms. Finally, the Privacy Act does not govern the massive amounts of personal information held by state and local agencies, and statutory protections at this level diverge widely.

There are other federal privacy laws that are concerned with protecting individuals from the disclosure of personal information that could be embarrassing if revealed to the

---

234 See Brown, supra note 58, at 428-29 (describing requirements of Privacy Act as they apply to TANF applicants). Brown also discusses important privacy issues surrounding immigration status. Id. at 430-32.
235 Id. at 428.
236 Schwartz, supra note 213, at 916-18 (discussing state statutes addressing privacy).
public. Thus, there are statutes that protect against disclosure of credit histories, student records, debts, bank records, tax returns, television viewing habits, health information, and (as discussed above) video rentals. Obviously, Americans from every social class benefit from these protections. Still, this bevy of statutes does not protect anyone from the embarrassment that occurs when the government or private entities gather information in an intrusive or demeaning manner in the first place. This mistreatment tends to happen disproportionately to the poor and other marginalized groups. Yet another group of statutes protects individuals from unwanted intrusion into their private affairs, including laws that limit hacking and unsolicited e-mails and that create do-not-call registries. Again, these statutes erect a wall; they do not mediate ongoing relationships between individuals and the government or corporations. As such, these statutes are not models for reconsidering surveillance of the poor.

Neither are employment laws. In our at-will system of employment, private employers face few restraints in monitoring their employees. Although employee monitoring is a subject for collective bargaining for those employees who are members of unions, this is an ever-decreasing share of the workforce. While the Electronic Communications Privacy Act of 1986 prohibits the interception of data transmitted by electronic means, it is riddled with exceptions that essentially take private employers out of its reach. As a result, companies can


239 O’Gorman, supra note 27, at 217-18. Some notable exceptions are laws prohibiting lie detector tests as a condition of employment and laws limiting employers from making employment decisions based on arrest records that disproportionately impact minorities. MADELEINE SCHACHTER, INFORMATIONAL AND DECISIONAL PRIVACY 41 (2003). Employers are also subject to the Fair Credit Reporting Act if they conduct background credit checks of applicants or employees. Id.

240 Video surveillance, physical examinations, drug and alcohol testing, and polygraph testing are all mandatory subjects of collective bargaining under the NLRA, and notice must be provided. Alexandra Fiore, Note, Undignified in Defeat: An Analysis of the Stagnation and Demise of Proposed Legislation Limiting Video Surveillance in the Workplace and Suggestions for Change, 25 HOFSTRA LAB. & EMP. L.J. 525, 540-41 (2008). Penalties for employers, however, are merely a “slap on the wrist.” Id. at 542.

monitor their employees as they work on computers or engage in phone calls. Over the years, Congress and state legislatures have considered bills that would limit employer surveillance by, for instance, giving employees greater notice of when they were being monitored or limiting monitoring of long-term employees, but none has passed. This lack of statutory protections falls hardest on low-wage employees who are monitored most extensively.

C. The Common Law

The entire body of privacy law emerged from the common law, largely as a result of a path-breaking law review article written in 1890 by Samuel Warren and Louis Brandeis called *The Right to Privacy* which articulated a “right to be let alone.” Concerned about an overzealous and sensationalistic press coupled with instantaneous photography, Warren and Brandeis asserted that the “common law secures to each individual the right of determining, ordinarily, to what extent his thoughts, sentiments, and emotions shall be communicated to others.” This new conception of privacy influenced state courts to recognize a common law right to privacy, and by the mid-twentieth century most states recognized four distinct privacy torts: (1) intrusion upon seclusion, (2) public disclosure of embarrassing private facts, (3) publicity that places a person in a false light in the public eye, and (4) commercial appropriation of a person’s name or likeness.

As few poor people are celebrities, the most relevant tort for this discussion is the tort of intrusion upon seclusion, which protects an individual from intrusion upon his “solitude or seclusion... or his private affairs or concerns... if the intrusion would be highly offensive to a reasonable person.”

---

242 *Id.* at 409-10.

243 Even public employees have limited protections because private employment law concepts have seeped into the constitutional analysis such that there is not much of a difference. *Id.* at 400; Sprague, *supra* note 135, at 114 (“Both the constitutional and common law rights to privacy require an underlying expectation of privacy; so, in this regard, the analysis is the same in both the public and private employment scenario.”). Moreover, the Fourth Amendment “does not address questions of the intensity or impersonality of the surveillance.” Rothstein, *supra* note 241, at 401.


245 Sprague, *supra* note 135, at 98.


Employees have occasionally been successful in asserting this tort against highly offensive privacy violations committed by employers, such as video surveillance of bathrooms and locker rooms.\textsuperscript{249} Being videotaped covertly while engaged in private acts is distressing and causes psychological trauma.\textsuperscript{250} Of course, this sort of conduct occurs disproportionately in low-wage workplaces,\textsuperscript{251} and most of it is never the subject of legal action.\textsuperscript{252} In addition, overt employer monitoring is not actionable because employees usually consent to it as a condition of employment.\textsuperscript{253} If employees refuse consent, they protect their privacy but lose their jobs.

Moreover, the tort of intrusion upon seclusion is essentially impotent against electronic monitoring. As one court summarized,

\begin{quote}
When courts have considered claims in the workplace, they have generally found for the plaintiffs only if the challenged intrusions involved information or activities of a highly intimate nature. . . . Where the intrusions have merely involved unwanted access to data or activities related to the workplace, however, claims of intrusion have failed.\textsuperscript{254}
\end{quote}

Today, electronic monitoring by employers includes keystroke loggers that trace every key pressed on a keyboard, phone monitoring, and video surveillance, as well as smart ID cards and GPS enabled cell phones and vehicles that track employee movements.\textsuperscript{255}

Tort challenges to these practices usually fail because the tort protects only reasonable expectations of privacy (thus mirroring Fourth Amendment standards).\textsuperscript{256} Under the common law, it is not reasonable to expect privacy in a public

\begin{footnotesize}
\begin{enumerate}
\item[(249)] Fiore, supra note 240, at 547 ("It is extremely difficult for an employee to succeed on an intrusion claim in all but the most egregious circumstances.").
\item[(250)] Robert I. Simon, Video Voyeurs and the Covert Videotaping of Unsuspecting Victims: Psychological and Legal Consequences, 42 J. FORENSIC SCI. 884, 884 (1997).
\item[(251)] NAT'L WORKRIGHTS INST., PRIVACY UNDER SIEGE: ELECTRONIC MONITORING IN THE WORKPLACE,\ available at http://epic.org/privacy/workplace/e-monitoring.pdf (listing reports of intrusive video surveillance, many of which occur in the service industry and manufacturing plants).
\item[(252)] At least three states have codified this common law protection and ban video recordings in locker rooms and restrooms and the like. See Fiore, supra note 240, at 543.
\item[(256)] See Corbett, supra note 255, at 110; O’Gorman, supra note 27, at 227-30.
\end{enumerate}
\end{footnotesize}
context, and the workplace is considered public. Even if an employee can demonstrate a cognizable privacy interest, the courts then proceed to balance employee and employer interests—a test employees rarely win because their injuries are considered isolated and individualized. While all employees face these common law limitations, the privacy intrusions for white collar workers are less visible and less humiliating. This may in part explain the lack of public outrage over employer monitoring.

The limits of the common law for securing privacy for the poor can be traced to its roots. The right to be left alone was conceived to protect society’s elites (such as Warren and Brandeis) from the glare of public scrutiny. It was grounded in property law conceptions; people “own” their own identity and should be able to decide how they present themselves to the world. In the United States, “[p]rivate is territorial and is seen as a possessive right that may be alienated preemptively and wholesale.” Once you enter the workplace or ask for governmental assistance, you leave that right at the door.

Notably, at the time Brandeis and Warren wrote their article, the poor were subject to “scientific charity,” a movement that relied on middle-class “friendly visitors” to enter the homes of the poor and to provide them moral and religious counseling. Prior to the scientific charity movement, the eighteenth-century poor were warehoused in poorhouses, which required “the poor to live within the walls of a total institution, often in uniform, and under strict rules of behavior and mandates of forced labor.” In the colonial era, the poor were

---

257 O’Gorman, supra note 27, at 237 (“[A]ctivities that are work-related are generally not considered private vis-à-vis one’s employer.”).
259 Warren and Brandeis bemoaned the “idle gossip” of the daily papers and the resultant “mental and pain and distress” suffered by the subjects. Brandeis was eventually a Supreme Court justice, while Warren was a wealthy lawyer about whom stories appeared in the local press. SOLOVE ET AL., supra note 213, at 10-11.
260 Courts have interpreted privacy as a component of property, but a countervailing narrative asserts that Warren and Brandeis understood that privacy derived from “inviolable personality,” meaning “the individual’s independence, dignity, and integrity.” Rothstein, supra note 241, at 407 (quoting Edward J. Bloustein, Privacy as an Aspect of Human Dignity, in PHILOSOPHICAL DIMENSIONS OF PRIVACY: AN ANTHOLOGY 6, 163 (Ferdinand D. Schoeman ed., 1984)) (internal quotation marks omitted).
261 Id. at 382.
262 Id.
264 GILLIOM, supra note 37, at 24.
often bound out as servants to wealthier members of the community.\textsuperscript{265} Today, as explained \textit{supra}, the welfare system continues to control the lives of poor mothers. Simply put, there has never been a historical conception of privacy for the poor. Surveillance has always been the government’s prerogative and a tool for social control. Thus, the common law right to be let alone is of little use to people who rely on public assistance for survival and who navigate an ongoing relationship with government officials and/or corporations.

III. \textsc{Reconceptualizing Privacy and Poverty}

The legal system provides scant privacy protections for the poor. Our body of privacy law is built upon the “right to be left alone,” which ill-fits the nature of the intertwined relationships between poor people and larger institutional actors. The idea of being left alone creates a class differential that shelters those who can afford it. The result is that the poor are often subject to humiliating and stigmatizing data collection practices. This raises the question of how the law can better equalize privacy among all citizens. This part explores various remedies for the class differential in privacy law and policy. First, it examines and rejects nonlegal solutions such as improved “customer” relations, on the one end, and increased automation, on the other. Then, this part explores other values that give meaning to privacy. Privacy is not an end to itself; rather, it fosters and furthers other values. Accordingly, this part suggests how norms of dignity, respect, and trust—as articulated by criminal-justice scholars—can enhance privacy for the poor in the civil realm. While each of these approaches has limitations, they provide valuable guideposts for reconceptualizing privacy for the poor.

A. \textit{Nonlegal Solutions}

1. Service with a Smile

One optimistic, but ultimately infeasible, solution is to make interactions between low-income people and larger institutions more pleasant and less humiliating. Indeed, some welfare offices have policies that encourage caseworkers to treat “clients” with respect. Still, it is difficult to mandate

politeness and questionable whether new rights-based regimes are desirable or effective, especially with populations that do not view themselves as rights-bearing individuals.

Susan Bennett has articulated how and why welfare bureaucracies use privacy-stripping tactics to discourage poor people from applying for assistance, with devastating consequences for the needy.\textsuperscript{266} Churned out of public bureaucracies, the poor suffer hunger and homelessness, as well as a “dampening of the spirit, a lowering of expectations of any kind of fair treatment.”\textsuperscript{267} In her study of a waiting room at the District of Columbia’s Office of Emergency Shelter and Support Services, she found an “ethos of undisclosed information, unexplained delays, and, above all, endless waiting, punctuated by humiliating demands for information.”\textsuperscript{268} This ethos arose from a variety of factors, such as a vague regulatory regime that permitted workers to demand extreme forms of proof;\textsuperscript{269} the front-line workers’ fear of “being spare-changed” and distrust of applicants;\textsuperscript{270} bureaucratic pressures to prevent the needy from filing applications so as to reduce welfare rolls and avoid providing due process protections;\textsuperscript{271} “external demands for fraud control,”\textsuperscript{272} and an inability to cope with rising demand in the face of decreasing resources and reduced staffing.\textsuperscript{273} In this environment, norms of “customer service” seem laughably optimistic.

On the employment side, low-wage workers experience a different workplace than white collar workers. They usually lack employment benefits such as health insurance or retirement accounts. They face inflexible or unpredictable work hours that limit access to child care and transportation. They have few opportunities for career advancement. They are more likely to work in unsafe or unhealthy conditions.\textsuperscript{274} They also have less privacy. These features of the low-income workforce are deeply structural, and require far more change than an attitude adjustment. As David Yamada explains, employment

\textsuperscript{266} See generally Susan D. Bennett, “No Relief but Upon the Terms of Coming into the House”—Controlled Spaces, Invisible Disentitlements, and Homelessness in an Urban Shelter System, 104 YALE L.J. 2157 (1995).
\textsuperscript{267} Id. at 2182.
\textsuperscript{268} Id. at 2158.
\textsuperscript{269} Id. at 2187.
\textsuperscript{270} Id. at 2188.
\textsuperscript{271} Id. at 2193-94.
\textsuperscript{272} Id. at 2194.
\textsuperscript{273} Id. at 2198-99.
law “cannot force organizations to care about the health and well-being of their employees, require workers to vote for union representation, or simply order everyone to be ‘nice’ to one another.”  

And even if the law could mandate politeness, the stigma of surveillance would stick. The sweetest social worker in the world might search your medicine cabinets with a smile. An employer may hand over a cup for a urine test with a polite request. The effect is still demeaning.

2. Automation

In lieu of mandated courtesy, it might be tempting to move in the opposite direction and limit interpersonal interaction between low-income Americans and larger institutional actors. Arguably, the more things are automated, the less opportunity there is for insult. Supporters of automation promote technology as a way to save money and ensure consistent decisions.  

However, as Danielle Citron explains, automation can and does fail. To begin with, computer programmers struggle to properly translate complex public benefits programs into code, resulting in violations of federal and state law. For instance, Citron describes how programmers in Colorado incorrectly coded nine hundred different public benefits rules into an automated system, resulting in people wrongfully losing food stamps and Medicaid. Furthermore, computers are not fail-safe; they “misidentify individuals [with] . . . same or similar names,” send faulty notices, and terminate benefits without warning. At the same time, many computer programs fail to maintain audit trails of decisions, which then make it impossible for individuals to challenge automated decisions in due process hearings.

At due process hearings, hearing officers are biased in favor of automated systems, with their veneer of objectivity and

---

275 David C. Yamada, Human Dignity and American Employment Law, 43 U. RICH. L. REV. 523, 554 (2009). Yamada notes, however, that the law can help workers by “safeguarding the rights of association and collective bargaining.” Id.


277 Id. at 1256-58 (listing examples of automation failures).

278 Id. at 1267-68.

279 Id. at 1268, 1271-72.

280 Id. at 1273.

281 Id. at 1275-76; see also Willse, supra note 227, at 240 (discussing homeless information management systems and stating that “database programs of course also fail short or fail—systems crash, networks go down, files get mysteriously deleted”).
correctness, and laypersons are hard-pressed to challenge the source code underlying automated programs.\textsuperscript{282} The end result is that computer programmers are, in effect, rewriting regulations without notice and comment.\textsuperscript{283} Automation can also be dehumanizing. As Virginia Eubanks writes, “the structure of technological systems erase the embodied contexts and knowledge of the people described in them.”\textsuperscript{284}

Even if the errors inherent in automation could be erased, the poor would likely still face differential treatment. Consider the Earned Income Tax Credit (EITC).\textsuperscript{285} The EITC, enacted in 1975, provides low-income, working Americans with a refundable tax credit, which amounted to as much as $5666 in 2010, depending on family size.\textsuperscript{286} The program lifts five million families above the poverty line each year.\textsuperscript{287} The program has significantly increased employment among single mothers and simultaneously lowered the receipt of welfare cash assistance.\textsuperscript{288} Notably, the tax credit is granted via an “impersonal and invisible process” that is far less demeaning than public benefits programs.\textsuperscript{289}

Nevertheless, low-income taxpayers claiming the EITC receive far greater scrutiny than middle-class taxpayers.\textsuperscript{290} They are audited at higher rates,\textsuperscript{291} even though over two-thirds of audited EITC claims are ultimately found to be proper.\textsuperscript{292} The IRS believes that the EITC has a high

\begin{itemize}
\item \textsuperscript{282} Citron, supra note 276, at 1283-84.
\item \textsuperscript{283} Id. at 1288.
\item \textsuperscript{284} Eubanks, supra note 49, at 99.
\item \textsuperscript{285} I.R.C. § 32 (2008).
\item \textsuperscript{286} EITC—Don’t Overlook It, INTERNAL REVENUE SERV. (Jan. 28, 2011), http://www.irs.gov/newsroom/article/0,,id=106429,00.html.
\item \textsuperscript{287} Dorothy A. Brown, Race and Class Matters in Tax Policy, 107 COLUM. L. REV. 790, 792 (2007). Brown notes, “More children are lifted out of poverty as a result of the credit than any other governmental program.” Id.
\item \textsuperscript{288} Id. at 799-800.
\item \textsuperscript{290} “No other taxpayers are subject to such scrutiny.” Brown, supra note 287, at 791-92.
\item \textsuperscript{292} See Brown, supra note 287, at 791. The government focuses disproportionate time and energy on EITC audits—since 1998, the IRS has poured over $1 billion into EITC audits. See id. at 792. However, the data shows that audits of wealthier taxpayers are far more productive. See id. at 808 (these audits “generally result in at least four times the recommended additional tax than audits of low-income taxpayers”).
\end{itemize}
overpayment rate due to fraud, but studies show that the vast majority of errors result from other, unintentional factors such as the “notoriously confusing” complexities of the program, the turnover as claimants move in and out of poverty, low literacy rates among low-income taxpayers, a lack of access to professional tax preparers, and a fear of turning over personal information to the government.

Due to the distrust of EITC taxpayers, the IRS rolled out a precertification process between 2003 and 2005 among selected portions of the EITC population, which required claimants to provide third party affidavits and documentation in support of their EITC tax filings. By contrast, the rest of the tax system relies on self-reporting. Thus, critics of the precertification process charged that this differential treatment for low-income filers was unfair and unduly burdensome.

Dorothy Brown explains the discrepancy in audit rates and filing requirements for the poor. She states that politicians publicly equate the EITC with welfare, thus knowingly triggering racialized welfare stereotypes in which, “low-income taxpayers are viewed as lazy former welfare recipients who . . . will lie and cheat in order to line their pockets with government money.” By contrast, “government subsidies that flow to predominantly white beneficiaries are not considered to constitute welfare,” such as farm subsidies. Ironically, most EITC recipients are white, which means the racial stereotypes harm all EITC-eligible taxpayers. EITC enforcement shows that perceptions of class and race lead to differential treatment even in programs that involve little face-to-face interaction. Automation alone is not the answer.

---

294 Leifeld, supra note 291, at 547. “[T]he IRS publication associated with the EITC is over fifty pages long with six separate worksheets.” Brown, supra note 287, at 792.
295 See Leifeld, supra note 291, at 550.
296 See Brown, supra note 287, at 809-10.
297 See, e.g., id. (explaining that “[n]o other tax provision requires precertification, while all welfare-type programs require it”).
298 Id. at 799-810.
299 Id. at 793-95.
300 Id. at 814
301 Id. at 820.
B. Giving Content to Privacy Under Law

Privacy is not an end to itself; it supports and enhances other important values. Accordingly, enhancing the relationship between the poor and larger institutions requires more than leaving the poor alone. In related contexts, scholars have been working to shape a richer conception of privacy, particularly under the Fourth Amendment as it relates to criminal searches. Frustrated with the malleable reasonableness standard and its failure to restrain government surveillance, these scholars have strived to give content to the meaning of privacy. These approaches, developed mostly in the criminal law context, are helpful in considering the welfare-to-work continuum because they focus on the individual’s relationship to the state. They are also relevant because welfare has become increasingly criminalized, and because criminal law jurisprudence most impacts poor communities. These theories focus on dignity, respect, and trust as values secured by privacy. Privacy is not an end to itself; rather, it secures higher values.

To be sure, there is a tension between seeking governmental assistance and simultaneously demanding privacy from the state.304 Yet wealthier Americans would recoil in horror if the government put them through similar scrutiny as a condition of receiving governmental subsidies, such as tax deductions for mortgages and retirement plans, and childcare tax credits.305 Moreover, poor people do not need to be “left alone” by the state or employer to benefit from privacy. Rather, the advantage of the dignity-respect-trust models discussed below is that they can accommodate intertwined relationships in a way that privacy has not been able to bear. In the end, surveillance needs to be proportional to its purposes. The goal is not to shut out the state or employers, but to make them partners with low-income Americans in a flourishing democracy.

305 See MARTHA ALBERTSON FINEMAN, THE NEUTERED MOTHER, THE SEXUAL FAMILY AND OTHER TWENTIETH CENTURY TRAGEDIES 191 (1995) (“[M]iddle-class families benefit from extensive entitlement programs, be they FHA or VA loans at below mortgage market rates or employer health and life insurance. These families receive untaxed benefits as direct subsidies.”).
1. Dignity

John Castiglione argues that dignity is an equally important Fourth Amendment value as privacy.\textsuperscript{306} As he points out, under the reasonableness standard’s balancing test, it is almost impossible for an individual’s “abstract, indeterminate” privacy interest to outweigh the state’s concrete interest in law enforcement and social control.\textsuperscript{307} Accordingly, he posits that dignity can support the scaffolding of the Fourth Amendment in a way that privacy cannot.\textsuperscript{308} He notes, people (such as prisoners) can completely lack privacy but still claim “a legitimate expectation of being treated with dignity.”\textsuperscript{310} In philosophical terms, dignity is the “right to be treated as an end, not as a means.”\textsuperscript{311} In practical terms, it is the opposite of “unnecessarily degrading, humiliating, or dehumanizing government behavior.”\textsuperscript{312} In legal terms, the Supreme Court often uses the concept of dignity to inform constitutional interpretation, so it is recognized as a constitutional commitment even if rarely enforced in the Fourth Amendment context.\textsuperscript{313} Castiglione helpfully contrasts privacy, which protects access to the self, with dignity, which “generally concerns a limitation on the manner in which an individual is interacted with.”\textsuperscript{314} Under his proposal, government searches and seizures would be unlawful if they degrade or humiliate individuals without a sufficient countervailing law enforcement interest.\textsuperscript{315}

Similarly, David Yamada has posited that dignity should replace “markets and management” as the framework for American employment law.\textsuperscript{316} The current employment law regime is dominated “by a belief system that embraces the idea of unfettered free markets and regards limitations on


\textsuperscript{307} \textit{Id.} at 664.

\textsuperscript{308} \textit{Id.} at 660.

\textsuperscript{309} \textit{Id.} at 674-75.

\textsuperscript{310} \textit{Id.} at 675.

\textsuperscript{311} \textit{Id.} at 678.

\textsuperscript{312} \textit{Id.} at 687.


\textsuperscript{314} Castiglione, \textit{supra} note 306, at 688-89.

\textsuperscript{315} \textit{Id.} at 696.

\textsuperscript{316} Yamada, \textit{supra} note 275, at 524. See generally RANDY HODSON, \textit{DIGNITY AT WORK} (2001) (describing threats to dignity in the workplace).
management authority with deep suspicion." Yet this framework benefits the rich at the expense of the poor, as income inequality grows and workers face increasing job insecurity, stress, and negative health consequences—all without adequate legal recourse. At the same time, the law fails to protect against humiliation “by having a remedial structure that is arbitrary, expensive, and difficult.” By contrast, a dignitary conception of the workplace would, among other things, support unions and collective bargaining “as an invaluable source of countervailing power in society,” because they give workers a voice as well as leverage to demand a better workplace. Notably, collective bargaining is one of the few ways in our system to limit employer surveillance. Although low-wage workers often do find dignity in their work, they do so in spite of the law.

Other authors have contrasted the American system—in which “dignity is denied by treating the employee as a mere factor of production . . . and ignoring . . . the worker’s individuality”—to European workplace law, which emphasizes dignity and sharply limits surveillance. In Europe, dignity is connected to “notions of community and citizenship [rather] than property.” Under this conception, private power is seen as great a threat to dignity as public power. Thus, privacy is considered a fundamental human right. For instance, workers in France have a say in when and how employer monitoring occurs, they have the right to be informed about the automated treatment of their personal

317 Yamada, supra note 275, at 523; see also Richard Sennett, The Corrosion of Character 31 (1998) (describing how the “new capitalism” emphasis on flexibility “loosens bonds of trust and commitment, and divorces will from behavior”). Sennett writes that information systems “give individuals anywhere in the network little room to hide.” Id. at 55.

318 Yamada, supra note 275, at 530-31.

319 Fisk, supra note 133, at 92.

320 Moreover, unions are particularly important in the low-wage workforce, where exploitation is rampant. Yamada, supra note 275, at 557. Dignitary norms would also provide protections against unjust or unfair dismissal, id. at 558-61, limit workplace bullying, id. at 562-65, and improve dispute resolution procedures for employment-related conflicts, id. at 566.


323 Rothstein, supra note 241, at 383; Nehf, supra note 1, at 81-82.

324 Rothstein, supra note 241, at 386.

325 Nehf, supra note 1, at 81-82.
information, they receive notice about the scope of monitoring, and monitoring must be proportional to the employers’ objectives. Reforming privacy to reflect dignitary values could be helpful to low-income Americans because dignity is an inviolable, core human right that cannot be bought or sold by those with more access to wealth. At the same time, the individual emphasis on dignity can obscure the class-based motivations for and consequences of surveillance.

2. Respect

Andrew Taslitz suggests a class-conscious approach that hinges upon respect as a central value underlying privacy. He explains that courts envision the “reasonable person” interacting with the police from the perspective of a white middle-class person “rather than the poor person familiar with police abuse.” Taslitz is particularly attuned to the disparate impact of current Fourth Amendment doctrine on racial minorities entangled in the criminal justice system. As he defines respect, it “is also about inclusion, about being considered full members of the wider political community.” He explains how a lack of governmental respect impacts communities, which suffer when they are targeted for suspicionless searches. Those searches “send a message to their victims that they are unworthy of the government’s respect.” Accordingly, he urges courts to expand their perspectives to better understand and acknowledge minority group experiences. This idea of respect would be helpful in shaping the experiences of low-income Americans as they interact with welfare offices and low-wage employers. Under his conception of respect, courts would have to consider both the impact of privacy law from the perspective of those under surveillance as well as the costs to civil society and


328 Id. at 56.

329 Id. at 21.

330 Id. at 27.

331 Id. at 23.

332 Id. at 23-24.

333 Id. at 92-97.
democratic participation when entire groups are demeaned and subordinated.

The challenge to this approach is that the poor have long been deemed undeserving of respect. The majority of Americans believe that this country is a meritocracy, by which anyone can lift themselves up by their bootstraps. This myth leads to cultural explanations for poverty that blame the poor for their own predicament.\textsuperscript{334}

As I have explained elsewhere, the cultural explanation of poverty is founded on conjecture masquerading as common sense, but has no empirical support.\textsuperscript{335} Nevertheless, it has had remarkable staying power because it demands less from government and it appeals to the economically insecure middle class.\textsuperscript{336} By contrast, the real causes of poverty are far more complex. An amalgamation of economic and demographic factors contribute to poverty, including declining labor market opportunities, the erosion of the minimum wage and low-wage income, deindustrialization, technological changes in the economy, globalization, the decline of unions, and the increased use of contingent workers who are low-wage, part-time, and lack benefits.\textsuperscript{337} In addition, governmental urban policies have segregated poor minority communities into areas of concentrated poverty.\textsuperscript{338} Nevertheless, the entrenched myth of the meritocracy makes it difficult to build a legal theory around respect.

3. Trust

Scott Sundby is similarly dismayed by the Fourth Amendment balancing test, and he advances a reciprocal government-citizen trust model under which the government could not “intrude into the citizenry’s lives without a finding that the citizenry has forfeited society’s trust to exercise its

\textsuperscript{334} See Handler & Hasenfeld, supra note 25, at 159-61 (describing culture of poverty theories).

\textsuperscript{335} See Michael B. Katz, The Price of Citizenship 320 (2008) (“In the welfare debates of the 1990s, conservative accounts of research simply misrepresented the evidence.”).


\textsuperscript{337} See id. at 66-72; Iceland, supra note 40, at 77-78; see also Michael B. Katz, Improving Poor People: The Welfare State, the “Underclass,” and Urban Schools as History 77-78 (1997); Joel Handler, “Ending Welfare as We Know It”—Wrong for Welfare, Wrong for Poverty, 2 GEO. J. ON FIGHTING POVERTY 3, 10-12 (1994).

freedoms responsibly.” In other words, citizens would not be searched unless they did something to raise suspicion of wrongdoing. His citizen-trust model applies most appropriately to situations in which the state initiates an intrusion into individual privacy (as is the case with public benefits regimes) rather than responds to perceived wrongdoing. This model recognizes that “[g]overnment action draws its legitimacy from the trust that the electorate places in its representatives by choosing them to govern.” Right now, poor Americans do not feel “they have the opportunities and capabilities to participate meaningfully in society,” in part due to privacy invasions that signal a lack of trust by government. “Rights are not simply enclaves of protection from government interference but also affect the citizen’s view of his or her role in society.”

Unfortunately, welfare mothers and low-wage workers have almost no conception of rights—they are “the inverse of the rights-bearing individual.” Moreover, a citizenship based approach to privacy does not extend to the private market, which often has just as much power over individuals as does the state. Nor does it protect noncitizens, who are nevertheless entitled to the human right of dignity. Finally, poor people who are citizens are often excluded from mainstream norms of citizenship. As Dorothy Roberts explains, the welfare system treats its recipients as subjects rather than citizens for reasons of class, gender, and race. Whereas citizens receive government benefits such as social security as an entitlement free of stigma, subjects “receive inferior, inadequate, and stigmatizing relief at the government’s discretion.” She concludes, “While welfare for citizens enables them to be self-ruling persons, welfare for subjects enables the government to rule them.” Until our conception of citizenship includes the poor, a citizen-trust model may not advance their privacy.

---

339 Sundby, supra note 154, at 1777.
340 Id. at 1787.
341 Id. at 1777.
342 Id. at 1784.
343 Gilliom, supra note 37, at 91.
345 Id.
346 Id.
C. The Fragile Nature of Privacy

All three of these models—dignity, respect, and trust—add to our understanding of why privacy is valuable, particularly for our most vulnerable citizens. Adoption of any of these theories could transform the experiences of poor citizens as they interact with the state and their employers. Nevertheless, the courts appear reluctant to weave these strands into privacy jurisprudence, even when courts rule in favor of privacy.

Consider Ferguson v. City of Charleston,\(^347\) in which the Supreme Court used the Fourth Amendment to strike down a public hospital’s policy (created in conjunction with local police and prosecutors) of drug testing poor, pregnant patients suspected of using drugs.\(^348\) The hospital implemented the policy in the wake of a perceived crack epidemic in which thousands of babies were reportedly being born to drug addicted mothers.\(^349\) Under the policy, patients who tested positive were referred to substance abuse treatment programs under the threat of arrest if they did not comply or if they tested positive in a subsequent screening.\(^350\) The policy was enforced only in the hospital’s Medicaid maternity clinic; it was not used in other hospital departments or for pregnant women who paid for their care.\(^351\) The class, race, and gender differential in privacy was clear.

The Court first ruled that the urine tests constituted searches within the meaning of the Fourth Amendment.\(^352\) As the Court stated, “The reasonable expectation of privacy enjoyed by the typical patient undergoing diagnostic tests in a hospital is that the results of those tests will not be shared with nonmedical personnel without her consent.”\(^353\) Further, prior cases approving drug testing involved a special need disconnected from law enforcement.\(^354\) By contrast, in Ferguson, “the central and indispensable feature of the policy from its inception was the use of law enforcement to coerce the patients into substance abuse treatment.”\(^355\)

\(^{348}\) Id. at 70-73.
\(^{349}\) Id. at 70 n.1.
\(^{350}\) Id. at 71-72.
\(^{351}\) Campbell, supra note 91, at 69; Andrew E. Taslitz, A Feminist Fourth Amendment?: Consent, Care, Privacy, and Social Meaning in Ferguson v. City of Charleston, 9 DUKE J. GENDER L. & POL’Y 1, 23 (2002) (“[T]he Policy was applied almost entirely to economically disadvantaged African-American women.”).
\(^{352}\) Ferguson, 532 U.S. at 76.
\(^{353}\) Id. at 78.
\(^{354}\) Id. at 79.
\(^{355}\) Id. at 80.
While the hospital argued that it had identified criteria that raised suspicion of cocaine use, the Court found no evidence in the record “that any of the nine search criteria was more apt to be caused by cocaine use than by some other factor, such as malnutrition, illness, or indigency.” 356 In so doing, the Court implicitly recognized the harmful effects of poverty and refused to punish the mothers for being poor. Thus, the positive narrative of Ferguson is that it took into “account the experiences and values of women,” and “gave patient and parental autonomy great weight” in a manner that “was offended by paternalistic notions.” 357 The Court appraised the context of the searches, recognized the stigma that attached to the women who were drug tested, and limited the criminalization of poverty.

However, the real-life impact of Ferguson remains limited. The Court’s holding hinges entirely upon its disapproval of police involvement in crafting and enforcing the hospital’s policy; at no time does the Court articulate any values underlying the patients’ claim for privacy. This absence increases the difficulty of stretching Ferguson to cases lacking police involvement, as the Court seemed to intend with its narrow holding.

Thus, the scrutiny of the lives of poor, pregnant women remains. In her ethnographic study, Khiara Bridges describes how poor, pregnant patients seeking Medicaid coverage for prenatal health care costs are subject to mandatory interviews that probe the most intimate corners of their lives. 358 The information gathered goes far beyond what wealthier pregnant women are expected to divulge to their doctors: it involves detailed assessments regarding nutrition, psycho-social factors, and finances conducted by a range of nurses, social workers, and other professionals. 359 The women receive education about contraceptive options throughout their pregnancies and access to long-acting contraception following the birth of their babies. 360 As a result, the state has “all the information necessary to sweep poor families within the ambit of child protective services, the foster care system, Immigration and Customs Enforcement, and, if deemed necessary, the criminal justice system.” 361

---

356 Id. at 77 n.10.
357 Taslitz, supra note 351, at 3.
358 Bridges, supra note 52, at 113, 114.
359 Id. at 124-34. Among other things, the inquiries involve “women’s sexual histories, experiences with substance use and abuse, histories of sexual and domestic violence, and strategies for preventing the conception and birth of more children . . . .” Id. at 163.
360 Id. at 151-32.
361 Id. at 132.
Moreover, Ferguson does not change the fact that many states criminally and civilly penalize a woman’s drug use during pregnancy. In addition, at least three states have mandatory drug testing of pregnant women in some circumstances. One of the harshest state statutes is in South Carolina, where the Ferguson case originated. In South Carolina, a positive drug test at the time of birth results in a presumption of parental neglect that warrants removal of a child from the mother’s custody. South Carolina also stands alone in approving criminal prosecutions of women who have used drugs during pregnancy. Although an estimated two hundred women with drug addictions have been criminally prosecuted in thirty states for fetal abuse, most appellate courts have overturned those convictions. In South Carolina, by contrast, the state Supreme Court held that “fetuses are ‘person[s]’ under the state’s criminal child endangerment statute,” and affirmed the conviction of a mother who had used cocaine during her pregnancy. Not surprisingly, drug treatment programs in South Carolina have seen their admissions of pregnant women drop, infant mortality rates have risen, and there has been a 20 percent increase in abandoned babies. Without a value-driven approach to privacy, the class differential in privacy policies is likely to continue. Dignity, respect, and trust are values that can inform privacy law and policy.

IV. EQUALITY IN PRIVACY LAW

As Ferguson demonstrates, the courts do not explicitly equate spatial or informational privacy with the values of dignity, respect, or trust, even though these principles would provide meaningful guideposts to distinguishing appropriate

362 See Lynn M. Paltrow, Governmental Responses to Pregnant Women Who Use Alcohol or Other Drugs, 8 DePaul J. Health Care L. 461, 464-65 (2005). In some jurisdictions without specific laws on the subject, government officials have nevertheless implemented policies that extend civil child abuse laws to pregnant women. Id. at 467, 474.
363 Id. at 467.
364 Id. at 465.
365 Id. at 485.
366 Id. at 488.
367 Id. at 490-91. These punitive, privacy-stripping approaches are fueled by media hysteria rather than science. Id. at 475. The science shows that cocaine is not always harmful to children; that drug abuse is a treatable addiction, rather than a moral failing; that a single drug test does not predict parenting ability; and that removing children from their mothers inflicts grave harm on children. Id. at 475-82.
data collection from demeaning surveillance.\textsuperscript{368} Interestingly however, there is one consistent strain in privacy cases that emerges over the years: class equality. In his Wyman dissent in 1971 (upholding welfare home searches), Justice Douglas pointed out the class differential in privacy law, stating, “No such sums are spent policing the government subsidies granted to farmers, airlines, steamship companies, and junk mail dealers, to name but a few.”\textsuperscript{369} Over thirty years later, the dissenting judge in Sanchez v. San Diego (a modern welfare home visit case) wryly pointed out, “I doubt my colleagues in the majority would disagree that an IRS auditor’s asking to look in [‘medicine cabinets, laundry baskets, closets and drawers for evidence of welfare fraud’] within their own homes to verify the number of dependents living at home would constitute snooping.”\textsuperscript{370} Similarly, the district court in Marchwinski (overturning drug testing of welfare applicants) quoted Justice Marshall to ask, “Would the majority sanction, in the absence of probable cause, compulsory visits to all American homes for the purpose of discovering child abuse?”\textsuperscript{371} As Justice Marshall stated in 1971, “Such a categorical approach to an entire class of citizens would be dangerously at odds with the tenets of our democracy.”\textsuperscript{372}

Justice Marshall’s warning is today’s reality. Poor people continue to suffer privacy invasions that generate stigma and humiliation. In addition to individual harms, this class differential in privacy harms democracy. Two in-depth studies demonstrate the link between surveillance and decreased democratic participation by the poor. First, Joe Soss, in his study of welfare, has explained that poverty “strip[s] individuals of the ability and time needed to follow or participate in political affairs,” as well as “the autonomy needed for self-government.”\textsuperscript{373} Moreover, interactions with welfare workers leave recipients “pessimistic about government’s responsiveness and the efficacy of political action.”\textsuperscript{374} As a result, being a welfare

\textsuperscript{368} In cases involving decisional privacy, the Court has moved away from privacy rhetoric, instead relying on liberty. See generally Jamal Greene, The So-Called Right to Privacy, 43 U.C. DAVIS L. REV. 715 (2010).


\textsuperscript{370} Sanchez v. Cnty. of San Diego, 464 F.3d 916, 936 (9th Cir. 2006) (quoting Wyman, 400 U.S. at 342 (Marshall, J., dissenting)) (internal quotation marks omitted).


\textsuperscript{372} Wyman, 400 U.S. at 342.


\textsuperscript{374} Id. at 164.
recipient “reduces the odds that a person will vote to slightly less than half of what it would have been otherwise.” For poor women, public benefits programs are their “most direct exposure to . . . [political] institution[s],” and the lessons welfare recipients learn there “have significant consequences for broader patterns of political action.” Surveillance is one way that the welfare system “position[s] clients as objects of paternalism,” which in turn leaves them feeling that government is not responsive to their concerns.

Second, John Gilliom has also studied how welfare surveillance suppresses political action. He found that while welfare mothers resist state control over their lives, they do not use litigation or democratic processes to do so. Rather than viewing privacy violations from a rights-based perspective, they “focus on need and on their duties to care for their families.” Thus, their methods of opposition are more subtle; for instance, they earn unreported income to supplement meager welfare checks. While this defiance sustains individual autonomy in the face of the state’s power, it also causes stress to welfare mothers who fear getting caught and possibly losing benefits or being punished criminally. Moreover, it obviates political organizing and leaves surveillance structures intact.

Likewise, among her low-wage coworkers, Barbara Ehrenreich found a complete lack of political consciousness or defiance of workplace privacy intrusions. As she explains, her coworkers could not just “get-up-and-go”; rather, their mobility was constrained by lack of transportation or child care, as well as a lack of information with which to compare employers. As a result, there is no competitive market pushing employers to treat employees more fairly. Further, management engenders a lack of self-respect in employees by subjecting them to random searches. Resistance in the low-wage workforce often comes in

---

375 Id. at 162.
376 Id. at 184.
377 Id. at 4.
378 Id. at 200.
379 Gilliom found that welfare mothers do not assert rights; there is “no sign of the mobilization and empowerment which might follow.” GILLIOM, supra note 37, at 85. “Wherever there is unjust power, resistance inevitably follows.” Hodson, supra note 316, at 42 (citing MICHEL FOUCAULT, POLITICS, PHILOSOPHY, CULTURE: INTERVIEWS AND OTHER WRITINGS (1988)).
380 GILLIOM, supra note 37, at 96.
381 See id. at 99-106, 113.
382 See id. at 87-88.
383 EHRENREICH, supra note 77, at 205-06.
384 Id. at 208.
the form of lower productivity; sometimes, workers quit in frustration. There is little political resistance. Ehrenreich comments, “We can hardly pride ourselves on being the world’s preeminent democracy, after all, if large numbers of citizens spend half their waking hours in what amounts, in plain terms, to a dictatorship.” The “dictatorship” is maintained, in part, by surveillance.

This class differential in privacy should be a concern to all Americans. It harms individuals by denying them full respect as citizens and limiting their autonomy. It decreases democratic participation, which in turn means there is less political check on privacy violations. The privacy differential also exacerbates income inequality by reinforcing class lines and disempowering people who in turn become too downtrodden—or busy trying to survive—to challenge public policies. The result is a downward cycle of disempowerment and class division. These costs are incurred with almost no countervailing benefits to individuals or society, other than the societal satisfaction that comes from censuring the “moral laxity” of people “unable to thrive within a capitalist economy.”

Of course, an equality approach means that privacy is only as secure as the majority has it. If Americans cross class lines and advocate together to preserve privacy and resist surveillance, we are more likely to restrain abuses of power. The scientists in Nelson v. NASA objected to intrusive government questioning conducted for what they believe are bogus security reasons. Poor Americans can relate; they have long been exposed to the prying eyes of the state. Middle-class and wealthy Americans need to realize that novel surveillance techniques are typically used first on the poor. By the time these strategies spread beyond controlling the poor, any “reasonable expectations” against their use have dissolved.

What can equality add to privacy law and policy? It has limited constitutional legs due to the lack of equal protection for the poor. Laws that distinguish on the basis of wealth are

---

385 Hodson, supra note 316, at 42-43, 60-68. Hodson describes how employees resist employer abuse by withdrawing cooperation and violating cumbersome rules. Id. at 42. “All of these forms of resistance are attempts to regain dignity in the face of organizations that violate worker’s interests, limit their prerogatives, or undermine their autonomy.” Id.; see also Carol Cleveland, A Desperate Means to Dignity: Work Refusal Amongst Philadelphia Welfare Recipients, 6 Ethnography 35, 54 (2005) (stating that low-wage employees find their dignity under assault and often quit rather than submit to perceived abuse and disrespect).

386 Ehrenreich, supra note 77, at 210.

387 See Bridges, supra note 52, at 163 (critiquing the use of poverty as a moral index).
subject to mere rational basis review, and the state can usually assert a societal reason for intruding on individual privacy. It is a difficult scale to tip. Nevertheless, equality can inform other constitutional provisions. For instance, under the Fourth Amendment, a government search may not be reasonable if targeted at individuals solely on class grounds.\textsuperscript{388} This is intimated in \textit{Ferguson} and articulated in \textit{Marchwinski}. Moreover, equality norms could be folded into the Fair Information Principles that inform privacy laws and policies. These principles currently focus on preventing unknowing disclosure to third parties. However, they could be expanded to cover the data collection phase and to forbid collection practices that intrude on privacy in a demeaning, humiliating, or stigmatizing way. In addition, statutory protections for workplace monitoring could be enacted that give employees greater notice about surveillance and limit tactics that are disproportionate to achieve employer objectives. At the same time, employers should recognize and replicate successful monitoring programs that give employees notice about monitoring, as well as a voice in surveillance strategies. Finally, the common law could recognize the privacy interests of and harms to low-wage employees as a group, rather than disaggregating claims down to an individual level. Rather than asking if a specific person has a reasonable expectation of privacy in a specific workplace, courts could consider whether employees have any reasonable expectation to be treated with dignity by employers. In short, our privacy policies and practices should be examined to ensure that they do not single out a specific class of persons for stigma. The ideal of equality can help provide a check against such targeted surveillance.

CONCLUSION

Due to advanced technologies, all Americans face corporate and governmental surveillance. However, poor Americans face different and more intense privacy violations than do wealthier Americans. While most Americans are vaguely aware that they are subject to surveillance, they do not feel its effects concretely, and they are willing to relinquish some privacy for increased security and for the conveniences of technology. Yet for the poor, surveillance is neither vague nor invisible. Rather, along the welfare-to-work continuum, poor

people face privacy intrusions at the time that the state or their employers gather data. This data collection tends to stigmatize and humiliate, not only compounding the harmful effects of living in poverty, but also dampening democratic participation by the poor.

Yet privacy law is focused on middle-class concerns about limiting the disclosure of personal data so that it is not misused. By contrast, the poor interact with the government and low-wage employers in ways that are ongoing and interpersonal, and as a result, the “right to be left alone” embodied in current privacy law does not protect their interests in dignity and autonomy. Privacy law, in its constitutional, statutory, and common law dimensions, protects reasonable expectations of privacy, but courts have long held that people give up expectations to privacy when they seek help from the government or go to work. The law thus reinforces the existing class differential in privacy practices. This class differential has costs not only for the poor, but for all citizens. The poor do not need to be left alone; they need to be treated with dignity. Privacy should not be for sale. All Americans would benefit from enhancing the privacy rights of the poor, and united, we can provide a powerful check on expanding surveillance that impacts the poor and rich alike.
Democratic Inclusion, Cognitive Development, and the Age of Electoral Majority

Vivian E. Hamilton†

INTRODUCTION

Who should vote in the modern democratic state? The question implicates the core of democratic government—popular political participation. And the vote is the archetypical participatory mechanism. For centuries, voting was a privilege limited to few, but democratic norms now require that electoral inclusion be presumed, and exclusion justified. Accordingly, few exclusionary rules remain. Among them are citizenship, law-abidingness, and minimum age requirements. The last of these, all but ignored by legal and political theorists, is this article’s focus.

The age of electoral majority has declined, over time and across the globe. At the beginning of the twentieth century, the average voting age worldwide was just under twenty-four; today, it is just over seventeen. More than a dozen nations have recently lowered local, state, or national voting ages to sixteen.

† Associate Professor of Law, William & Mary School of Law, J.D. Harvard Law School, B.A. Yale College. I am grateful to Ludvig Beckman, Emily Buss, Naomi Cahn, June Carbone, Neal Devins, James Dwyer, Catherine Ross, and Timothy Zick for their generous comments on earlier drafts. Michael Green and Eric Kades provided thoughtful guidance in the project’s earliest stages. I also thank Research Librarian Fred Dingledy and Research Assistants Jana Robinson and Rebecca Sandler for their excellent research assistance. Finally, the editors of the Brooklyn Law Review provided sophisticated and insightful substantive feedback and have been a true pleasure to work with.

Others, including Australia and the United Kingdom, are seriously considering doing the same. The voting age in every U.S. state is eighteen, but the United States is not among the growing number of democracies deliberating the electoral inclusion of some cohort of their younger citizens. It should be.

Presumably, eighteen is a proxy for voters’ attainment of desirable characteristics—e.g., maturity of judgment, knowledge of civics, and understanding of political processes. Yet there has been no sustained scholarly effort to examine whether age eighteen is a good, or even good-enough, indicator of the attainment of those or other relevant characteristics. Academic inattention persists despite widespread acceptance that the franchise is the core of modern representative democracy; its “free and unimpaired [exercise] preserv[es] . . . other basic civil and political rights.”

I argue that presumptive electoral inclusion places on the state the burden of justifying the exclusion of a category of persons. Assessing the legitimacy of any exclusion requires a minimum standard for electoral inclusion. That standard legitimately includes competence. Assessing competence likewise first requires a conception of it, but none currently exists. Classic democratic theory describes the decision making of the ideal citizen-voter as both well informed and rational. The decision making of the actual citizen-voter, however, is often neither. The classic account thus cannot define competence, which contemplates a minimum standard of adequacy, not an aspirational ideal rarely attained. I thus argue for a conception of electoral competence first, informed by behavioral decision theory and studies of voter decision making, and second, characterized by the reliable attainment of the relevant cognitive processes (cognition/learning, information processing, and decision making) and maturity of judgment.

Converging research from several disciplines within the developmental sciences has established a reliable connection

---

2 Stefan Olsson, Children’s Suffrage: A Critique of the Importance of Voters’ Knowledge for the Well-Being of Democracy, 16 INT’L J. CHILD. RTS. 55, 55 (2008) (“That children should not have the right to vote is something that most people think . . . is so obvious that almost none of the prominent democratic theorists have given it any serious consideration. It is a non-issue.”).

3 Kramer v. Union Free Sch. Dist. No. 15, 395 U.S. 621, 626 (1969) (“Since the right to exercise the franchise in a free and unimpaired manner is preservative of other basic civil and political rights, any alleged infringement of the right of citizens to vote must be carefully and meticulously scrutinized.”); see also Reynolds v. Sims, 377 U.S. 553, 555 (1964) (“The right to vote freely for the candidate of one’s choice is of the essence of a democratic society, and any restrictions on that right strike at the heart of representative government.”).
between age range and the attainment of certain cognitive competencies. Research in developmental psychology and cognitive and social neuroscience explains not only that adolescents make notoriously bad decisions under certain conditions, but also why it is they do so. This research explains that by midadolescence, when making unpressured, considered decisions—like those required to privately cast a ballot in an election that has unfolded over time—their cognitive competencies are mature.

States can thus no longer justify the electoral exclusion of midadolescents by claiming that they lack the relevant competencies. Absent other legitimate bases for their exclusion, the democratic presumption of inclusion obliges the states to adjust downward the age of electoral majority.

In the United States, the individual states retain broad power to establish electoral qualifications, subject to certain constitutional and other federal law constraints. The Twenty-Sixth Amendment, for example, prohibits states from setting the age of electoral majority above eighteen. No constitutional or other federal law provision, however, prohibits states from lowering the age of electoral majority; each state retains that power. Yet other than a few states that allow seventeen-year-olds to vote in a primary election so long as they will turn eighteen in time for the general election, no state has

---

4 Lassiter v. Northampton Cnty. Bd. of Elections, 360 U.S. 45, 50-51 (1959) (“The States have long been held to have broad powers to determine the conditions under which the right of suffrage may be exercised absent of course the discrimination which the Constitution condemns.” (citations omitted)). The U.S. Constitution gives state legislatures the authority to establish “[t]he Times, Places and Manner of holding Elections for [U.S.] Senators and Representatives,” but it reserves to Congress the right to “at any time by Law make or alter such Regulations, except as to the Place of chusing Senators.” U.S. CONST. art. I, § 4, cl. 1; Harper v. Va. Bd. of Elections, 383 U.S. 663, 665 (1966) (recognizing the franchise as a conditional fundamental right by providing that “once the franchise is granted to the electorate, lines may not be drawn which are inconsistent with the Equal Protection Clause of the Fourteenth Amendment”); id. at 670 (holding that state laws restricting individuals’ right to vote will be subject to the Court’s strict scrutiny).

5 U.S. CONST. amend. XXVI, § 1 (“The right of citizens of the United States, who are eighteen years of age or older, to vote shall not be denied or abridged by the United States or by any State on account of age.”). 

6 U.S. CONST. art. I, § 2 & amend. XVII (providing that in statewide elections for congressional representatives and senators, “the Electors in each State shall have the Qualifications requisite for Electors of the most numerous Branch of the State Legislature”); Lassiter, 360 U.S. at 51 (“[W]hile the right of suffrage is established and guaranteed by the Constitution, it is subject to the imposition of state standards which are not discriminatory and which do not contravene any restriction that Congress, acting pursuant to its constitutional powers, has imposed.” (citations omitted)).

7 See infra Part I.C.4.
exercised—nor seriously considered exercising—its power to lower the voting age.

The Twenty-Sixth Amendment’s exclusion of individuals under eighteen from explicit constitutional protection undermines constitutionally grounded arguments for lowering the voting age. Any court that held the enfranchisement of citizens younger than eighteen constitutionally compelled would thus depart from well over a century of established precedent. Constitutional compulsion, however, is not the sole—or even primary—justification for most state action.

I argue here that the core democratic principle of inclusion (embraced by democracy theorists and political entities, and from which derives the concept of “universal . . . suffrage”) places on states the burden of justifying electoral exclusions. Moreover, in light of decades of research on voter decision making and significant advances across various scientific disciplines in knowledge of cognitive and psychological development, the time has come to revisit and if warranted, readjust, the age of electoral majority.

We can surmise the reasoning of one who believes that the voting age merits little attention: (a) age is a reasonable—
perhaps the only reasonable—proxy for the development-related attainment of the capacities required for competent voting; (b) the setting of any voting age will inevitably involve some slippage (i.e., will exclude some individuals below the set voting age who will have nonetheless attained voting competence and include others who will not have attained such competence despite having reached the set voting age), but that is the nature of proxies, and of bright-line legal rules more generally; and (c) because age eighteen is a common proxy for legal competence generally, and there appears to be community consensus for this notion, eighteen seems to be as good a proxy for electoral competence as any.

This reasoning, while superficially plausible, suffers serious flaws that this article exposes and corrects. Here, I briefly highlight those flaws, as well as the core elements of my argument.

First, this reasoning ignores the presumption of electoral inclusion to which persons subject to a democratic government’s authority are entitled. Presumptive inclusion is a broadly accepted normative commitment flowing from basic principles of democratic theory. While the presumption does not foreclose the possibility of legitimate exclusions, it does shift to the state the burden of justifying electoral exclusion.

Second, assessing whether a state has met its justificatory burden requires some principled criteria or of the demos that governs the state. An eight-year-old child can hardly be enlightened enough to participate equally with adults in deciding on laws to be enforced by the government of the state,” but failing to discuss older children or adolescents). Or, second, despite minors’ incompetence, their interests should be registered through the use of some proxy voting method. See, e.g., Olsson, supra note 2, at 55 (arguing that parents should be designated their children’s representatives for purposes of voting); Jane Rutherford, One Child, One Vote: Proxies for Parents, 82 MINN. L. REV. 1463, 1502 (1998) (proposing that children’s interests can be registered by “creating proxies, so that their interests and voting power are expressed through others”). Two scholars, legal theorist David Archard and political theorist Francis Schrag, do consider age-related differences among those younger than eighteen, although each author only briefly addresses the issue. See David Archard, Children: Rights and Childhood 103 (2d ed. 2004) (discussed infra note 295); Francis Schrag, Children and Democracy: Theory and Policy, 3 POL., PHIL. & ECON. 368-70 (2004) (discussed infra note 162).

11 See Paul Arshagouni, “But I’m an Adult Now . . . Sort of”: Adolescent Consent in Health Care Decision-Making and the Adolescent Brain, 9 J. HEALTH CARE L. & POL’Y 315, 333 n.110 (2006) (“Wherever one draws a bright line, there will inevitably be a certain number of individuals who in truth should have fallen on the other side of the line. Epidemiologically, these are known as false positives and false negatives. The greater the number of false positives and false negatives, the less usefulness we have for a given bright line rule.”).

12 In other words, some individuals younger than eighteen will have attained the relevant competence yet be denied the franchise, and other individuals who have reached eighteen but failed to attain the relevant competence will nonetheless be extended it.

13 See infra Part II.A.
standards for inclusion. Democracy theory and liberal constitutional principles have long supported two such criteria: ongoing interest in and connection to the political community and electoral decision-making competence.14

Because no principled conception of electoral competence exists, I develop a concept of it here, informed by behavioral-decision research on voter decision making and by research from various disciplines in the developmental sciences on adolescent cognitive development. I argue for a definition of electoral competence as *the attainment and application of adultlike cognitive-processing capacities in the electoral context*. These capacities include the abilities to acquire information and knowledge, to assess and process information, and to make and justify a decision.15

And third, although age eighteen may have been the best available proxy for electoral competence when it became the national voting age in 1972, research in the developmental sciences in the intervening years lays the groundwork for a better-informed assessment of the attainment of voting competence. I survey research in behavioral and developmental psychology and in cognitive and social neuroscience. This research has expanded our understanding of the development of a full range of cognitive capacities. It establishes that adolescents reliably reach adultlike cognitive-processing capacities by ages fifteen or sixteen, but that numerous factors (e.g., situations involving high levels of emotion or stress, peer pressure, or time pressure) will predictably compromise their cognitive performance. Adolescent cognitive-processing competence is thus domain- or context-specific.16

While scientific research cannot dictate policy, it can inform policy. Armed with more nuanced understandings of both voter decision making and the development of adolescent cognitive processing and decision making, I conclude that voting is the sort of decision-making context in which midadolescents will reliably demonstrate competence.17 In light of midadolescent electoral competence, states fail to meet their justificatory burden in the absence of other reasons for continued midadolescent electoral exclusion.

---

14 *See infra* Part II.B.
15 *See infra* Part III.A.
16 *See infra* Part III.
17 *See infra* Part III.B.
The article proceeds in three parts. Part I situates the U.S. voting age within a conceptual, historical, and modern global context. It first explains why the voting age in the United States was originally set at twenty-one. It goes on to document the lowering of the voting age to eighteen in the mid-twentieth century. Finally, it describes and explains the growing global trend of lowering the voting age to sixteen. Part II identifies and justifies basic voting criteria. It argues that democratic principles require members of a political community to be presumptively entitled to political/electoral participation. Members may nonetheless be excluded from participating if they fail to possess certain characteristics—namely, ongoing interest in and connection to the relevant community, and electoral competence. Part III argues for a conception of electoral competence informed by political- and behavioral-decision theorists’ understanding of voter decision making, and psychologists’ understanding of the cognitive processes required to competently make decisions in the electoral context. It then demonstrates that the converging research of developmental scientists in several disciplines provides new evidence of the age-related attainment of relevant cognitive processes.

This article concludes that there is strong empirical evidence that the cognitive processes required for competent voting reliably mature by age sixteen. A reexamination of the voting age is necessary to account for the evolution of our understanding of electoral competence and its achievement. Only then can the modern democratic state ensure that the continued disfranchisement of a category of citizens remains consistent with its foundational political commitments.

I. CONCEPTUAL, HISTORICAL, AND MODERN GLOBAL CONTEXT

This part discusses the political and historical forces that initially set the U.S. voting age at twenty-one, the age of legal maturity in the English common-law tradition. The voting age remained unchanged in the United States for more than two centuries. Concerted efforts to lower it began in 1942, when Congress lowered the age of conscription from twenty-one to eighteen. Those efforts bore fruit in 1971, when a war that was especially unpopular with young people sparked a nationwide movement to lower the national voting age to eighteen. In the four decades since, the voting age has received
essentially no attention domestically. It has, however, become an increasingly visible issue abroad, and a growing number of countries have lowered their voting ages from eighteen to sixteen or are considering doing so. This part concludes by surveying the contemporary global context and examining the factors that are driving this downward trend.

A. England and the British Empire Through the Eighteenth Century

The primary historical influence on early American voting rules was, unsurprisingly, the English common-law tradition. Arguments made during a series of important debates in mid-seventeenth-century England, both in favor of and against widespread suffrage, have echoed at voting rights debates ever since; they provide a conceptual framework for the discussion of democratic inclusion that follows in Part II.

The age of majority has fluctuated throughout history. Under Roman law, the age of majority was fourteen for males, and twelve for females.18 The law presumed that by age fourteen, males would have attained the intellectual capacities necessary to exercise full citizenship, which required “understanding and judgment as to acts in law, in particular in relation to property rights.”19 In France, Germany, and throughout the northern parts of Europe between the ninth and eleventh centuries, the age of majority for males was fifteen.20 Though nearly identical to the Roman age of majority, the requisite capacities in northern Europe that signaled legal maturity were not intellectual but instead physical—namely, the physical ability to participate in warfare.21

The age of majority for English knights, who fought on behalf of the crown, increased during the Middle Ages.22 Legal historians attribute the increased age requirement to the changing characteristics of war making. Armies increasingly included mounted cavalry, which required knights skilled in the use of horses in battle—skills that required a lengthy training period to develop. Along the same lines, improvements

19 Id. at 25.
20 Id.
21 Id. ("The test applied in selecting this age seems to have been different from that applied in Rome[—]namely, the capacity to bear arms.").
22 James observes that there is no "clear authority" that the English age of majority in the ninth and tenth centuries was fifteen, but concludes that it is a reasonable assumption. Id. at 26.
in defensive armor also increased the armor’s weight, requiring additional strength by those wearing it. Young men thus became eligible for knighthood at twenty-one, because not until then would they have completed the training and acquired the strength and endurance required of the armored warriors who fought in the heavy cavalry.

English suffrage originated in 1215, when English barons forced King John’s accession to the Magna Carta. Initially the exclusive privilege of the English nobility, the franchise gradually expanded to other property owners. Early suffrage provisions imposed only residence and property (“freehold”) qualifications. Nothing explicitly restricted the franchise to males, or to people of a certain age; these restrictions were so sufficiently obvious that they remained unstated well into the nineteenth century.

English historical and common law traditions eventually became law throughout the British Commonwealth, and indeed in much of the Western world. However arbitrary its genesis may seem in retrospect, age twenty-one remained firmly entrenched as the age of legal and electoral majority for centuries in England, as well as in the nations across the globe that incorporated English traditions.

---

23 Id. at 22-23, 30; see also William Arthur Shaw, The Knights of England: A Complete Record from the Earliest Time to the Present Day of the Knights of All the Orders of Chivalry in England, Scotland, and Ireland, and of Knights Bachelor, Incorporating a Complete List of Knights Bachelor Dubbed in Ireland Vol. I ii-iii (1906), available at http://openlibrary.org/books/OL7047747M/The_Knights_of_England. For “tenants in socage” (generally farmers who held land of feudal lords, to whom they owed rent), however, the age of majority remained fifteen (or sometimes fourteen). James, supra note 18, at 30.

24 James, supra note 18, at 28.


26 Beginning in late fourteenth-century England, “franchise” referred to a political privilege or entitlement granted at the will of the governing body, who was the ultimate sovereign—“a special privilege or exclusive right to perform some public function, granted by a sovereign power to any person or body of people.” Oxford Eng. Dictionary (2d ed. draft rev. 2010), available at http://www.oed.com/ (search “franchise”). A well-known 1430 English statute extended to “forty-shilling freeholders” the right to elect members to parliament’s House of Commons. Donald Grier Stephenson, Jr., The Right to Vote: Rights and Liberties Under the Law 35-36 (2004); Chilton Williamson, American Suffrage from Property to Democracy 1760-1860, at 5-6 (1960).

27 The Reform Act of 1832 was the first English voter qualification statute that explicitly specified that the franchise extended only to “male person[s] of full age” who met other qualifications. Wendell W. Cultice, Youth’s Battle for the Ballot: A History of Voting Age in America 72 (1992).

28 Id.

29 Id. at 2.

30 James, supra note 18, at 22, 33. James characterizes the common law age of majority of twenty-one as “a curious development from the older systems requiring
1. Concepts of Electoral Inclusion: The Putney Debates

The early franchise extended only to a tiny fraction of Englishmen. Following the end of the English Civil War in the mid-seventeenth century, soldiers who had fought in Oliver Cromwell’s victorious parliamentary army joined with political activists known as Levelers to demand equal (“level”) political rights. Among their then-radical demands were the elimination of the freehold requirement and the adoption of near-universal male suffrage. The group submitted a proposed “Agreement of the People” to Cromwell and other parliamentary officers, who then met with them in 1647 to discuss the proposal. The men held a series of meetings in Putney, near London, and their discussions were memorialized as what became known as the Putney Debates.

The debates frame modern concepts of political inclusion and justifiable limits to inclusion. The concerns underlying the moderates’ arguments for limited suffrage have helped justify and define modern boundaries of democratic inclusion; the Levelers’ arguments in favor of widespread suffrage echo, too, as the standard justifications for presumptive democratic inclusion and universal suffrage.

The more moderate spokesmen who defended narrow property-based suffrage made two basic arguments. First, because voters’ choices eventually shaped the laws that would bind everyone in the community, it was important for voters themselves to have “a permanent fixed interest in th[e]
Kingdom...[and to] comprehend the local interest...". A property-ownership requirement guaranteed that would-be voters possessed these characteristics. Property gave a man a personal stake in and knowledge of the community. The relatively permanent and ongoing nature of property ownership meant that the freeholder would himself be affected by and subject to the community’s laws, both current and future. The freeholder’s interests were thus linked to, if not identical with, those of the community, and he could be trusted by his fellow citizens to vote in a manner reasonably consistent with its interests—distinguishing him from transients who, lacking the rootedness that came with property ownership, could be “here today, and gone tomorrow.”

Second, only “men freed from dependence upon others” could be trusted to vote. The landless, servants, and women would too easily be influenced or manipulated by those they were economically dependent on. Economic independence alone could guarantee that persons’ votes would reflect intellectual independence.

The radicals who argued in favor of universal suffrage, on the other hand, tapped into the antimonarchical and growing egalitarian sentiments of the post-Civil War, pre-Enlightenment years. They argued that all men equally possessed certain “natural rights” by virtue of their humanity—rights neither derived from, nor dependent on, the government or property ownership. “Has not the meanest He,”

35 STEPHENSON, supra note 26, at 36-37 (quoting Statement of Commissary General Henry Ireton, in 1 THE CLARKE PAPERS, NEW SERIES XLIX, 299-303 (C.H. Firth ed., 1891)). Ireton was Oliver Cromwell’s son-in-law and a senior military officer in the parliamentary army who led a moderate faction within the military. He was the primary author of the Declaration of the Army, which made more modest demands of Parliament. Id. at 34.


37 WILLIAMSON, supra note 26, at 64 (citing Statement of Commissary General Henry Ireton, in PURITANISM AND LIBERTY: BEING THE ARMY DEBATES (1647-9) FROM THE CLARKE MANUSCRIPTS WITH SUPPLEMENTARY DOCUMENTS 58 (A.S.P. Woodhouse, ed., 2d ed. 1951)).

38 KEYSSAR, supra note 1, at 5 (quoting Statement of Commissary General Henry Ireton, in PURITANISM AND LIBERTY, supra note 37, at 82).

39 STEPHENSON, supra note 26, at 37 (quoting Statement of Colonel Rich, in THE CLARKE PAPERS, supra note 35, at 315) (arguing that in the Roman Republic, “the people’s voices were bought and sold, ... and thence it came that he that was the richest man ... made himself a perpetual dictator”). At a time when much voting was conducted viva voce—not by secret ballot—the economically vulnerable might understandably hesitate to cast a vote that would displease their economic superiors. Id. at 38-39, 47. At the time of the U.S. founding, voting in some locales, particularly in the South, continued to be an oral and public act. KEYSSAR, supra note 1, at 23-24.

40 KEYSSAR, supra note 1, at 38-39.

41 STEPHENSON, supra note 26, at 40; WILLIAMSON, supra note 26, at 62-67.
asked one of their leading spokesmen pointedly, “as much a life to live as the greatest He?" Men, regardless of station or education, were equally endowed with “human reason.” In light of their fundamental equality, no man could legitimately be made subject to the will of another without giving his consent, nor could a man legitimately be made subject to the laws of a government, unless he “first by his own consent . . . put himself under that Government.” Those “bound by laws in which they have no voice at all” are a people “enslave[d].”

The arguments made by the radical Levelers and their moderate counterparts were not original; indeed, many of the ideas that swirled at Putney could be traced to ancient thinkers. The considerable influence of the Putney Debates on American political thought more likely derived from two factors: first, the timing of the debates resulted in their dissemination throughout the American colonies in the decades preceding the American Revolution, as colonists increasingly chafed at their political inequality with respect to their mother country; second, the debates’ emergence out of Britain itself magnified their importance in the American colonies, given that colonists identified with the British or as British. The intellectual maelstrom of the period immediately following the English Civil War thus became the more “immediate origin[]” of American political thought and was “more instructive and influential in understanding” later American political developments than more historically distant antecedents.

2. The American Colonies

The rhetoric at Putney had enduring influence but little immediate effect, either in England or in the American colonies. Instead, English law retained property and income qualifications for nearly three more centuries. The English

42 Stephenson, supra note 26, at 37-38 (quoting Statement of Colonel Thomas Rainborough, in The Clarke Papers, supra note 35, at 300-01, 304); Williamson, supra note 26, at 64 (quoting Statement of Colonel Thomas Rainborough, in Puritanism and Liberty, supra note 37, at 53).
43 Williamson, supra note 26, at 65.
45 Id.
46 Id. at 33.
47 Id.
48 England expanded the franchise to include all men twenty-one and older who met a six-month residence requirement in the Reform Act of 1918. Cultice, supra note 27, at 72.
age of legal and electoral majority would remain unchanged even longer—well into the twentieth century.

When England established the thirteen American colonies in the seventeenth and eighteenth centuries, however, it imposed on them no uniform voting rule. Each colonial assembly thus enacted voter qualification rules that would govern elections within its respective territory. Although the colonies’ rules varied, they all adopted property-based electoral systems that reflected prevailing British practice. They also retained twenty-one years as the near-universal age of electoral majority. An occasional and singular exception to general voter requirements was militia service, and some colonies occasionally enfranchised militiamen younger than twenty-one. This exception was in keeping with longstanding opinion (whose endurance seems to come more from its emotional appeal than its logical integrity) that those who risked their lives defending their country earned a voice in its governance.

B. United States, Through the Twentieth Century

This section traces the political history of the U.S. voting age from the founding through the twentieth century. American revolutionaries echoed the Levelers’ claim that political membership was a birthright, and that voting was the fundamental political act. It was an ideology that made citizenship and suffrage inseparable. Just as the Levelers’

49 Stephenson, supra note 26, at 41-47.
50 Id. at 41-45. “[W]ithin some colonies, some cities possessed charters issued by the Crown.” Id. at 42. In these cities, royal decree, not the colonial assembly, established the rules governing the franchise. Id.
51 Id. at 41.
52 Cultice, supra note 27, at 2. Several colonies adopted rules allowing men who failed to meet certain qualifications to vote upon reaching an age higher than twenty-one. In Massachusetts and New Hampshire, for example, nonfreemen and those not members of the church could vote upon reaching age twenty-four. Id. at 4.
53 Id. at 2-3. The Virginia House of Burgesses in 1619, for example, pronounced that every male over sixteen was to serve in the militia, pay taxes, and vote. Id. at 3.
54 Akhil Reed Amar, America’s Constitution: A Biography 19 (2005); Amar noted that “[i]n classic republican theory, the rights of collective self-government stood shoulder to shoulder with the responsibilities of collective self-defense.” Id.; see also Keyssar, supra note 1, at 36. More pragmatically, men denied the franchise might balk at the call to military service. Id. at 12-13; Pamela S. Karlan, Ballots and Bullets: The Exceptional History of the Right to Vote, 71 U. CIN. L. REV. 1345, 1348 (2003).
55 Judith N. Shklar, American Citizenship: The Quest for Inclusion 45 (1991) (arguing that the Putney “debates have a permanent significance, especially for American political thought,” and that “[t]he future American citizen was born in the course of these exchanges”).
ideas echoed in support of expanding the franchise, however, the opponents of widespread suffrage at Putney remained equally important to the political evolution of the nation. Political theorist Judith Shklar has noted that the moderates’ “arguments were repeated over and over again whenever yet another group of Americans demanded the right to vote.”56

The Constitution submitted to the states for ratification contained no uniform national suffrage law.57 The Framers’ omission was intentional: not only were they themselves ideologically divided, but they also feared that any national suffrage law they proposed would generate sufficient controversy to derail ratification altogether.58 The states, as had the colonies before them, thus retained the power to determine voter-qualification standards for both state and national elections,59 and all retained twenty-one as the age of electoral majority.60

The Fourteenth Amendment, ratified after the Civil War in 1868, formally extended the benefits of citizenship to African Americans, and its Reduction-of-Representation Clause sought to secure their enfranchisement, albeit indirectly.61 It warned that a state disfranchising “any of the male inhabitants . . . being twenty-one years of age, and citizens of the United States, . . . except for

56 Id.
57 KEYSSAR, supra note 1, at 18.
58 Id. at 18-20. Alexander Hamilton explained,

To have reduced the different qualifications in the different States to one uniform rule, would probably have been as dissatisfactory to some of the States, as it would have been difficult to the Convention. The provision made by the Convention appears, therefore, to be the best that lay within their option. It must be satisfactory to every State; because it is conformable to the standard already established, or which may be established by the State itself.

59 The Framers sought, however, to make the House of Representatives the federal legislative body most responsive to, and representative of, the common citizen. Article I thus provides that “the People of the several States” who meet “the Qualifications requisite for Electors of the most numerous Branch of the State Legislature” shall elect the members of the House. U.S. CONST. art. I, § 2. In this way, the Framers ensured that the most liberal electoral standard adopted by a given state would apply to state voters’ selection of their House delegates.

The Constitution also imposes age qualifications for various federal offices: twenty-five-years-old for the House, thirty for the Senate, and thirty-five for the presidency. Id. art. I, § 3 & art. II, § 1. Tench Coxe argued that the Constitution’s age requirements obligated the wealthy or politically connected individual, otherwise able to ascend to federal office at an early age, to first gain necessary experience and also to demonstrate “his merits to his country—a more rational ground of preference surely than mere property.” Tench Coxe, An Examination of the Constitution for the United States (II), in PAMPHLETS ON THE CONSTITUTION OF THE UNITED STATES 141 (Paul Leicester Ford ed., Da Capo ed. 1888) (1787-88).

60 CULTICE, supra note 27, at 12.
participation in rebellion, or other crime,” would lose representation in Congress and in the electoral college. The clause was interpreted as establishing something of a national voter-qualification norm, divesting the states of the power to disfranchise twenty-one-year-old male law-abiding citizens.

The voting age received no national attention to speak of until the United States entered World War II at the end of 1941. Soon after declaring war, Congress began to debate amending the Selective Service and Training Act to lower the draft age from twenty-one to eighteen. As they debated that measure, legislators made another proposal—a constitutional amendment to lower the national voting age, also to eighteen. Republican Arthur Vandenberg introduced the measure in the Senate, invoking again the idea that “if young men are to be drafted at 18 years of age to fight for their Government, they ought to be entitled to vote at eighteen years of age for the kind

---


63 Minor v. Happersett, 88 U.S. (21 Wall.) 162, 174-75 (1874) (relying on the clause’s explicit mention of “male” citizens to support its conclusion that the Fourteenth Amendment did not require extending the franchise to women).

Two years after ratifying the Fourteenth Amendment, the states ratified the Fifteenth, which sought to prevent the disfranchisement of newly freed African American men. The amendment states that “[t]he right of citizens of the United States to vote shall not be denied or abridged by the United States or by any State on account of race, color, or previous condition of servitude.” U.S. CONST. amend. XV, § 1. It was hugely, but only briefly, successful. Nearly a million African American men voted, but once Union occupation of the South ended, southern states used a variety of legal and extralegal practices to again disfranchise African Americans. A whole range of barriers, both formal and informal, continued to limit the political participation of African Americans and others formally entitled to vote. Social and economic barriers such as poverty and illiteracy, as well as seemingly innocuous aspects of election administration (e.g., preregistration procedures, identification requirements at polling stations, choice of election day and hours) have presented, and continue to present, hurdles to political participation. DENNIS F. THOMPSON, JUST ELECTIONS 28 (2002).

Over the following decades, activists sought both to enfranchise women and reenfranchise African American men. The passage of the Nineteenth Amendment in 1920 accomplished the former; U.S. CONST. amend. XIX; years of litigation and the passage of the Voting Rights Act of 1965 went a long way toward accomplishing the latter. THOMPSON, supra.

64 CULTICE, supra note 27, at 7, 13-14. The first “serious” consideration given to a proposal to lower a statewide voting age from twenty-one to eighteen probably occurred at the Missouri Convention in 1820. Id. at 7. After the end of the Civil War, delegates to the New York Constitutional Convention of 1867 considered a similar proposal; both conventions rejected the proposals. Id. at 7, 13-14.

65 Republican Senator Arthur Vandenberg, from Michigan, and Democratic Congressman Jennings Randolph, from West Virginia, introduced joint resolutions proposing the constitutional amendment. S.J. Res. 166, 77th Cong. (1942); H.R.J. Res. 354, 77th Cong. (1942).
of government for which they are best satisfied to fight." Congress lowered the draft age to eighteen in the fall of 1942 but adjourned without taking action on the proposed amendment.

Between 1942 and 1944, members of Congress introduced more than a half-dozen similar joint resolutions to lower the national voting age; in the states, lawmakers began doing the same. Georgia became the first state to lower its statewide voting age to eighteen, amending its constitution in 1943, just one year after eighteen-year-olds became eligible for the draft. Between 1945 and 1952, state lawmakers introduced nearly 100 bills in their legislatures proposing reductions in the voting age, and federal lawmakers sent more than a dozen bills to congressional committees. It was more than ten years after Georgia’s constitutional amendment, however, before any other proposals would succeed. Then, in 1955, Kentucky became the second state to lower its voting age. When new states Alaska and Hawaii adopted their constitutions soon thereafter, they adopted compromise voting ages of nineteen and twenty, respectively.

Unlike earlier movements to enfranchise African American and women citizens, the disfranchised—young people themselves—were not at the forefront of the youth-vote movement. The nation’s youth did not begin to mobilize until the early 1960s, when the nation’s involvement in the Vietnam War galvanized their efforts. Student organizations on college

---

67 CULTICE, supra note 27, at 22.
68 Id. at 20, 22.
69 Id. at 24-25.
70 Id. at 25-26.
71 Id. at 30.
72 Id. at 55. In 1954, President Dwight Eisenhower, a former military general, became the first U.S. president to endorse a constitutional amendment extending the vote in federal elections to eighteen-year-olds. Eisenhower’s reasoning was to the point: “[I]f a man is old enough to fight he is old enough to vote . . . .” KEYSSAR, supra note 1, at 225.
73 CULTICE, supra note 27, at 59-60.
74 See Lowering the Voting Age to 18; Hearings Before the Subcomm. on Constitutional Amendments of the Comm. on the Judiciary, 90th Cong. 23 (1968) (statement of Rep. Spencer Oliver). As of 1968, approximately 25 percent of U.S. troops, and nearly 30 percent of U.S. casualties, had been younger than twenty-one years old. Id. In 1968, students founded Let Us Vote (LUV) on the campus of the University of the Pacific in Stockton, California. Within months, the organization expanded to include chapters at more than three thousand high schools and four hundred colleges across the country. CULTICE, supra note 27, at 97-98. LUV joined the Youth Franchise Coalition, comprising twenty-three civil rights and educational organizations working to extend voting rights, both at the state level and through a federal constitutional amendment. Id. at 98.
campuses and in high schools then joined various civil-rights organizations already working to extend voting rights to eighteen-year-olds.75

In an effort to bypass the cumbersome process of amending the U.S. Constitution, Congress in 1970 moved to lower the nationwide voting age through federal legislation.76 Lawmakers, before voting to extend the Voting Rights Act, which was set to expire, inserted into the Act a provision lowering the voting age to eighteen in both federal and state elections. President Richard Nixon, who explained that he supported a lower voting age not because eighteen-year-olds were old enough to fight, but because “they were smart enough to vote,” signed the amended Act into law.77 As he did so, however, Nixon expressed skepticism that Congress’s power extended beyond regulating federal elections to also include setting voter qualifications for statewide elections.78

Nixon’s skepticism was well-founded, and four states immediately challenged the law.79 The case Oregon v. Mitchell split the Supreme Court.80 Four justices would have held that Congress was without power to set voter qualifications for either federal or state elections.81 Four justices would have held that Congress had power to set voter qualifications for both federal and state elections.82 Justice Hugo Black, the swing vote, believed that Congress had the power to set voter qualifications in federal—but not state—elections.83 His plurality opinion, upholding the provision in part and invalidating it in part, became the judgment of the Court.84

---

75 CULTICE, supra note 27, at 99.
76 The amendment extended the Voting Rights Act and, in addition to its provision lowering the voting age, abolished residency requirements over thirty days for national elections and prohibited literacy tests for a period of five years. Voting Rights Act Amendments of 1970, Pub. L. No. 91-285, 84 Stat. 314.
77 CULTICE, supra note 27, at 115.
78 The Voting Rights Act was set to expire if Congress failed to pass the amendments extending it. Unwilling to allow that to happen, Nixon signed the amendment, indicating that he was “leaving the decision on the disputed provision to what I hope will be a swift resolution by the courts.” Id. at 139.
80 The justices filed five separate opinions in the case. Id. at 115.
81 They were Justices Blackmun, Harlan, and Stewart, and Chief Justice Burger. See id. at 152 (Harlan, J., concurring in part and dissenting in part); id. at 281 (Stewart, J., concurring in part and dissenting in part).
82 They were Justices Brennan, Douglas, Marshall, and White. See id. at 229 (Brennan, J., concurring in part and dissenting in part); id. at 135 (Douglas, J., concurring in part and dissenting in part).
83 Id. at 117.
84 Id.; see also CULTICE, supra note 27, at 172-73.
Oregon v. Mitchell thus left intact the provision in the Voting Rights Act lowering the voting age to eighteen in federal elections, but by invalidating the same provision’s application to the states, returned the voting ages for statewide elections to their respective pre-1970 status quo. The forty-seven states that did not permit eighteen-year-olds to vote scrambled either to modify their voting systems to accommodate two voting lists—one for federal elections and another for state elections—or to change their laws to lower the voting age in both federal and state elections. Modifying state voting systems presented significant administrative challenges and promised to be expensive. Lowering the voting age statewide required amending state constitutions, and in every state but Delaware, constitutional amendments required voter referenda. Because of the requirements of their constitutional amendment processes, only fifteen states could possibly have changed their statewide voting ages in time for the November 1972 elections.

Spurred by the prospect of election-day chaos, in March 1971 Congress hurriedly approved the proposed Twenty-Sixth Amendment. The amendment provided that the right of citizens “eighteen years of age or older, to vote shall not be denied or abridged by the United States or by any State on account of age.” The state legislatures all met in session or special session to ratify the amendment, and they did so in record time. The Twenty-Sixth Amendment became law—in time for the 1972 elections—when Ohio became the thirty-eighth state to ratify it in June 1971.

Significantly, World War II did not spark the voting-age debate in the United States alone: it ignited a “global suffrage age reduction movement.” By the early 1970s, the world was nearly evenly divided; sixty-nine countries had voting ages of twenty-one or older, while sixty-eight countries had adopted

---

85 CULTICE, supra note 27, at 180-81.
86 Id.
87 Id. at 185.
88 Id.
89 Id. at 191.
90 U.S. CONST. amend. XXVI, § 1.
91 CULTICE, supra note 27, at 191.
92 After lowering the age of conscription, U.S. allies also confronted pressures to enfranchise their youngest soldiers; several did so, at least temporarily. Id. at 78. During both world wars, for example, Britain, Canada, and several other Commonwealth members lowered the wartime voting age from twenty-one to nineteen, restoring their voting ages to twenty-one after the wars. Id. at 76. In 1936, Russia became the first European nation to permanently lower its voting age to eighteen. Id.
voting ages of less than twenty-one (with fifty-two of those adopting eighteen as the national voting age).93

C. Global Context, into the Twenty-First Century

This section surveys the continuing expansion of the franchise to younger citizens in other countries, and the status of the voting age in the United States. More than a dozen countries have lowered local, state, or national voting ages to sixteen, driven primarily by efforts to increase youths’ political engagement and counter the disproportionate political influence of older citizens (who vote at higher rates than the young, and whose numbers have grown as a result of demographic factors). Other nations have begun to consider doing the same. In the United States, a handful of state legislatures have considered proposals to lower statewide voting ages, but the issue has not generated widespread attention.

At the beginning of the twentieth century, the average age of the electoral majority worldwide was just under twenty-four years.94 Approximately half of Europe’s nations (including Britain, France, and Italy) set the age of electoral majority at twenty-one, and half at an older age.95

Today, most nations have adopted some form of representative democratic government.96 Over 80 percent have set the voting age at eighteen without exception.97 Twelve countries extend the franchise to sixteen-year-olds, although two do so only for those who are employed, and three only for those who are married.98 Five countries have set the voting age

93 Id. at 79. More than a dozen of the countries that retained twenty-one as the voting age allowed individuals younger than twenty-one to vote under certain circumstances—e.g., those serving in the military, those who were married, or those participating in provincial or local elections. Id. at 78-79.
94 Id. at 76.
95 Id. The original nineteenth-century constitutions of more than a dozen European nations, for example, set the age of electoral majority at twenty-five years; several other nations set it at thirty. Id. at 89.
98 Countries permitting sixteen-year-olds to vote include Austria, Bosnia and Herzegovina, Brazil, Croatia, Cuba, Dominican Republic, Ecuador, Guernsey, Indonesia, Isle of Man, Jersey, Nicaragua, and Slovenia. Id. Bosnia and Herzegovina, Croatia, and Slovenia permit only employed sixteen-year-olds to vote; otherwise, the
The average voting age worldwide is between seventeen and eighteen years. In Western democracies and the Commonwealth nations, countries that have lowered or are considering lowering the voting age are doing so to counteract the aging of their electorates and to increase more generally the political participation of young people. These continue to be the nations with which the United States is most closely aligned politically, and I discuss them at greater length below.

1. Canada and Australia

Canadian political parties all allow members as young as fourteen to vote for the parties’ candidates for Canadian prime minister. The national voting age for all other elections, however, is eighteen.

Over the last decade, the Canadian Parliament has considered numerous proposals to lower the national voting age to sixteen, the most recent coming before the House of Commons in 2005. The bill’s supporters advanced their proposal as an instrumental measure that would reverse declining voter participation and reengage youth in the nation’s politics. Members of Parliament from each of the four federal

voting age is eighteen. Id.; Danish Youth Council, DUF Factsheet: The Suffrage Age— in Europe 2 (Jan. 2011), http://duf.dk/uploads/tx_templavoila/Suffrage_in_Europe.pdf. The Dominican Republic and Indonesia permit married individuals regardless of age to vote; otherwise the voting age is eighteen in the Dominican Republic and seventeen in Indonesia. The World Factbook, supra note 97. Hungary allows married individuals aged sixteen or seventeen to vote. Danish Youth Council, supra, at 2. In Sweden and Finland, sixteen-year-olds may vote in parochial church council elections. Id. at 1; see also BRAZIL: A COUNTRY STUDY 300-01 (Rex A. Hudson ed., 1997), available at http://countrystudies.us/brazil/. 99 Countries permitting seventeen-years-olds to vote include Indonesia, North Korea, Seychelles, Sudan, and Timor-Leste. The World Factbook, supra note 97.

100 See generally The World Factbook, supra note 97 (listing most nations’ voting ages between seventeen and nineteen).

101 House of Commons Debates (Hansard), No. 47 (Feb. 1, 2005), at 1735 (Can.), available at http://www.parl.gc.ca/content/hoc/House/381/Debates/047/HAN047-E.PDF (statement of Private Member Mark Holland on Canada Elections Act) (“We give youth the opportunity at leadership conventions to select the leaders of our respective parties, who become prime ministers. That certainly is something that we all think is acceptable. In fact, in all of our nominations youth as young as 14 are allowed to select who their local candidate will be.”).

102 The World Factbook, supra note 97.


104 See House of Commons Debates (Hansard), No. 47, supra note 101, at 1730 (statement of Private Member Mark Holland on Canada Elections Act). Holland, who sponsored the bill, suggested that because younger people lack the vote, legislators
political parties, including the caucus leaders of each party, joined in multipartisan support of the proposal, but the bill failed to pass the House.\(^{105}\)

The Australian federal government, also with the goal of increasing voter engagement and participation, raised the possibility of lowering the national voting age to sixteen in a 2009 green paper.\(^{106}\) As of this writing, however, the proposal has not advanced further in the lawmaking process.

Like the United States, Australia is a federal democracy, and its six states retain power over state voting rules. The ignore them and their issues, which contributes to their disengagement. He argued that “by the time they get to 18, . . . they are often disengaged and they are not in a general education environment any more. Their patterns have already been established.” Id. Younger adolescents still in school could join a political party, participate in debates, and engage with legislators and candidates for office in a meaningful way. Because they still live at home (as opposed to high-school graduates who have gone to college or moved out of their parents’ home), it would be easier for them to register and actually vote. Early voting and political engagement will potentially establish a lifetime pattern; Holland claimed that “if we can get them to vote once they will vote again and again.” Id.


Victorian Electoral Commission (VEC), which manages state and local elections in the state of Victoria and also conducts electoral research, published a study of the age of electoral majority in a 2004 research paper. The VEC concluded that research conducted both in the United Kingdom and in Germany suggested that many of the arguments for excluding sixteen- and seventeen-year-olds from the franchise “might not be valid.” It then noted numerous advantages to lowering the voting age. Foremost among them, the VEC noted that “lowering the voting age to an age when people are still in school would allow more effective education programmes due to them being more relevant to students’ immediate lives,” and thus potentially “reduce voter ignorance overall.” Consistent with reported studies, the research paper suggested that participating in elections at a young age could establish lifelong participatory habits and reduce the likelihood of apathy at later ages. The VEC concluded that “democracy will be enhanced by the inclusion of additional viewpoints.”

2. Continental Europe

A growing number of Western European nations have either lowered their voting ages or are considering doing so. In general, supporters of the change view it as a policy instrument to increase youth political representation and civic engagement, counterbalance the overrepresentation of older voters among the electorate (caused by the higher turnout of older voters and the aging of the population), and improve democracy more generally.

In 2009, a group comprising members of Parliament from nine European countries proposed that the Parliamentary Assembly of the Council of Europe study whether to “lower[] . . . the voting age to sixteen in all [forty-seven] member

---

108 Id. at 10.
109 Id. at 10-11.
110 Id. at 11. The VEC reasoned that if voter involvement begins “when people are young and enthusiastic, . . . fewer young people may develop an attitude of ‘politicians don’t care about me’ which continues into later life.” Id.
111 Id. Among the initiatives advanced by a Youth Summit held in 2008 was a proposal to lower the voting age to sixteen. Susanna Dunkerley, Youth Speak: Make the Voting Age 16, ADVERTISER (Austl.), Apr. 14, 2008, at 7, available at 2008 WLNR 6923734.
The proposal expressed the signers’ concern over current low voting rates among young voters across Western Europe and observed that the aging of the population would only worsen youths’ marginalization in political processes. It argued that the trends threatened “the future stability of European democracy,” and that lowering the voting age could help reengage young people in the democratic process.113

Austria lowered its national voting age to sixteen by constitutional amendment in 2007, becoming the first European Union nation to do so.114 In large part, the change sought to counterbalance the increasing percentage of voters aged sixty-five and older, whose numbers have been growing due to declining birth rates. Because individuals under eighteen could not vote at all, and because older citizens vote at rates higher than do younger citizens, the growing demographic imbalance prompted concerns that government would become less responsive to the interests of the nation’s young people. Sixteen- and seventeen-year-olds voted for the first time in the 2008 national elections.115 Although the Austrian government does not track voter participation by age, one government-funded study found that sixteen- to eighteen-year-olds voted in the 2008 national elections at the same rate as the rest of electorate—approximately 73 percent.116

Several states in Switzerland and Germany have lowered the voting age for local elections to sixteen; nearly half of the sixteen German states have done so. In Norway, the ombudsman for children has published a report advocating


113 Id. ¶ 5.


lowering the national voting age to sixteen. As a pilot project, the Norwegian Parliament has authorized twenty municipalities to enfranchise sixteen-year-olds in municipal and county elections in 2011. Finland, too, has recently appointed a government group to study the issue.

3. The British Islands

In the British Islands, the self-governing Crown dependencies—the Isle of Man and the Bailiwicks of Guernsey and Jersey—lowered their voting ages from eighteen to sixteen in 2006, 2007, and 2008, respectively. In both Scotland and Wales, Parliamentary Assemblies voted in 2008 in favor of lowering the voting age to sixteen. Neither is currently able to implement the change, however, since the authority to establish voter qualifications, even for local elections, rests with the central U.K. Parliament in Westminster. The Scottish Parliament has called on Westminster to transfer to Scotland the legislative and executive power needed to effectuate a change in the voting age. It passed legislation in 2009 to lower

---

118 Id.
121 Robbie Dinwoodie, SNP in Move to Lower Voting Age to 16: Change in Law Required, HERALD (Glasgow, Scot.), June 13, 2008, at 7, available at 2008 WLNR 11182161 (detailing how Scotland lowered the voting age to sixteen); Record of Plenary Proceedings, NAT’L ASSEMBLY FOR WALES (Feb. 6, 2008), http://www.assemblywales.org/bus-home/bus-chamber/bus-chamber-third-assembly-rop.htm?act=dis&id=81636&ds=2/2008 (approving by a vote of 44–4 a motion stating “that the National Assembly for Wales . . . [b]elieves that in order to engage young people in the democratic process, 16-year-olds should be entitled to vote”); AMs Call to Lower Voting Age, WALESONLINE (Feb. 6, 2008), http://www.walesonline.co.uk/news/wales-news/2008/02/06/ams-call-to-lower-voting-age-91466-20449924/.
122 Dinwoodie, supra note 121. The U.K. Parliament has devolved designated powers to the governments of Scotland and Wales, retaining others. See Which Responsibilities are Devolved?, CABINET OFFICE, http://www.cabinetoffice.gov.uk/content/which-responsibilities-are-devolved (last visited Mar. 9, 2012). The power to set electoral rules is among those retained by the U.K. Parliament in Westminster. See Dinwoodie, supra note 121; see also Voting at 16, WELSH NAT’L ASSEMBLY, http://exploretheassembly.org/lang/en-uk/get-involved/elections-etholiadau/voting-at-16 (last visited Sept. 4, 2010) (acknowledging that, despite the Welsh National Assembly’s vote in favor of lowering the voting age, “deciding on the age of voting isn’t one of the Devolved Fields, so it is down to Westminster to decide”).
the voting age to sixteen in pilot health board elections, and the pro-independence Scottish government permitted sixteen- and seventeen-year-olds to vote in a referendum on Scottish independence in 2010.124

In Westminster, members of Parliament’s House of Commons have introduced four bills to lower the voting age in the last six years, suffering defeats by increasingly narrow margins.125 Former British Prime Ministers Tony Blair and Gordon Brown both announced their support while in office for lowering the voting age.126

In 2002, the Electoral Commission, an independent commission charged by the British Parliament to review public election law and policy, began a review of minimum voting and

---


The commission undertook its review in response to declining election turnouts, noting that the citizens least likely to vote in the United Kingdom are the youngest group of eligible voters—eighteen- to twenty-four-year-olds. Its review followed those of independent commissions from England, Scotland, and Wales that had all recommended lowering the voting and candidacy ages as a way of increasing young people’s interest and participation in government. The commission reached five major conclusions. First, while the fact that a clear majority of countries have a voting age of eighteen does not conclusively preclude the adoption of a lower age in the United Kingdom, it does shift the burden of persuasion to those seeking the change. Second, the age at which young people attain various legal rights and responsibilities varies; therefore, while this information is useful, the nation should assess the age of electoral majority in its own context. Third, there is no consensus on the definition of “maturity” and “what it means in relation to electoral participation and minimum voting and candidacy ages.” Fourth, in research the commission conducted, young people reported that they did not vote primarily because they were insufficiently informed; the commission also noted that citizenship education was only in its infancy in the U.K. And finally, while the majority of direct respondents to the commission’s public consultation paper favored a reduction of the voting age to sixteen, the general public favored retaining the current age of

---

128 Id. at 7.
129 Id. at 8 (citing the reports of commissions in England, Scotland, and Wales that had made such recommendations in reports published in 2002). In a paper published in 2003 reporting its research and seeking public comment, the commission also noted that the Human Rights Commission in Northern Ireland also recommended a reduction in the voting age, to seventeen. The Electoral Comm’n, How Old Is Old Enough?: The Minimum Age of Voting and Candidacy in UK Elections 29 (2003).
131 Id. at 59.
132 Id. at 59-60.
133 Id. at 60.
134 Id.
eighteen, and there was “no significant or even consistent majority of young people calling for the right to vote . . . .”

The commission concluded that “there does not seem to exist a sufficiently strong argument that change now would affect the level of political engagement between young people and the political process,” and looking for “clear evidence on which to base any change in the current voting age, . . . [the commission] to date has found insufficient justification for such change.”

The commission planned to revisit the issue again in 2010 or 2011, and said that two factors in particular could “change the social context to a sufficient degree to make a lower voting age appropriate in the future.” The first factor was the continued development of citizenship education across the United Kingdom (public schools had only recently introduced a new citizenship-education program). The second factor was whether there was any change in public opinion regarding the preferred general age of majority. At the time of this writing, the commission has not yet revisited the issue. The Labour Party in 2008 submitted to Parliament a bill to lower the voting age to sixteen for all U.K. elections. Although the bill garnered some support from other parties, opponents successfully blocked its progression through Parliament.

4. The United States

The voting age everywhere in the United States is eighteen, although nineteen states permit seventeen-year-olds to vote in primaries if they will turn eighteen in time for the general election. Nothing suggests imminent change. A number of state legislatures in recent years have considered—and rejected—occasional bills or proposed constitutional

---

135 Id. at 61.
136 Id.
137 Id. at 62.
138 Id.
139 Id.
amendments that would lower state or local voting ages. The Massachusetts legislature has not yet voted on a proposal to lower the voting age to seventeen, but in 2009, three of the four Democratic U.S. Senate candidates in Massachusetts supported the measure. A proposal that would permit sixteen-year-olds to vote in limited circumstances, such as for school-district elections, is also pending before the Michigan legislature.

II. PRINCIPLES OF DEMOCRATIC ELECTORAL INCLUSION

Voter qualification rules determine the categories of individuals included in and excluded from an electorate. This part looks to democratic and liberal theories to identify standards for establishing these rules and assessing their legitimacy.

Part II.A introduces the principle of presumptive inclusion, advanced by democracy theorists and widely accepted as a foundational normative commitment of democratic states. Presumptive inclusion places on the state the burden of justifying electoral exclusions.

---

142 States that have considered proposals to lower the voting age are Arizona (2001, 2002, 2008, and 2009 proposals to lower age to sixteen); California (2004 proposal to lower voting age to fourteen, with votes of fourteen- and fifteen-year-olds counting for one-quarter vote, and votes of sixteen- and seventeen-year-olds counting for one-half vote); Hawaii (2004 proposals to lower voting age to sixteen and seventeen); Illinois (2008 and 2010 proposals to lower voting age to seventeen); Iowa (2004 and 2008 proposals to lower voting age to seventeen in school-district elections); Michigan (2004 proposal to lower voting age to seventeen, in addition to 2008 proposals to lower voting age to sixteen for all elections, and for school-district elections); Minnesota (2002, 2004, and 2009-2010 proposals to lower the voting age to sixteen or seventeen); Texas (2001 and 2003 proposals to lower voting age to fourteen); Washington (2006 and 2008 proposals to lower voting age to sixteen); and Wisconsin (2009-2010 proposal to lower voting age to seventeen). See Election Reform Legislation, NAT'L CONF. OF STATE LEGIS., http://www.ncsl.org/legislatures-elections/elections/2001-2010-database-of-election-reform-legislation.aspx (last visited Mar. 9, 2012).


Inclusion in democratic/electoral processes is presumptive, but it is not absolute. Part II.B examines democratic exclusions and the conditions that render them justifiable. It evaluates longstanding standards that have generally required, for inclusion in the electorate, both (1) ongoing connection to the community and (2) vote decision-making competence. Individuals lacking these characteristics (or indicia of them) are commonly excluded from political participation.

While these standards for inclusion (or some variation of them) have long enjoyed near-universal acceptance, few democracy theorists have sought to justify them. In other words, the basic standards for inclusion have widespread intuitive appeal and seem correct, but it has been difficult to say why they are correct.

I advance a new argument that the twin standards for inclusion can derive, not solely from democratic principles, but also from the foundational commitment to individual liberty of the liberal constitutional democratic state.

Part III turns to the standard that is of central relevance here—electoral decision-making competence. This section first develops a conception of electoral competence, since none currently exists, and next assesses the age range by which young would-be voters have reliably attained that competence.

A. Presumptive Electoral Inclusion

There are many conceptions of democracy, and each has normative implications for the democratic legitimacy of a given political system. Assessing the nature of political inclusion required for democratic legitimacy thus first requires a conception of democracy itself. The conception advanced below is a fairly typical one that describes the minimum requirements for a democratic system—i.e., the type of political participation required to render a process democratic, and the scope of political inclusion required to render a community democratic.

---

1. Defining Democratic Inclusion and Democratic Participation

A typical account of democracy provides that, in order for a political system to qualify as democratic, the people subject to its laws must collectively authorize them.\(^{146}\) A democratic government thus derives its authority from the “the people” who are the individual members of the political community. Although minimalist, this normative account supplies a particular conception of “the people” who are entitled to political participation,\(^{147}\) and it implies a conception of political participation itself.

Under this account, the people are the legal subjects of a government.\(^{148}\) This conception is narrower than that advocated

\(^{146}\) Although this is a fairly conventional conception of democracy, this formulation draws directly from the nearly identical definitions of political philosopher David Estlund and political theorist Albert Weale. For Estlund, “[d]emocracy is the authorization of laws collectively by the people who are subject to them and is inseparable from voting.” DAVID M. ESTLUND, DEMOCRATIC AUTHORITY: A PHILOSOPHICAL FRAMEWORK 66 (2008). Weale’s formulation of democratic legitimacy is that a government must, at a minimum, guarantee that “important public decisions on questions of law and policy will depend, directly or indirectly, upon public opinion formally expressed by citizens of the community, the vast bulk of whom have equal political rights.” ALBERT WEALE, DEMOCRACY 18 (2d ed. 2007).

Political theorist Iris Marion Young is among those who articulate a conception that is decidedly more robust, embracing a “minimalist understanding of democracy . . . in which democratic politics entails a rule of law, promotion of civil and political liberties, and free and fair election of lawmakers.” IRIS MARION YOUNG, INCLUSION AND DEMOCRACY 5 (2000).

Political scientist and economist Joseph Schumpeter has famously adopted what remains an atypical conception of the minimum requirements for a democratic political system. To Schumpeter, democracy exists so long as there is widespread political competition. He argues against “defin[ing] democracy by the extent of the franchise.” JOSEPH A. SCHUMPETER, CAPITALISM, SOCIALISM, AND DEMOCRACY 276 n.16 (Taylor & Francis e-Library 2003) (1943). He thus denies the centrality to democratic systems of a widespread franchise, decrying what he terms the “classic[al] doctrine of democracy.” Id. Schumpeter believes that typical voters lack sufficient political knowledge to make reliable decisions, denies the possibility of a “uniquely determined common good,” id. at 251, and argues that the expression of a public opinion “from the infinitely complex jumble of individual and group-wise . . . volitions . . . of the ‘democratic process,’ . . . lacks not only rational unity but also rational sanction.” Id. at 253.

Political theorists refer to the question of “what persons have a rightful claim to be included in the demos” as the “problem of inclusion.” ROBERT A. DAHL, DILEMMAS OF PLURALIST DEMOCRACY: AUTONOMY VS. CONTROL 98 (1982) [hereinafter DAHL, DILEMMAS]; see also DAHL, supra note 10, at 119. See generally BECKMAN, supra note 10, at 10-15 (discussing various aspects of the “problem of inclusion”).

\(^{147}\) Political theorists refer to the question of “what persons have a rightful claim to be included in the demos” as the “problem of inclusion.” ROBERT A. DAHL, DILEMMAS OF PLURALIST DEMOCRACY: AUTONOMY VS. CONTROL 98 (1982) [hereinafter DAHL, DILEMMAS]; see also DAHL, supra note 10, at 119. See generally BECKMAN, supra note 10, at 10-15 (discussing various aspects of the “problem of inclusion”).

\(^{148}\) See, e.g., ESTLUND, supra note 146, at 66 (defining the people entitled to authorize laws as “collectively[,] the people who are subject to them”). Weale defines it as “the vast bulk of . . . citizens of the community,” where “citizen” appears to refer to the legal dimension of citizenship. Weale acknowledges that citizenship is not always a necessary condition for securing political rights such as the franchise, but he observes that “it is invariably the principal basis.” WEALE, supra note 146, at 208; see also
by some theorists, who argue that “the people” should include “[e]veryone who is affected by the decisions of a government [and who] should [thus] have the right to participate in that government.”149 The broader conception is the more inclusive of the two, but also the less useful. Because there are innumerable ways in which governments’ decisions affect people, the scope of the affected by conception is difficult to delimit. And since causal connections cross national borders, species membership, and time, the conception’s implementation (i.e., the method by which the preferences of geographically scattered or remote persons, other species, future generations, et cetera, would be identified and registered) poses nearly insurmountable challenges.150 The narrower, legal conception of “the people” delimits the notion of “affected” by extending rights of participation only to those individuals who are legal subjects “bound by” or “subject to the government and its laws.” 151 The scope of government’s authority to directly regulate an individual’s behavior or status thus bounds the relevant conception of “affected.”152

This account of democracy also requires “the people” to authorize the laws that govern them. Members of a political community can participate in and influence government’s decisions in any number of ways. 153 Voting is one method, but others include political demonstrations, participatory town...
meetings or deliberation, canvassing, community or Internet organizing, letter/op-ed writing, et cetera.\textsuperscript{154} Whatever the relative merits or efficacy of other forms of political activity, pragmatic and equitable concerns generally require eventual resort to a model that registers the collective will as the aggregation of individuals’ preferences, as expressed through their vote decisions. Participatory or deliberative processes, for example, may aim to achieve consensus of some sort, but they are also apt to result in continued disagreement (even when numbers do not make such deliberation impractical). Voting provides a method for reaching legitimate, collectively binding decisions by registering and weighing equally individuals’ expressed preferences.\textsuperscript{155} Voting has thus long been the primary means by which “the people” authorize the laws—sometimes directly, by voting on policy questions, but more commonly indirectly, by electing legislators.\textsuperscript{156}

\textsuperscript{154} Id. at 105 (listing “the main forms of political participation”). Civil society theorists have argued more generally that widespread citizen participation in a range of organizations and associations—not only political associations, but also churches, athletic clubs, etc.—contributes to democracies’ flourishing. See generally ROBERT D. PUTNAM, BOWLING ALONE: THE COLLAPSE AND REVIVAL OF AMERICAN COMMUNITY (2000). The deliberative model of democracy places at the core of democratic processes not voting, but instead reasoned public deliberation aimed at achieving rational consensus. See, e.g., Joshua Cohen, Deliberation and Democratic Legitimacy, in THE GOOD POLITY 17, 22 (Alan Hamlin & Phillip Pettit eds., 1989) (arguing that political decisions are legitimate “if and only if they could be the object of free and reasoned agreement among equals”); JÜRGEN HABERMAS, BETWEEN FACTS AND NORMS: CONTRIBUTIONS TO A DISCOURSE THEORY OF LAW AND DEMOCRACY 32 (William Rehg, trans., MIT Press 1996) (1992) (arguing that political legitimacy requires “the concurring and united will of all free and equal citizens”).

\textsuperscript{155} See, e.g., ROBERT E. GOODIN, REFLECTIVE DEMOCRACY 12, 227 (2003) (acknowledging that processes of deliberation will likely end in a vote and observing that aggregation is not intrinsically bad, but instead what is objectionable about a merely aggregative model is the “mechanistic meat-grinder aspect of the aggregation of votes into collective decisions”). James Fishkin has argued in favor of methods that integrate the deliberative model into representative democracy. JAMES S. FISHKIN, DEMOCRACY AND DELIBERATION: NEW DIRECTIONS FOR DEMOCRATIC REFORM 1 (1991). Fishkin is perhaps best known for advocating “deliberative” polling, a form of opinion poll in which a small but representative group of citizens gathers, receives briefing materials on a specific issue (policies, candidates, etc.), deliberates, and is then polled. The polling and their deliberations are publicly broadcast. The process aims to reveal what public opinion would be on a given issue, were the public well-informed and engaged. Id. at 81; see also JAMES S. FISHKIN, THE VOICE OF THE PEOPLE: PUBLIC OPINION AND DEMOCRACY (1995).

\textsuperscript{156} Both Goodin and Fishkin allow that representation, “direct or indirect” or by “authorizing” laws, does not delegitimize democracy. See supra note 155. Not all democratic theorists agree. Radical democrats, for example, treat direct democracy as normative and tend to view political representation as inconsistent with democratic values, because it “impairs the community’s ability to function as a regulating instrument of justice . . . .” BENJAMIN BARBER, STRONG DEMOCRACY: PARTICIPATORY POLITICS FOR A NEW AGE 145-46 (1984). These ideas echo Rousseau’s account of unmediated popular rule, which required that citizens assemble and decide law and
I thus adopt from democratic theory and advance here the (quite modest) normative premise that a democratic government derives its authority from the individuals governed by it—i.e., the individuals subject to and bound by its laws. Those individuals are the legal subjects of the political system and thus members of the political community. They are “the people” who must collectively authorize the laws in order for that government to claim democratic legitimacy. Correspondingly, an individual’s status as a legal subject of the government, and thus a member of the political community and one of “the people,” presumptively entitles the individual to participate in the governance of a democratic system. This is what I will call the democratic principle of presumptive inclusion.

The next sections first develop the normative implications of the principle of presumptive inclusion, and then they explore its boundaries. Every political system excludes some individual members of “the people” from the most basic form of political participation—electoral participation. The “demos,” which comprises those persons within a given community entitled to political participation through the franchise, never includes all of “the people.” I explore whether and when democratic exclusion is legitimate, then make the new argument that liberal democratic theory can provide justification for the exclusion from the demos of certain categories of “the people.”

2. Implications: Presumptive Inclusion and the Burden of Justifying Exclusion

The previous section argued that the individual members of a political community—“the people”—have presumptive claims to inclusion in the demos. Presumptive inclusion, however, neither mandates electoral participation nor precludes the possibility of legitimate exclusions. In a democratic process, the aggregation of individuals’ votes

---

157 See infra notes 166-92 and accompanying text.

158 More broadly, the term demos can indicate the members of a populace, but I use its narrower construction, in which it refers to people within a given community entitled to political participation through the franchise. Demos Definition, OXFORD ENG. DICTIONARY.COM, http://www.oed.com/view/Entry/498659 (last visited Mar. 9, 2012).
determines the collective preference. But aggregated votes will accurately reflect collective community preference so long as some critical mass of the people—sufficiently representative of the whole—participate. Because the number of voters is typically large, the aggregation of votes will accurately reflect the collective preference even without the actual participation of every member of the community. Democratic legitimacy thus requires widespread electoral inclusion, but it survives tolerable levels of nonvoting and does not foreclose the possibility of some legitimate exclusion.

Their presumed inclusion entitles legal subjects to political participation through the franchise, absent some legitimate reason for treating them unequally by excluding them from it. Would-be voters subject to a government’s authority thus ought not bear the burden of demonstrating that they merit full political participation. To the contrary—the state bears the burden of demonstrating the legitimacy of its exclusion.

The report of the U.K. Electoral Commission described above is a recent and explicit example of official failure to assimilate the principle of presumptive inclusion. The commission’s report almost certainly represents the most comprehensive and balanced examination of the voting age to be conducted by any public entity to date. Yet, the commission

---

159 But see, for example, Young, supra note 146, at 6, for a sustained argument that “voting equality is only a minimal condition of political equality,” and arguing for “additional and deeper conditions of political inclusion and exclusion, such as those involving modes of communication, attending to social difference, representation, civic organizing, and the borders of political jurisdictions;” see also Susan E. Clarke, Splintering Citizenship and the Prospects for Democratic Inclusion, in THE POLITICS OF DEMOCRATIC INCLUSION 210, 211-13 (Christina Wolbrecht & Rodney E. Hero eds., 2005) (“[I]nclusive citizenship encompasses opportunities for collective problem solving and deliberation.”).

160 What some of these legitimate reasons might be is the subject of Parts II.B.1 and II.B.2, infra.

161 Elizabeth Fraser, Democracy, Citizenship and Gender, in DEMOCRATIC THEORY TODAY: CHALLENGES FOR THE 21ST CENTURY 75 (April Carter & Geoffrey Stokes eds., 2002) (“Democracy . . . has progressively come to imply the rightness of universal suffrage . . . . So, any barriers to participation, or any exclusions, have explicitly to be justified.”).

162 Part II.B, infra, discusses criteria for inclusion. Francis Schrag argues broadly that contemporary democratic theorists who argue for universal suffrage cannot adequately account for the exclusion of children. He thus argues for a system of universal suffrage in which the voting age would be lowered “substantially,” and where younger children’s interests would be represented either by a proxy vote exercised by their parents, or by an appointed “Guardian” who would represent the interests of all children. Schrag, supra note 10, at 376.

163 See supra notes 127-39 and accompanying text; see also Part I.C.3.
presumed the legitimacy of youth exclusion. Observing that most countries currently retain a voting age of eighteen, the report explicitly shifted the burden of persuasion to those seeking to change the status quo by lowering the voting age to sixteen. In other words, the commission has imposed on proponents of the enfranchisement of sixteen- and seventeen-year-olds the burden of demonstrating their entitlement to political inclusion. In so doing, it relieved the state of the obligation of justifying its exclusion. Presumptive inclusion requires the reverse: the burden rests firmly with the state to justify voter qualifications that operate to exclude any category of persons subject to its authority.

B. The Boundaries of Electoral Inclusion

The previous section argued that the state bears the burden of justifying electoral exclusions. At the same time, no demos is fully inclusive, and every democracy has adopted voter qualification rules that exclude some members of the community from electoral participation. In this section, I first describe the two basic voting criteria that electoral qualification rules typically—

---

164 Presumptive electoral inclusion embodies the norm of universal suffrage. I avoid using the term universal suffrage, however, because, as conventionally used, it describes an electoral system that imposes reasonable restrictions on the franchise, not one that is literally universal. Scholars routinely claim that nearly all of the world’s nations are democracies, and that all democracies now provide for universal suffrage, although no democratic nation allows everyone to participate in elections. See, e.g., Tatu Vanhanen, Democratization: A Comparative Analysis of 170 Countries 65 (2003) (concluding that more than 85 percent of all countries “provided for universal suffrage”); see also Beckman, supra note 10, at 2; Dahl, Dilemmas, supra note 147, at 97; Keyssar, supra note 1, at xxvi (“[T]he United States was one of the last countries in the developed world to attain universal suffrage.”); L. Massicotte et al., Establishing the Rules of the Game: Elections Laws in Democracies 26 (2004).

The use of universal suffrage to refer to what is actually less-than-universal suffrage is widely enough understood that it does not result in confusion, but it is an error nonetheless. It subtly infuses a normative judgment—that certain exclusions from the franchise are justifiable and thus ought not count as democratic deficits—into what purports to be a descriptive term (“universal suffrage”). Doing so risks an elision of both the fact and significance of exclusion. Decisions about whom to include and exclude from political participation require normative arguments. A political system should turn to normative principles to explain why a certain exclusion from the franchise is a justified exclusion; it should not simply redefine exclusions—just because they happen to be common—as nonexclusions.


166 I reiterate here that this is not a constitutional analysis of voter qualifications. It is instead a normative analysis of the obligations of the liberal democratic state, grounded in political and democracy theory. Were the analysis grounded in the obligations imposed on the federal and state governments by the U.S. Constitution, it would be necessary to consider at some length the standard to which the state would be held in order to satisfy its burden of justifying its rules.
and legitimately—aim to ensure. I then propose a new justification of these criteria grounded in liberal democratic theory.

1. Identifying Criteria for Electoral Inclusion: Interest and Competence

Few political theorists explicitly address the basic voting criteria (e.g., connection to, or interest in, a given political community) that explain and may legitimize specific electoral exclusionary rules (e.g., rules excluding nonresidents from the franchise, because they will presumably lack this connection or interest).167 Theorists have tended more generally to argue in the political liberal tradition that distinctions made between groups of individuals, in order to be just, must be reasonable and consistent with norms of equal treatment.168

Identifying and justifying the basic criteria that all voters should satisfy, however, are important endeavors. Criteria can serve as a useful standard, both for establishing voter qualification rules and against which to evaluate the legitimacy of existing rules.169

Those theorists and activists who have sought to identify basic voting criteria have reached near consensus, even across the centuries.170 Most have designated criteria that exclude from the demos individuals who lack (1) a significant

---

167 THOMPSON, supra note 63, at 20 (“[T]he electoral process should provide citizens with equal opportunities to have their votes equally counted, unless respectful reasons justify unequal treatment. The reasons are respectful if they could be mutually accepted by free and equal citizens.”); see also BECKMAN, supra note 10, at 12; Donald W. Rogers, Introduction: The Right to Vote in American History, in VOTING AND THE SPIRIT OF AMERICAN DEMOCRACY 3 (Donald W. Rogers ed., 1992).


169 DAHL, DILEMMAS, supra note 147, at 98 (concluding simply that there is “no definitive answer” to questions involving which criteria, if met, would establish the legitimacy of any given exclusion); see also BECKMAN, supra note 10, at 9 (suggesting that the reluctance of political scientists to undertake the task might be due to its “not being perceived as interesting enough or because the issue has been associated with seemingly intractable normative and conceptual muddles”). Beckman argues for a set of criteria, consistent with the liberal tradition, in the recently published Frontiers of Democracy. Id. at 8. Beckman applies his criteria to common voting rules disfranchising criminals, noncitizens, and minors. Id. He does not, however, apply his criteria to minors of different ages. See id.

170 Beckman groups existing voter qualifications as requiring “competence, belonging [in the relevant community], and independence.” BECKMAN, supra note 10, at 8. Weale argues that the franchise may be limited to those with a commitment through a “nexus” to the community “rooted in the circumstances of the lives of individuals.” WEALE, supra note 146, at 215. He argues generally that “[t]he general principles of interest, qualification and commitment through a nexus to the community therefore provide a basis for the allocation of the franchise.” Id. at 217. Youths’ lack of qualification, or competence, Weale concludes, is “[t]he sole ground for excluding children from the vote.” Id. at 214.
and ongoing connection to the community and (2) vote decision-
making competence.\footnote{See supra Part I.A.1 (recounting arguments made at the seventeenth-
century Putney Debates). Two centuries after the Putney Debates, John Stuart Mill advocated these same basic standards for inclusion—interest in community and decision-making competence. To ensure that voters had the requisite competence and to improve the quality of voting generally, Mill (somewhat infamously) proposed literacy and mathematical tests as voting criteria, as well as the allocation of additional votes to those with higher levels of education. John Stuart Mill, \textit{Considerations on Representative Government, in John Stuart Mill: On Liberty and Other Essays} 329-31 (J. Gray ed., 1991) (1861).}

While basic voting criteria have largely remained constant, notions of reliable indicia of them reflected in specific voter qualification rules have changed significantly. For example, the seventeenth-century moderates at Putney sought to ensure that voters would cast their ballots in a manner consistent with community interests, and believed that only voters with a “permanent and fixed” interest in the community and its future would reliably vote in this way.\footnote{See supra Part I.A.1.} To them, property ownership was the best indicator that a potential voter possessed the requisite interest.\footnote{See supra Part I.A.1.} Today, not property ownership but instead citizenship, residence, and law-abidingness qualifications all seek to ensure the same criteria—ongoing community interest and connection. Next, the Putney moderates sought to ensure that voters would vote in a manner that reflected independent intellectual judgment. To them, intellectual independence could not exist in the absence of economic independence. Dependent voters, they reasoned, might be unwilling to vote in a way that accurately reflected their best independent judgment if doing so risked displeasing those to whom they were economically beholden. Today’s voting rules do not inextricably link economic and intellectual independence/judgment. But to the same end, states have adopted voter qualification rules that allow the disfranchisement of adults deemed mentally incompetent. The voting age, however, is the primary voter qualification rule whose aim is to ensure that voters have developed the requisite intellectual independence and decision-making competence.

The next subsection briefly examines the justifications for these two basic voting criteria.
2. Criteria for Electoral Inclusion in the Liberal Constitutional Democracy: A Noninstrumental Justification

To the extent that democracy implies absolute majority rule, no nation is—nor aspires to be—fully democratic. A system’s commitment to the democratic principle of popular rule coexists with, and is tempered by, other commitments. The United States is a liberal constitutional federal democracy. The democratic principle requires that those subject to a government’s authority collectively authorize its laws. But constitutionalism restrains popular sovereignty, limits the power of government, and establishes procedures for legitimate rulemaking; and, pursuant to constitutionalism, federalism aims to achieve an optimal balance between local and centralized governance. The nation’s foundational value and core political commitment, however—and that which undergirds the others—is individual liberty.


175 ROBERT A. DAHL, ON POLITICAL EQUALITY 7 (2006). Dahl explains that an “ideal” can serve two purposes—one empirical, one moral—and that the two are often confused. Although the “democratic ideal” may describe a system that is in some sense perfect, the function of “ideal” here is strictly definitional or descriptive. It does not necessarily follow that the more perfectly democratic, the better. The ideal system is not necessarily the best system. It may be, but deciding that the ideal is desirable is distinct from defining what the ideal is. The former is a normative judgment; the latter a descriptive definition. Id.

176 For a discussion of more participatory or republican democratic models (in which the electorate participates in making policy decisions) and more protective ones (in which the primary function of democracy is to protect individuals’ liberty by constraining the power of government), see David Miller, The Competitive Model of Democracy, in DEMOCRATIC THEORY AND PRACTICE 133 (G. Duncan ed., 1983).

177 Constitutionalism thus ensures that our individual and collective “better selves” will constrain our “more impulsive selves”—an instrument for “the people sober to keep in check the people drunk.” Karol Edward Soltan, Introduction to Citizen Competence and Democratic Institutions 6 (Stephen L. Elkin & Karol Edward Soltan eds., 1999).

178 Liberalism is the broad political philosophy that served as the nation’s founding principle—“that all men are created equal,” that among their inalienable rights are the rights to “Life, Liberty, and the pursuit of Happiness,” and that government is instituted “to secure these rights, . . . deriving their just powers from the consent of the governed . . . .” THE DECLARATION OF INDEPENDENCE para. 2 (U.S. 1776). The Constitution’s Preamble declares the document’s purpose to include “secur[ing] the Blessings of Liberty to ourselves and our Posterity . . . .” U.S. CONST. pmbl.

Political theorist Stephen Macedo describes as a “truisim” the idea that the liberal tradition is the foundation of the nation’s political identity. STEPHEN MACE.DO, LIBERAL VIRTUES 5-6 (1990). For a discussion of individual liberty as the nation’s core value, see Vivian E. Hamilton, Immature Citizens and the State, 2010 BYU L. REV. 1055, 1068-75.
Classical liberals consider individual liberty to be the central value of the liberal state, whereas modern liberals consider it part of a more complex core aimed at ensuring that “[e]ach person has an equal right to a fully adequate scheme of equal basic liberties which is compatible with a similar scheme of liberties for all.” Individual liberty is a core value for classical and modern liberals alike, however, and is thus a least common denominator of sorts.

While individual liberty is the nation’s core value, different conceptions exist of what that liberty itself entails. The thinnest of these is negative liberty—freedom as noninterference. Negative liberty permits a person to define and pursue his or her ultimate life course. And, as I have argued elsewhere, if liberty is the state’s core value, then safeguarding individuals’ liberty must be its primary end. Basic, life-deciding liberty is thus the minimum entitlement of individuals in the liberal state.

The complement of the individual’s basic life-deciding liberty is the absence in every other person of a liberty to decide that individual’s life course. Each person thus has a claim, or right, to have the state withhold from all other persons—popular or majority wishes notwithstanding—the right to be “other-determining.” This restraining function is one of the core purposes of constitutionalism and the institutional arrangements through which it limits popular and governmental power.

---

180 JOHN RAWLS, POLITICAL LIBERALISM 291 (expanded ed. 2005).
181 Hamilton, supra note 178, at 1169-70.
182 Three conceptions of individual liberty are most prominent—negative liberty, positive liberty, and republican liberty. For a brief description of each, see Samantha Besson & José Luis Martí, Law and Republicanism, in LEGAL REPUBLICISM: NATIONAL AND INTERNATIONAL PERSPECTIVES 3, 14-15 (Samantha Besson & José Luis Martí eds., 2009); Hamilton, supra note 178, at 1070.
183 Isaiah Berlin describes “[p]olitical liberty in this sense [as] simply the area within which a man can act unobstructed by others.” ISAIAH BERLIN, Two Concepts of Liberty, in FOUR ESSAYS ON LIBERTY 118, 122 (1969).
184 Hamilton, supra note 178, at 1073.
185 Id. at 1074.
186 This is the correlativity thesis advanced by Wesley Newcomb Hohfeld, the influential cataloguer of legal rights. See generally WESLEY NEWCOMB HOHFELD, Fundamental Legal Conceptions as Applied in Judicial Reasoning, in FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING AND OTHER LEGAL ESSAYS 65 (Walter Wheeler Cook ed., 1919).
187 Just as one party’s liberty correlates to another party’s absence of liberty in Hohfeld’s analysis of legal relations, a claim correlates to a duty. Id.; see also PAVLOS ELEFTHERIADIS, LEGAL RIGHTS 107-14 (2008) (discussing Hohfeld’s model of legal relations). The term “other-determining” in this article simply refers to the ability to determine or control the life course of another person.
These liberal political values and the institutional structures that actualize them have important implications for electoral inclusion/exclusion. One of these values is the democratic principle of presumptive inclusion, which aims at a minimum to secure individuals’ consent to being governed and to ensure that the laws directly or indirectly reflect the collective preference, defined as the aggregated preferences of the individual members of the political community.

By registering a vote, an individual expresses his or her will or preference. A vote can be thought of as representing (1) the allotment or share of the individual’s influence over governance,\textsuperscript{188} (2) the individual’s transfer or surrender to the government of some corresponding share of the individual’s liberty (i.e., a transfer of power or authority) for the purpose of effectuating and enforcing the combined wills of the people, and (3) tacit acceptance that, once transferred to it, the government’s exercise of its accumulated authority is legitimate and binding, whether or not consistent with the individual’s preference.\textsuperscript{189}

Every person governs and is governed in equal measure under this liberal democratic conception of the franchise.\textsuperscript{190} No one person wields greater influence than another in the development of rules, and the resulting rules bind all equally.

The categorical exclusion from the franchise of some members of the political community through voter qualification rules disrupts this symmetry. Excluded individuals are governed, yet they are denied a corresponding/offsetting share in influence over governance. In this sense, voter qualification rules that exclude some community members are democratic deficits. There is nonetheless general agreement that certain exclusionary rules can be legitimate,\textsuperscript{191} particularly when they ensure that voters meet the two basic criteria for inclusion—the relevant community interest, and competence. Rarely addressed, however, is why these criteria themselves are justified.

\textsuperscript{188} There are, of course, other means by which individuals might influence governance. See, e.g., \textit{Birch}, supra note 9, at 159-63 (discussing various forms of direct and indirect methods by which individuals might influence decisions of government agencies).

\textsuperscript{189} The abstaining nonvoter, not prevented from voting by external or illegitimate forces, might be thought to have implicitly transferred her quantum of influence or individual liberty to the people as a whole, deferring and agreeing to the collective judgment. This type of nonvoting thus does not necessarily represent a democratic deficit.

\textsuperscript{190} This is the Aristotelian symmetrical conception of democratic legitimacy. See \textit{Aristotle}, \textit{The Politics} of \textit{Aristotle} § 1275a8 (E. Barker trans., New York, Oxford Univ. Press ed. 1958).

\textsuperscript{191} \textit{Beckman}, supra note 10, at 5.
One justification for the criteria is instrumental—i.e., without a connection to (and knowledge of) the community, and without adequate intellect and judgment, individuals cannot be relied upon to cast votes consistent with community interests. This justification can explain the ex ante exclusion of certain individuals from the franchise. The instrumental justification for the criteria for inclusion is insufficient, though, and the following two examples illustrate why.

First, assume that an individual lives and works outside the political community and otherwise has no personal connection to it. Nonetheless present within it on Election Day, she casts a well-informed and public-minded vote. Why shouldn't her vote be counted? The instrumentalist rationale justifies the outsider-voter’s ex ante exclusion by assuming that she is more likely than an insider-voter to cast a bad vote. It does not explain or justify, however, the ex post exclusion of the outsider-voter’s good vote.

Second, assume that a member of the political community lacks electoral competence. The instrumental justification for excluding incompetent voters is weaker than the instrumental justification for excluding the outsider-voter. One might reason that the outsider-voter is more likely to cast a vote that considers short-term but not long-term consequences, or otherwise vote in a manner predictably at odds with the interests of the political community. The votes of incompetent individuals, however, will presumably be distributed randomly. Their random votes should thus cancel each other out and accordingly have no effect on electoral outcomes.

I argue that a noninstrumental justification can explain and legitimize the exclusion of both the “good” outsider-voter and the incompetent voter. Recalling the liberal democratic conception of a vote’s meaning, these voters’ ballots denote (1) a share of influence over governance, but not (2) the transfer or surrender of a corresponding share of individual liberty to the government or (3) acceptance of the government’s resulting legitimate authority over them. The outsider-voter does not surrender a share of her liberty to the government or necessarily accept the government’s legitimate authority over her because she remains beyond its reach, outside the political community. The incompetent voter has influenced governance, albeit without the capacity to do so purposefully. She does not

---

192 In Part III, infra, I argue for a conception of electoral competence, but here let us simply assume the absence of the relevant competence, however it is conceived.
surrender liberty nor accept the legitimacy of government’s authority over her, however, because she presumably lacks the capacity to (even implicitly) do either.

The outsiders’ and incompetent voters’ net influence over others would thus exceed that of others over them, and the degree by which the members of the political community are governed exceeds that by which they govern. By withholding the franchise from the community outsider and the incompetent individual, then, the state prevents the unequal distribution of liberty that would otherwise result and thus performs its liberty-preserving obligation.193

Whether citizenship, residency, law abidingness, etc. are sufficiently reliable indicia of community interest/connection so as to justify various voter qualification rules is debatable. Those debates, however, are beyond the scope of this article. But whether age eighteen is a sufficiently reliable indicator of electoral competence so as to justify rules excluding younger members of the political community from the franchise (recalling that the burden of justifying exclusion rests with the state) is the heart of this article’s inquiry.

III. ELECTORAL COMPETENCE

Part II argued that the state bears the burden of justifying electoral exclusion, but that it may legitimately adopt standards that would exclude those who lack the requisite community connections/interest or the requisite electoral competence. This part argues for a conception of the competence to which the state may hold voting members of the political community.

Democracy theorists today tend to shy away from the concept of competence, perhaps partly because political elites have historically, and notoriously, invoked the supposed incompetence of various groups—including women, African Americans, and the poor—to justify their categorical disfranchisement.194

193 See ELEFTHERIADIS, supra note 187, at 108-09.
194 Marion Smiley, Democratic Citizenship, in CITIZEN COMPETENCE, supra note 177, at 371, 380 (“[T]he very undemocratic history of the concept of competence in Western politics . . . has led most democratic theorists to steer away from the language of competence in discussions of citizenship.”). Smiley argues that the concept has reflected “inadvertent and unselfconscious biases . . . [as well as intentional] political machination.” Id. at 381; see also ERIC FONER, RECONSTRUCTION: AMERICA’S UNFINISHED REVOLUTION, 1863-1877, at 278-80 (1988) (demonstrating the use of the concept of competence to justify the ongoing disfranchisement of African Americans and immigrants).
Some conception of competence, though, must underlie voting-age requirements. The connection/interest-related criterion cannot explain age-based exclusion. One might argue that the young are not members of the political community at all, and are thus not among those presumptively entitled to electoral inclusion. The argument fails, because while they may receive different treatment than do their elders when they violate government’s mandates, the young are nonetheless equally subject to them. One might alternatively argue that the young lack the requisite interest in, and connection to, the governance of the political community and are thus legitimately excluded. That argument, too, fails. As community residents (and the generally more vulnerable among them), the young have the same, if not greater, interests as their elders in issues of public concern—including public health, safety, and education. And as young people, they are more likely to bear the long-term consequences of public policy. The young are, therefore, members of the political community, with significant interest in that community and ongoing connections to it.

It is thus young people’s lack of the relevant competence that must justify their electoral exclusion. There can be little dispute that newborns lack that competence, or that the typical person acquires it at some point over the course of his or her development. Age and cognitive development are predictably correlated. There is, then, a temporal element to the attainment of electoral competence, for which age is arguably the most reasonable proxy. The impracticality of widespread individual competence assessments, moreover, makes an age-based qualification reasonable.

A voting-age qualification thus helps ensure that voters will satisfy the criterion of electoral competence. What electoral competence entails, however, remains ill-defined, even among voting experts. One expert, for example, justified young people’s ongoing exclusion by stating that youth “under 18 do not have any competence to vote, [and they possess essentially] no knowledge. If they’re lucky, they have taken one civic course.” He added later, “[they are not] mature enough to

\[195\] Indeed, the Supreme Court asserted in Brown v. Board of Education that “education is perhaps the most important function of state and local governments.” 347 U.S. 483, 493 (1954); see also Wisconsin v. Yoder, 406 U.S. 205, 213 (1972) (“Providing public schools ranks at the very apex of the function of a state.”).

\[196\] See infra note 260 and accompanying text.

\[197\] Marilyn Rauber, Vote Early—And Young: It’s the Goal of Plans to Lower the Voting Age to 16, or Even 14, RICH. TIMES-DISPATCH, June 13, 2004, at A9,
make voting judgments because they don’t have any historical perspective and they don’t have any comparable civic responsibility.”198 These statements imply possible elements of a standard of competence: certain factual (perhaps civics-related) knowledge, maturity of judgment, historical perspective, life experience, civic responsibility, etc. But are these the correct, or even among the correct, elements of the criterion of voting competence? And how can we assess whether young people have actually achieved these (or some other) elements of competence? This part answers these questions.

Part III.A develops a new, cognitive-process-driven conception of electoral competence, informed by political science, behavioral decision research (including research on voter decision making), and developmental psychology. Part III.B describes the course of development of the relevant cognitive-processing capacities, reviewing research in developmental psychology and social and cognitive neuroscience. This research explains that, as well as why, adolescent decision-making competence is context-specific. By midadolescence (around age sixteen), young people have attained adultlike cognitive-processing capacities. The domains in which they reliably and competently exercise these capacities are limited, but these situations include those allowing for unpressured, considered decision making. I conclude by arguing that privately casting a ballot in an election that has unfolded over time is such a domain.

A. Conceptualizing Electoral Competence

In this section, I first consider the voting rights of adults with cognitive impairments and explain why it is reasonable to apply a different—more lenient—standard of electoral competence to adults with cognitive disabilities than to youth. I then examine whether political/civic knowledge ought to figure into a concept of electoral competence and conclude that it ought not. I conclude by deriving a cognitive-process-driven conception of electoral competence.

---

1. Adults with Cognitive Impairments and the Competence Assessment Tool for Voting

Many democratic systems disfranchise the cognitively impaired.199 Nearly forty U.S. states, for example, have constitutional or statutory provisions prohibiting people with cognitive impairments from voting.200 Beginning in the 1990s, however, states began imposing procedural protections to guard against the unwarranted deprivation of various rights of cognitively impaired persons under guardianship, including the right to vote. Over thirty states now provide for individualized judicial determinations of whether persons under guardianship nonetheless retain the competence to vote.201

In a 2001 case in which a group of cognitively impaired adults challenged their categorical disfranchisement under Maine law, a federal district court articulated a standard for voting competence that has since been widely cited and incorporated into a Competence Assessment Tool.202 The standard articulated in Doe v. Rowe requires simply that potential voters have “the mental capacity to make their own decision by being able to understand the nature and effect of the voting act itself.”203

Using the Doe standard, psychiatrists developed the Competence Assessment Tool for Voting (CAT-V), a questionnaire administered to individuals to assess their voting capacity.204 The questionnaire is brief (seven questions) and asks respondents to imagine that it is election day for the office of state governor. It then asks questions aimed to

---


200 Hurme & Appelbaum, supra note 199, at app. A.

201 Id. at 933.


203 Id. at 51 (citing Def. Mot. for Summ. J. at 8, No. 1:00 CV 00206 (Mar. 8, 2001)).

204 The study was reported in Raymond Raad, Jason Karlawish, & Paul S. Appelbaum, The Capacity of Vote of Persons with Serious Mental Illness, 60 Psychiatric Servs. 624 (2009).
ascertain respondents' understanding of the nature of voting ("What will the people of [person’s state] do today to pick the next Governor?"\textsuperscript{205}) and the effect of voting ("When the election for governor is over, how will it be decided who the winner is?"\textsuperscript{206}). The test then provides information about two hypothetical candidates and asks the respondent to compare them, choose one, and then discuss the potential consequences to the respondent of that candidate’s election.\textsuperscript{207}

If the \textit{Doe} standard or something like it defines electoral competence, then a large category of young people—including many preadolescents—could very well qualify as competent to vote. Still, young individuals (who may or may not have attained electoral competence) and cognitively impaired individuals (who have attained the age of presumptive electoral competence) differ in significant respects relevant to voting.

Age-qualified individuals with mental impairments are members of the in-group that has presumptively attained the development-related cognitive capacities required for electoral competence. The function of a standard by which to determine their competence is to assess whether the nature of individuals’ cognitive impairments are such that the presumption of competence ought not apply to them—i.e., their impairments have prevented them from developing the relevant cognitive capacities, or have caused them to lose the relevant capacities. The state, however, must overcome two presumptions before being justified in disfranchising a cognitively impaired adult. First, it bears the burden of justifying electoral exclusion. Second, all age-qualified individuals are members of the group that has satisfied the presumption of electoral competence. Adults with cognitive impairments should receive the benefit of that presumption. In other words, evidence of electoral incompetence must be persuasive in order to justify the competence-related disfranchisement of an age-qualified individual.

The purpose of a standard by which to measure the electoral competence of young people, on the other hand, is to assess as an initial matter their acquisition as a group of the array of cognitive capacities required for competent voting. Thus while the \textit{Doe} standard might suffice to indicate adequate electoral competence despite cognitive impairment, it is not


\textsuperscript{206} Id.

\textsuperscript{207} Id.
necessarily adequate to assess the initial development-related attainment of the array of cognitive capacities required for electoral competence.

The next subsections develop a normative standard of electoral competence, beginning by considering whether competence properly includes political or civics knowledge.

2. Electoral Competence and Political/Civics Knowledge?

Rousseau believed that a well-informed citizenry was necessary to determine and implement the public good. Modern democracy theorists, too, have argued that informed and watchful citizens help ensure a responsive, accountable government. There are several reasons, however, for excluding factual knowledge from our conception of electoral competence.

First, voluminous data methodically gathered since the 1930s have consistently shown the typical citizen to be far

---

208 See Nisbet, supra note 156, at 102-03.
209 Theorists have argued that citizens can perform this function even if they do little more than vote out of office those representatives who underperform. See Schumpeter, supra note 146, at 272 ("[E]lectorates normally do not control their political leaders in any way except by refusing to reelect them . . . ."). Probably the best-known account of a theory of what is now termed "retrospective voting" is Morris Fiorina, Retrospective Voting in National Presidential Elections (1981). Political theorists note that several of Fiorina’s empirical assumptions were mistaken. See, e.g., Michael X. Delli Carpini & Scott Keeter, What Americans Know About Politics and Why It Matters XI (1996); Richard R. Lau & David P. Redlawsk, How Voters Decide: Information Processing During Election Campaigns 72 (2006). Delli Carpini and Keeter use public-opinion surveys to highlight various characteristics of better-informed citizens, such as their improved abilities to "connect their enlightened self-interest to specific opinions about the political world," concluding that "informed citizens are better citizens in a number of ways consistent with normative and pragmatic notions of what constitutes good citizenship." Delli Carpini & Keeter, supra, at 19.

210 To argue that political knowledge ought not figure into our conception of electoral competence is not to say that political knowledge is unimportant. Delli Carpini and Keeter have identified broad categories of information likely to be relevant to voting, which citizens would thus ideally be familiar with. These are

(1) The rules of the game (the institutions and processes of elections and governance);
(2) the substance of politics (the major domestic and international issues of the day, current social and economic conditions, key policy initiatives, and so forth);
and (3) people and parties (the promises, performances, and attributes of candidates, public officials, and the political parties).

Delli Carpini & Keeter, supra note 209, at 14.

211 Id. at 62-67. Delli Carpini and Keeter gathered national survey data to assess Americans’ political knowledge over time. They note that the most comprehensive collection of public-opinion surveys dates to the 1930s; this collection is held at the Roper Center for Public Opinion Research at the University of Connecticut. Id. at 66; see also Public Opinion Archives, ROPER CTR., UNIV. OF CONN.,
removed from the ideal citizen of classic democratic theory. Instead, “a large segment of the public has been and remains woefully ignorant about virtually every aspect of American politics.” Studies find that public ignorance extends to knowledge of basic civics and government.

Widespread voter ignorance alone arguably renders infeasible the adoption of specific factual knowledge as a component of voting competence. Incorporating even basic levels of civics or political knowledge into a conception of electoral competence theoretically justifies voter qualification rules that would operate to disfranchise a significant proportion of the current (aged eighteen and over) electorate. Rates of disfranchisement would be unequally distributed across the population based on differences in knowledge among


See Richard R. Lau & David P. Redlawsk, Voting Correctly, 91 POL. SCI. REV. 585, 585 (1997) (“[O]nly a tiny minority of the citizens in any democracy actually live up to these ideals. Interest in politics is generally weak, discussion is rare, political knowledge on the average is pitifully low, and few people actively participate in politics beyond voting.”); see also Rick Shenkman, Just How Stupid Are We?: Facing the Truth About the American Voter 22 (2008).

A number of political scientists have suggested, however, that flaws in the way pollsters conduct public-opinion surveys can lead to the underreporting of political knowledge. See, e.g., Melissa K. Miller & Shannon K. Orr, Experimenting with a “Third Way” in Political Knowledge Estimation, 72 PUB. OP. Q. 768 (2008); Jeffrey J. Mondak, Developing Valid Knowledge Scales, 45 AM. J. POL. SCI. 224 (2001); Jeffrey J. Mondak, Reconsidering the Measurement of Political Knowledge, 8 POL. ANALYSIS 57 (2000); Markus Prior & Arthur Lupia, Money, Time, and Political Knowledge: Distinguishing Quick Recall and Political Learning Skills, 52 AM. J. POL. SCI. 169 (2008). Some have found that the number of correct responses increased somewhat when respondents were discouraged or prevented from responding “don’t know,” Miller & Orr, supra at 775-76, 779; given an incentive to respond correctly (such as one dollar for each correct answer), Prior & Lupia, supra at 169; or given extra time in which to respond (such as twenty-four hours as opposed to one minute), id. at 169, 171.


See Susan Jacoby, The Age of American Unreason 299-300 (2008) (reporting the results of surveys conducted by the National Constitution Center); Mark M. Blumenthal, The Political Professionals Respond, in The Electoral Challenge, supra note 213, at 83 (“[O]ne can almost never underestimate the level of information about politics and government possessed by the voters who typically decide the outcome of elections.”).

At least one study suggests that teens are even less knowledgeable than their elders: whereas about 50 percent of adults could name the three branches of government, for example, only 41 percent of teenagers could do so. Jacoby, supra, at 299-300.
various groups that have held steady over time—rates would likely be higher among women than men, African Americans than whites, high school graduates than college graduates, low-income earners than high-income earners, and people under thirty than those sixty-five and older.\(^{215}\)

Public educational policy should certainly endeavor to ensure that citizens will possess basic categories of civics and political knowledge. Formal requirements aimed at ensuring well-informed voting, however, would likely result in a better-informed electorate, but also a less representative and democratic one.

Excluding a factual-knowledge component from a conception of electoral competence is also reasonable in light of variability in the instrumental utility of political knowledge itself.\(^{216}\)

First, the utility of political knowledge is situational; its value will depend on the decision-making context.\(^{217}\) A voter’s intimate knowledge of campaign-finance legislation, for example, does not help him or her decide whether to vote in favor of a proposed school-redistricting plan. It is thus difficult to identify with any certainty the knowledge required for competent voting in a given election, and knowledge requirements are likely to change from one election to the next.

Second, the utility of political knowledge is collective, in that the greater the aggregate amount of such knowledge, the greater the likelihood that the citizenry’s decisions will accurately reflect the will of the people.\(^{218}\) Because random or


\(^{216}\) Delli Carpini & Keeter, supra note 209, at 12-16 (1996). Delli Carpini and Keeter define “political knowledge” as the range of factual information about politics retained in memory. Id. at 10. “Factual information” refers to (correct) knowledge, distinct from opinions, values, and other cognitive processes like reasoning. Id. at 10-11. Retained factual information provides a context for understanding, assimilating, and assessing newly acquired information. Id. at 337 n.3.

Delli Carpini and Keeter also argue that the instrumental value of knowledge is relative, in that, all other things being equal, more information is better than less information. Id. at 14-15. Behavioral decision research, discussed infra Part III.A.3, provides some evidence to the contrary. Id. Belief in the relative value of knowledge, however, also suggests that citizens ought to become relatively “better” voters over time (in that they more accurately identify and vote consistently with their own interests); lifelong accumulation and assimilation of information enables voters to refine their opinions and interests and vote accordingly. Id.

\(^{217}\) Id. at 14.

\(^{218}\) Id. at 15.
uninformed views cancel each other out, the “miracle of aggregation” generally results in collective decisions that reflect those beliefs that are well informed and coherent.219 Thus, while a broadly informed public is critical to the functioning of a democratic system, broadly informed citizens are less critical.

Finally, individuals may have a variety of goals when making their vote decisions. Not all voters necessarily seek to cast a vote for the candidate who “would, if elected, produce a better outcome set from [the voter’s] point of view.”220 Some voters may have less instrumental goals, regarding their vote as a “speech-act” with primarily expressive or symbolic (rather than instrumental) value.221 This use of one’s ballot is not irrational, given the negligible real-world influence of the individual’s vote on an election’s outcome. Since individual voters may want to cast their ballots so as to express any number of messages, values, or beliefs, this possible use of the vote, too, weighs against substantive standards of voter knowledge.

When considering both limited voter knowledge and arguments against adopting knowledge requirements as a condition of electoral competence, questions arise as to how (relatively uninformed) voters do go about making their vote decisions and how electoral competence ought to be defined and assessed. I address these questions in the next subsections by providing a descriptive account of voter decision making drawn from behavioral decision and cognitive performance research then arguing for a standard of electoral competence that is cognitive-process driven rather than knowledge based.


Normative decision theory prescribes a decision-making model whose rules lead the individual decision maker “to

---

219 Some studies have shown that errors do not always cancel each other out; instead, voters might share misperceptions that reflect lopsided biases. In these cases, low levels of political knowledge might indeed skew election results. Craig & Martinez, supra note 213, at 77-78, 81-82.

220 ALVIN GOLDMAN, KNOWLEDGE IN A SOCIAL WORLD 323 (1999). Goldman is a philosopher who has studied voting and voter knowledge as part of a larger project aimed at identifying the social practices and institutions “that would best advance the cause of knowledge.” Id. at 79.

221 See, e.g., GEOFFREY BRENNAN & ALAN HAMLIN, DEMOCRATIC DEVICES AND DESIRES 130-31, 136-47 (2000) (developing an expressive view of voting behavior in which individuals consider voting to be primarily a “speech-act,” as opposed to serving other instrumental interests).
choos[e] the option with the highest expected utility,” based on the individual’s beliefs and values. It describes decision making as a logical process that involves (1) identifying the relevant options, (2) predicting the possible outcomes associated with each option and the probability of each outcome’s occurrence, (3) determining the relative value/utility of each outcome, and (4) combining the probabilities of occurrence and the utility of each option to identify (and choose) the option that maximizes expected value.

Just as they rarely possess optimal levels of political knowledge, individuals also rarely make decisions using the value-maximizing, decision-making model. Normative analysis is thus just the starting point of behavioral decision research. That research instead recognizes that people are not always rational, that they can make choices that are rational without using rational processes, and that they may have goals other than making the most rational choice in a given decision-making context.

Empirical political scientists Richard Lau and David Redlawsk have extensively researched voters’ decision making and conclude that “classic democratic theory sets unrealistic standards for ideal citizens at least in part because it holds unrealistic expectations about the very nature of human cognition.” Individuals’ limited cognitive-processing abilities allow them to absorb and process only a small fraction of the barrage of information to which they are typically exposed, including political information. Limited information combined with limited time and motivation can impede rational decision making.

---

227 Id. at 73-74. Thus:

Citizens do not have all the information about politics that is required of them by classic democratic theory; neither do they process that information in as logical a way as those theorists hoped, in large part because of strict cognitive limitations. It is not so much that people do a particularly bad job of
From a behavioral economics perspective, voter ignorance makes good sense: the cost to voters of acquiring information about electoral politics (policy issues, candidates’ platforms, etc.) outweighs the expected benefit (the increased likelihood of casting a vote that accurately reflects voters’ values/preferences).

In lieu of incurring the cost of educating themselves, voters generally rely on more readily available “information shortcuts” such as party affiliation and a candidate’s or party’s past performance. These shortcuts (more broadly referred to by theorists today as heuristics) substitute for more complete information, allowing voters to make decisions reasonably consistent with their preferences while expending relatively little effort. Heuristics that voters commonly use in political processing political information, of course, but rather that we do an equally bad job of processing any other type of complex information.

Id.

In contrast to the rational actors identified in economic decision-making models, people tend to be what political scientist Herbert Simon terms “boundedly rational information processors.” Herbert Simon, *Rational Choice and the Structure of the Environment*, 63 PSYCHOL. REV. 129, 136 (1956). The actual decision-making process generally aims not to “maximize” or identify the optimal option, but instead more modestly to “satisfice” or identify an adequate or satisfactory option. See generally Herbert Simon, *A Behavioral Model of Rational Choice*, 69 Q.J. ECON. 99 (1955). Individuals’ desire to make a good decision thus competes with their desire to expend minimal cognitive effort in making that decision. LAU & REDLAWSK, supra note 209, at 29.

In his now classic treatise, economist Anthony Downs famously described voters as rationally ignorant. ANTHONY DOWNS, AN ECONOMIC THEORY OF DEMOCRACY 246 (1957).

Cognitive psychologists have identified three categories of judgment heuristics that decision makers use in order to simplify complex decisions and avoid burdensome information gathering and analysis. They are (1) availability—judging probability, frequency, and causality by how easily concrete examples come to mind (e.g., when a voter encounters an unfamiliar candidate who is a Democrat, the voter’s first thought may be that Democrats favor higher taxes, and she may then apply that attribute to the new candidate), Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124, 1124-27 (1974); (2) representativeness—assigning new information to broader preexisting categories (such as stereotypes or other schema) with which it best fits (e.g., applying stereotypes based on gender, race, or age to fill in an impression of a candidate, or partisan schemata to do the same), id. at 1127-28; and (3) anchoring and adjustment—using preexisting knowledge or judgment as a starting point or presumption, then adjusting by reviewing new information (rather than independently and fully evaluating new information), id. at 1128-30. People generally categorize new information into a preexisting schema or group with certain default characteristics. Categorization is cognitively efficient because it allows people to ignore details of the new information and assign it the default values already associated with the schema. LAU & REDLAWSK, supra note 209, at 26.

Kahneman, supra note 232, at 1124; see also LAU & REDLAWSK, supra note 209, at 13. Samuel Popkin elaborates on Downs’s model in *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns* (2d ed. 1994). He argues that, despite decades of studies that show low levels of civics and political knowledge, people acquire a great deal of information in their daily lives, such as information about the economy or their communities, which they then apply to political
decision making include party affiliation, group endorsements, person stereotypes such as gender, race, or age, and poll results indicating candidate viability.\textsuperscript{234}

Researchers have studied the effectiveness of heuristic use as a decision-making strategy. In what has become a widely used method for evaluating an individual’s vote decision, Lau and Redlawsk identify “a correct vote decision as one that is the same as the choice that would have been made under conditions of full information,” given the subjective beliefs and values of the individual voter.\textsuperscript{235} Voter ignorance, they argue, poses a less serious concern for democracy if people vote correctly most of the time, despite low levels of information and knowledge.\textsuperscript{236} They found “that limited information decision strategies not only may perform as well as, but in many instances may perform better than, traditional rational . . . decision strategies.”\textsuperscript{237}

Cognitive psychologists have conducted research in other decision-making contexts that confirms that in some cases, excess information—i.e., a volume of information that is beyond an individual’s cognitive-processing capacity—hurts decision making, presumably by confusing individuals or otherwise making it more difficult for them to identify and retain salient information.\textsuperscript{238} In certain decision-making contexts, a greater amount of preexisting knowledge can hinder rational analysis of a new set of facts.\textsuperscript{239} In making judgments, people generally bring to bear their prior knowledge; in many contexts, this improves decision making. But in decision-making tasks calling for “decontextualized” reasoning—which

\begin{footnotesize}
\begin{tabular}{l}
\textsuperscript{234} Particularly in primaries, poll results can inform voters which candidates seem to be gaining consensus support and which seem to be hopelessly behind. \textit{Id.} at 233-35. \\
\textsuperscript{235} \textit{Id.} at 74-75 (emphasis added). \\
\textsuperscript{236} \textit{Id.} at 74. \\
\textsuperscript{237} \textit{Id.} at 226. \\
\textsuperscript{238} \textit{Id.} at 212. \\
\textsuperscript{239} See generally Deanna Kuhn et al., \textit{Developing Reason}, 10 THINKING & REASONING 197 (2004) (reporting studies). \\
\end{tabular}
\end{footnotesize}
requires evaluation only of evidence presented—real-world knowledge hampers analysis. Decontextualized reasoning improves the evaluation of causation—e.g., a jury’s evaluation of evidence in order to reach a verdict—and deductive reasoning generally.240

Preexisting knowledge and beliefs can hinder cognitive performance in other ways. They can lead, for example, to “undue certainty in one’s judgments.”241 Persons with preexisting beliefs or theories tend to subject new information to inconsistent standards of evidence in order to protect their preferred theories.242 Cognitive psychologist Deanna Kuhn concludes that “the causal reasoning of average adults regarding everyday matters is in fact highly fallible. People frequently make unwarranted inferences with unwarranted certainty . . . .”243 “Undue certainty” in one’s beliefs in turn “underlies the rigidity in thinking that is a major contributor to human strife.”244

At least in some contexts, then, less knowledge leads to more objective analysis and thus improves cognitive performance.245 Research confirms common wisdom that, with age and experience, people can become less open minded, less objective when analyzing new evidence, and generally more “set in their ways.” Knowledge can lead to cognitive biases that

---

240 To give an example from one recently reported study, adults examined the following syllogism involving deductive reasoning:

Syllogism 1
Premise 1: All living things need water.
Premise 2: Roses need water.
Conclusion: Roses are living things.

Deanna Kuhn, Jumping to Conclusions, 18 SCI. AM. MIND 44, 49 (2007). Although the conclusion does not follow logically from the premises, 70 percent of adults studied accepted the syllogism as valid. Compare it with the second syllogism presented to them, identical in form to the first:

Syllogism 2
Premise 1: All animals of the hudon class are ferocious.
Premise 2: Wampets are ferocious.
Conclusion: Wampets are animals of the hudon class.

Id. at 50. Only 20 percent of adults accepted this conclusion as valid, with 80 percent (correctly) rejecting it. What explains the difference in performance? Researchers posit that, because adults know the conclusion of Syllogism 1 to be true in the real world, they easily accept it, although it does not follow logically from the premises. They were able to analyze Syllogism 2 more accurately, “however, because no obfuscating real-world knowledge” clouded their analysis. Id. at 49-50.

241 Id. at 50. This sort of misplaced certainty “reflects a failure” of metacognition, or “knowing what [one] know[s].” Id.

242 Id.

243 Id. at 51.

244 Id. at 50.

245 Id. at 49-50.
impede analysis. Youth and inexperience may, perhaps counterintuitively, contribute to superior cognitive performance.

Objective reasoning can improve with practice, however, even into adulthood. Other studies conducted by Kuhn and her colleagues demonstrated that both children and adults became more careful and critical causal reasoners when given frequent opportunities to practice evaluating evidence.246 “Early adolescents [and young adults who initially] show[ed] faulty multivariable causal reasoning” also showed significant improvement after engaging with similar problems over the course of several months.247

Therefore, for most voters, the cost of acquiring and processing full political information prior to casting a vote is prohibitive, or at least generally outweighs the benefits of doing so. The typical voter nonetheless generally reaches rational and “correct” decisions by acquiring and processing smaller, readily available bits of meaningful information that function as serviceable substitutes for full information. Finally, some research suggests that less preexisting knowledge or experience can, in some instances, improve objective analysis of new information.

In the next subsection I propose a normative standard of electoral competence that accounts for this more nuanced understanding of voter knowledge and voter decision making.

4. In Support of a Cognitive-Process-Driven Conception of Electoral Competence

I argue for a standard of competence that is process driven rather than knowledge based. As discussed above, incorporating factual knowledge into a normative standard of electoral competence risks disfranchising much of the current electorate and is unnecessary to ensure correct vote decisions. Even without requiring specific knowledge, however, it is possible to identify the cognitive processes, or mental operations, involved in—and required for—competent voting.

Cognitive processes include (1) learning and retrieving information; (2) encoding, which involves forming a mental representation of information or a situation; and (3) thinking, which is the goal-directed application and “coordination of

246 Id. at 51.
247 Id.; see also Deanna Kuhn, Education for Thinking 91-101 (2005) (summarizing studies).
inferences\textsuperscript{248} and includes various forms of reasoning.\textsuperscript{249} When thinkers deliberately constrain their inferences so as to conform to what they believe are appropriate inferential norms, they engage in reasoning.\textsuperscript{250} Forms of reasoning include deductive, inductive, and analogical reasoning, as well as decision making and problem solving.\textsuperscript{251} Reasoning supplies a person with reasons for his or her beliefs and actions, or justifiability. The ability to appropriately apply and coordinate various reasoning processes is a fundamental aspect of “[r]ationality, [which] in its oldest[ and] broadest . . . sense . . . [requires] good reasons for one’s beliefs and actions.”\textsuperscript{252}

Rationality does not necessarily require applying formal logic to a set of premises or adhering to a normative, value-maximizing decision-making process. These formal cognitive processes will indeed provide “good reasons” for one’s conclusions, but, as developmental psychologist David Moshman asserts, “Even in the absence of formal proof, we often have good enough reason to choose one belief or course of action over another. There is much more to rationality than formal logic.”\textsuperscript{253}

Consider a typical voter: in order to cast a nonrandom vote, she must go through the process of acquiring some relevant knowledge/information. In an election in which numerous candidates are vying for a number of offices, for example, the typical voter is unlikely to have gathered full information about all the candidates for each office. The voter might learn from the election ballot itself only the names and party affiliations of candidates seeking a certain local office.

Our typical voter has thus learned a limited amount of information. After acquiring that information (itself a cognitive process), the voter applies additional cognitive processes to it. The voter may recall that Republicans generally favor lower taxes. She makes the inference that the Republican candidate is more likely to favor lower taxes than the Democratic candidate, and infers further that electing a Republican to office makes it more likely that taxes will be reduced. She believes that a tax

\textsuperscript{248} David Moshman, Adolescent Psychological Development: Rationality, Morality, and Identity 25-26 (2d ed. 2005).
\textsuperscript{250} Id.
\textsuperscript{251} Id. at 227.
\textsuperscript{252} Moshman, supra note 248, at 16.
\textsuperscript{253} Id.
reduction would help her and other middle-class workers like her. She also notes that the Democratic candidate is a woman, and surmises that this candidate might be even more liberal and pro-government spending than the typical Democrat. As a result, she votes for the Republican candidate. Although the voter has not made a particularly well-informed decision, she has arguably made a minimally competent one, and one likely to be correct (i.e., consistent with the decision she would have made had she possessed full information).

Therefore, our typical voter acquired relevant information about the candidates, retrieved relevant encoded information from her long-term memory, and applied deductive reasoning to reach conclusions about the candidates that led her to make a choice that she could justify with a good-enough reason. I thus begin by suggesting that a minimally competent voting decision involves the appropriate application and coordination of various reasoning processes to make a choice that could be justified by a good-enough reason.

This definition of competent voting might usefully be refined further. One possible refinement would require that instead of employing merely “appropriate” reasoning processes, competent voting employs “mature” reasoning processes. Requiring “mature” reasoning processes may go too far, though. While the level of thinking of many individuals continues to develop through and beyond adolescence, developmental psychologists have determined that there is no universal state of maturity attained by all, or even most, adults. Instead, the development of thinking in and beyond adolescence is highly variable, depending on individual interests, activities, and contexts.254 A “mature” cognitive-processing requirement for electoral competence, then, like a factual-knowledge requirement, would exceed the capacities of—and thus disfranchise—many current voters.

A more pragmatic standard for electoral competence could modestly require “adultlike” cognitive-processing capacities—i.e., the minimum levels of thinking and processing attained by developmentally normal adults.

---

254 Id. at 24 (observing that while the concepts and forms of reasoning of many individuals continue to develop after childhood, “postchildhood developmental changes in thinking are not tied to age and do not culminate in a state of maturity”). Cf. Laurence Steinberg & Elizabeth Scott, Less Guilty by Reason of Adolescence: Developmental Immaturity, Diminished Responsibility, and the Juvenile Death Penalty, 58 AM. PSYCHOL. 1009, 1016-17 (2003).
The resulting standard of electoral competence thus provides that a minimally competent voting decision involves an adultlike application and coordination of various reasoning processes to make a choice that could be justified by a good-enough reason.

The next section surveys recent research on adolescent development and concludes that adolescents reliably attain the relevant cognitive-processing abilities by age fifteen or sixteen.

B. Assessing Electoral Competence

Researchers who study cognitive development cannot precisely identify every context in which developmentally normal citizens are likely to have decision-making competence, given both individual and situational, or context-specific, variability. But, scientists have made two critical findings. First, by midadolescence, individuals have the cognitive capacity to make competent decisions. Second, certain situations and factors can hinder the decision-making abilities that adolescents otherwise possess.

This section first canvasses research on adolescent cognitive development from various disciplines in the developmental sciences. I then conclude that the factors that characterize the vote decision-making context (time for deliberation, the absence of peers, etc.) render voting the type of domain in which midadolescents will capably exercise the relevant cognitive-processing capacities.


Adolescence, the developmental period between childhood and adulthood, is characterized by increases in both rational decision-making capacity and irrational risk-taking behavior. Developmental neuroscientists have begun developing a neurologically based model that has the potential to explain the simultaneous increases in adolescents’ risk taking and poor decision making on the one hand, and improved cognitive ability

255 Moshman, supra note 248, at 13.
256 See infra notes 271-74 and accompanying text.
on the other. A brief discussion of relevant aspects of adolescent development follows, first from the perspective of behavioral science and then from that of developmental neuroscience.

Cognitive capacity, including learning and reasoning from facts and information processing, improves more or less linearly throughout childhood, reaching adultlike levels by midadolescence. Researchers have consistently found “[t]he logical reasoning and basic information-processing abilities of 16-year-olds” to be “comparable to [or essentially indistinguishable from] those of adults.” By midadolescence, thinking processes are adultlike. According to developmental psychologist David Moshman, “[n]o theorist or researcher has ever identified a form or level of thinking routine among adults that is rarely seen in adolescents.”

Despite their apparently advanced cognitive abilities, universal characteristics of adolescent behavior include increased propensities for impulsivity, risk taking, and sensation seeking. Early behavioral decision models attributed these behavioral characteristics to cognitive deficiencies that caused adolescents to misperceive risks and fail to appreciate universal characteristics of adolescent behavior include increased propensities for impulsivity, risk taking, and sensation seeking. Early behavioral decision models attributed these behavioral characteristics to cognitive deficiencies that caused adolescents to misperceive risks and fail to appreciate.

---

258 Casey et al., supra note 257, at 63 (discussing cognitive and neurobiological hypotheses that fail to adequately account for adolescent decision-making behavior).

259 Laurence Steinberg recently emphasized the importance—to all disciplines within developmental science—of research in developmental neuroscience, suggesting that this research has the “potential to structure a new, overarching model of normative . . . adolescent development.” Laurence Steinberg, A Behavioral Scientist Looks at the Science of Adolescent Brain Development, 72 BRAIN & COGNITION 160, 162 (2010) [hereinafter Steinberg, Adolescent Brain Development]. See generally Laurence Steinberg, A Social Neuroscience Perspective on Adolescent Risk-Taking, 28 DEV. REV. 78 (2008) [hereinafter Steinberg, Adolescent Risk-Taking].

260 For a summary of cognitive development from early childhood through early adulthood, see Hamilton, supra note 178, at 1099-1116.

261 Steinberg, Adolescent Risk-Taking, supra note 258, at 80.

262 Moshman, supra note 248, at 24.

263 Sara B. Johnson et al., Adolescent Maturity and the Brain: The Promise and Pitfalls of Neuroscience Research in Adolescent Health Policy, 45 J. ADOLESCENT HEALTH 216, 218 (2009). Compared with adults over twenty-five, adolescents and young adults are more likely to binge drink, commit crimes, engage in violence, have casual sex, and cause serious or fatal automobile accidents. Steinberg, Adolescent Risk-Taking, supra note 258, at 79; see also Michael Windle & Rebecca C. Windle, Alcohol and Other Substance Use and Abuse, in BLACKWELL HANDBOOK, supra note 249, at 450-63 (surveying empirical studies on the onset and escalation of substance use among adolescents).

Developmental scientists reason that evolutionary processes would have selected for these characteristics, which presumably motivated adolescents (of all cultures and species) to leave their natal environments and seek out mates. Steinberg, Adolescent Brain Development, supra note 258, at 161.
the long-term consequences of their decisions. Intervention programs implemented to counteract these cognitive deficiencies by correcting adolescents’ misperceptions about common risks, however, were largely ineffectual in changing adolescent behavior. Studies, moreover, revealed no cognitive differences between adolescents and adults that could explain their different propensities for risk taking.

Behavioral scientists thus reached the counterintuitive conclusion that adolescents engage in higher rates of risky, seemingly irrational behavior than do adults despite being as “knowledgeable, logical, reality-based, and accurate in the ways in which they think about risky activity . . . as their elders.” Cognitive deficiencies do not account for adolescents’ propensity for risky and impulsive decision making. Studies instead consistently confirm that adolescents have the cognitive competence to make rational decisions about risks. Researchers have endeavored to determine why adolescents nonetheless frequently make irrational, risky decisions.

Behavioral scientists examined more closely the real-world contexts in which adolescents make decisions, and have gained valuable insights into adolescent decision-making processes. Their findings confirmed adolescents’ competence to make rational decisions—at least when making decisions in the artificially ideal conditions of the research laboratories in which they complete tasks involving minor, symbolic risks. The real-world contexts in which adolescents usually make decisions, however, can drastically affect the quality of their decision making.

265 Id. at 33 (surveying studies of education interventions aiming to seek adolescents’ misperceptions by educating them about commonly encountered risks).
266 Id.
267 Id.
268 Id.
269 Behavioral scientists define a “context [as] a culturally defined situation that (a) occurs in a particular time and place and (b) contains actors who perform culturally defined roles.” James P. Byrnes, The Development of Self-Regulated Decision Making, in The Development of Judgment, supra note 223, at 5, 7.
270 Id.; Reyna & Farley, supra note 264, at 2.
271 Id.; Reyna & Farley, supra note 264, at 2.
Studies found that contexts that predictably compromise adolescent decision-making include those requiring them to make decisions “[i]n the heat of passion, on the spur of the moment, in unfamiliar situations, . . . and when behavioral inhibition is required for good outcomes . . . .” In other words, adolescents tend to make bad decisions in emotionally charged or pressured situations, and they struggle to control impulses that lead to undesirable behavior.

Developmental neuroscientists also study adolescent cognitive development and have begun creating a neurologically based model primarily oriented around development in two neural systems of the brain—the system associated with cognitive control, and the one associated with socio-emotional maturity. The core insight of this dual-systems model is that these two neural systems develop along different timelines. This temporal disjunction has the potential to explain adolescents’ risk-taking and poor decision-making despite their improved cognitive ability, as well as other aspects of adolescent psychology and behavior. An overview of the model’s features follows.
The socio-emotional system within the dual-systems model includes neural circuitries across regions of the brain implicated in social-information-processing and reward-seeking/processing. When certain neurons (nerve cells that transmit information throughout the brain in the form of electrical or chemical impulses) are stimulated by a chemical impulse, they trigger the release of neurotransmitters that then chemically stimulate the next neuron in the circuit. In the socio-emotional system, the neurotransmitter dopamine modulates the neural reward circuitry. The mechanisms underlying dopamine neurotransmission continue to mature during adolescence. Dopaminergic activity peaks rapidly and dramatically in early adolescence, around the time of pubertal maturation.

Researchers believe that this peak in dopaminergic activity makes adolescents experience a potentially rewarding stimuli as even more rewarding “than would be the case during either childhood or adulthood.” The resulting heightening of reward salience leads to increased sensation seeking—a “tendency to seek out novel, varied, and highly stimulating experiences, [coupled with a] willingness to take risks in order to attain them.” Consistent with this theory, studies show that sensation seeking, risk preference, susceptibility to deviant or antisocial peer influence, and reward sensitivity all follow a curvilinear, (“\(\wedge\)”) shaped trend. These behavioral characteristics begin to increase at age ten, peak around ages fourteen to fifteen (depending on the study and measure used), and then decline.

The second neural system in the dual-systems model is the cognitive control system. Cognitive control refers to the abilities to voluntarily coordinate and engage in goal-directed behavior. This system includes the prefrontal cortex, which is

---

277 The socio-emotional system includes the “amygdala, nucleus accumbens, orbitofrontal cortex, medial prefrontal cortex, and superior temporal sulcus.” Steinberg, Adolescent Risk-Taking, supra note 258, at 83.


279 Geier & Luna, supra note 257, at 216.

280 Id. at 216-17; Steinberg et al., supra note 275, at 1764-66.

281 Id.; Steinberg, Adolescent Risk-Taking, supra note 258, at 85.

282 Id.; Steinberg et al., supra note 275, at 1765.

283 Steinberg, Adolescent Brain Development, supra note 258, at 163; Sindy R. Sumter et al., The Developmental Pattern of Resistance to Peer Influence in Adolescence: Will the Teenager Ever Be Able to Resist?, 32 J. ADOLESCENCE 1009-10 (2009); see also Steinberg, Adolescent Risk-Taking, supra note 258, at 89 (ages thirteen to sixteen); Steinberg et al., supra note 275, at 1774 (ages twelve to fifteen).

284 Luna et al., supra note 272, at 101.
involved in executive, decision-making, and self-regulatory functions, and “association” areas, which connect different regions of the brain and thus support the complex integration of functions.285 The cognitive control system follows a more gradual and linear developmental trajectory than does the socio-emotional system.286 Three structural changes in the brain characterize the maturation of cognitive control during adolescence.

The first structural change involves a process known as synaptic pruning, by which synapses (the point of contact between two nerve cells in a given neural circuit) that have not been stimulated (due to lack of use) are eliminated, and remaining synaptic connections stabilize and strengthen. Synaptic pruning begins during childhood and accelerates in adolescence, with the prefrontal cortex maturing in midadolescence.287 This correlates with the maturation of basic cognitive processes by the age of sixteen.

Second, myelination (a process involving the insulation of existing connections between neurons with a fatty layer that improves neural connectivity) continues within the regions of the cortex and between the different cortical regions through adolescence and into the twenties.288 This change correlates with observed behavioral improvements in higher-order and executive functions (future orientation, planning, response inhibition, spatial working memory, etc.) associated with the integrated functioning of multiple prefrontal regions of the brain.289

285 Steinberg, Adolescent Risk-Taking, supra note 258 at 93-94. The cognitive control system also includes parts of the corpus callosum, which connects the left and right hemispheres. Beatriz Luna, Developmental Changes in Cognitive Control Through Adolescence, in ADVANCES IN CHILD DEVELOPMENT AND BEHAVIOR 233, 240 (Patricia Bauer ed., 2009).

286 Steinberg, Adolescent Risk-Taking, supra note 258, at 93.

287 Nitin Gogtay & Paul M. Thompson, Mapping Gray Matter Development: Implications for Typical Development and Vulnerability to Psychopathology, 72 BRAIN & COGNITION 6, 7 (2010); Tomas Paus, Mapping Brain Maturation and Cognitive Development During Adolescence, 9 TRENDS IN COGNITIVE SCI. 60, 62 (2005); Arthur W. Toga et al., Mapping Brain Maturation, 29 TRENDS IN NEUROSCIENCES 148, 149-50 (2006). There is also some evidence of synaptic pruning in the association areas (areas throughout the brain which connect its different regions and support the complex integration of interregional function). Luna, supra note 285, at 238.

288 Geier & Luna, supra note 257, at 216; Gogtay & Thompson, supra note 287, at 7; Luna, supra note 285, at 237-41; Tomas Paus, Growth of White Matter in the Adolescent Brain: Myelin or Axon?, 72 BRAIN & COGNITION 26 (2010); Steinberg, Adolescent Risk-Taking, supra note 258, at 94-96. Since myelination involves the “gradual enhancement of established connections”—as opposed to the initial establishment of such connections—the “changes in white matter [represent] a refinement of executive control processes that are in place earlier in development.” Luna, supra note 285, at 239-40.

289 Steinberg, Adolescent Risk-Taking, supra note 258, at 94-96.
Third, myelination also continues between the cortex and other regions of the brain, including connections between regions involved in social and emotional information processing, and those involved in cognitive control processes (especially the prefrontal regions).\textsuperscript{290} The increased connectivity between these regions correlates with coordination of affect (the external expression of emotions) and cognition; emotional regulation and impulse control both improve through the midtwenties as a result.\textsuperscript{291} Strategic planning, anticipation of future consequences, and resistance to neutral (as opposed to antisocial) peer influence and peer influence in general all follow the same trajectory, increasing linearly from preadolescence through late adolescence and early adulthood.\textsuperscript{292}

In summary, adolescents’ basic cognitive abilities are mature by the age of sixteen, giving them the capacity to process information and make rational decisions. But the heightened sensitivity to reward that increases and peaks around midadolescence inclines young people towards risk taking, sensation seeking, and impulsivity. These inclinations may dominate or overwhelm their cognitive processes and shape their behaviors, especially in situations triggering heightened emotion or pressure.\textsuperscript{293}

Adolescents’ susceptibility to the confounding influence of heightened-reward salience on their decision making begins to decline after midadolescence, while their ability to exercise cognitive control increases, ultimately reaching mature levels in their twenties.\textsuperscript{294}

2. Domain-Specific Competence: The Vote Decision

By ages fifteen or sixteen, adolescents have attained adultlike cognitive-processing capacities.\textsuperscript{295} In other words, they

\textsuperscript{290} Id. at 94-98. Important social and emotional information-processing regions of the brain include the limbic and paralimbic regions. Id. at 94-95.
\textsuperscript{291} Id.
\textsuperscript{292} Id.; see also Sumter et al., supra note 283, at 1016 (reporting “a steady increase in resistance to general peer influence with age”). See generally Luna et al., supra note 272, at 101.
\textsuperscript{293} Luna, supra note 285, at 257; Steinberg, Adolescent Risk-Taking, supra note 258, at 96-98. Researchers have generally found the following personality traits and contextual factors to correlate with suboptimal choices: sensation seeking, impulsivity, competitiveness, overconfidence, and the presence of peers. Byrnes, supra note 269, at 31-32.
\textsuperscript{294} Luna, supra note 285, at 257; Steinberg, Adolescent Risk-Taking, supra note 258, at 97-98.
\textsuperscript{295} David Archard has addressed the voting age in his now-classic volume on children’s rights. He argues for a minimalist concept of voter competence, conceivably
are as able as adults to acquire, retain, and retrieve relevant information and apply reasoning processes that lead to justifiable conclusions.

But while they have adultlike abilities to think and reach rational judgments, their capacities are more susceptible to being confounded by the real-world contexts in which they make decisions. When they must either make decisions quickly or under time pressure, or when they are highly emotional or stressed, adolescents’ performance suffers. In contexts in which adolescents are likely to make poor decisions—especially when their decisions will have negative externalities—the state properly constrains their decision-making liberty.

One example of such a real-world context is driving. Driving provides ready opportunities for risk taking and thrill seeking—especially in the presence of encouraging peers. At the same time, responsible driving frequently requires rapid decision making in response to unpredictable situations, in what is still a new and unfamiliar context to the inexperienced adolescent driver. Data on adolescent collisions and motor vehicle-related fatalities provide compelling evidence of the challenges adolescent drivers face. In other work, I discuss additional contexts likely to impair adolescents’ otherwise-mature cognitive-processing abilities, as well as those contexts in which their cognitive-processing abilities (and the rationality and maturity of their resulting decisions) are likely to remain uncompromised.

Elections, on the other hand, are a decision-making domain in which midadolescents’ adultlike cognitive-processing abilities should remain uncompromised. Elections unfold over a period of time, giving voters the opportunity to deliberate and evaluate options without undue pressure. Many sources of information are readily available over a period of time as well,
which voters can use as a kind of scaffolding or heuristics to help them evaluate their choices—broadcast debates, endorsements of candidates, party affiliations, etc. Voting itself is done anonymously and in private, which diminishes the concern that adolescents’ choices will be unduly pressured or influenced by peers.298

By the age of sixteen, adolescents meet the standard of electoral competence defined above, where “a minimally competent voting decision involves an adultlike application and coordination of various reasoning processes to make a choice that could be justified by a good-enough reason.”299

Finally, Lau and Redlawsk’s test for correct voting300 may help assess adolescents’ vote decisions. Lau and Redlawsk’s mock-election study predicted that 70 percent of voters vote correctly; their study of the nine actual presidential elections from 1988 to 2004 showed that the mean number of correct voters was just over 75 percent.301 Empirical studies of correct voting might thus be used as a benchmark or test of adolescent voting competence: if adolescents cast correct votes between 70 to 75 percent of the time, then they have achieved adultlike levels of competence.302

CONCLUSION

Even without including the numerous policy considerations that support lowering the voting age (for example, making tangible and relevant to young people the civic education they receive in middle and secondary schools, and encouraging interest in, and habits of, civic participation), compelling reasons to do so—grounded in foundational democratic principles—have emerged. I have argued that democratic legitimacy requires the presumptive electoral

298 See Catherine J. Ross, A Stable Paradigm: Revisiting Capacity, Vulnerability and the Rights Claims of Adolescents after Roper v. Simmons, in LAW, MIND AND BRAIN (Michael Freeman & Oliver R. Goodenough eds., 2009) 183, 184-87, 193-96 (discussing adolescent decision making by the Supreme Court in Roper and arguing for a paradigm of rights that accounts both for adolescents’ vulnerabilities and capacities).

299 See supra Part III.A.4.

300 To determine a correct vote, Lau and Redlawsk ask whether “[i]rrespective of how the vote decision is actually reached, how frequently do voters vote correctly.” See supra notes 226-37 and accompanying text; LAU & REDLAWSK, supra note 209, at 88; see also Richard R. Lau et al., An Exploration of Correct Voting in Recent U.S. Presidential Elections, 52 AM. J. POL. SCI. 395 (2008).

301 LAU & REDLAWSK, supra note 209, at 85, 88; Lau et al., supra note 300, at 406.

302 An arguably more radical approach might rely on aggregation models to find adolescent voters competent once they have reached a significantly lower threshold of correct voting—presumably something greater than 50 percent.
inclusion of members of the political community. Democratic systems may nonetheless legitimately impose competence-related electoral qualifications. Voter qualification rules excluding citizens younger than eighteen from the electorate are justified by the presumed electoral incompetence of that category of citizens, but the requirements of electoral competence remain unspecified. By studying voter decision making and the development of cognitive-processing skills, it is possible to derive a pragmatic conception of electoral competence. Research demonstrates that young people reliably attain electoral competence by the age of fifteen or sixteen. Thus, labeling them incompetent is error and can no longer justify their continued exclusion.

States should thus lower the age of electoral majority to sixteen, by which age it is safe to say that adolescent citizens will be competent voters.
NOTES

For Sale

THE THREAT OF STATE PUBLIC ACCOMMODATIONS LAWS TO THE FIRST AMENDMENT RIGHTS OF ARTISTIC BUSINESSES

We begin with the proposition that the right of freedom of thought protected by the First Amendment against state action includes both the right to speak freely and the right to refrain from speaking at all.1

INTRODUCTION

Over the last fifty years, the changing landscape of the American economy and the continued evolution of state public accommodations laws toward protection of a greater number of suspect classes in a wider variety of places2 have created an environment of potentially widespread First Amendment violations.3 Public accommodations laws are the modern conception of the common-law principle that innkeepers and other common carriers could not refuse service to customers without good reason.4 Historically, this principle was a narrow

1 Wooley v. Maynard, 430 U.S. 705, 714 (1977) (invalidating a New Hampshire regulation requiring noncommercial vehicles to carry license plates inscribed with the State's motto because it improperly forced individuals to publicly disseminate the State's ideological message).


3 Lauren Rosenblum, Note, Equal Access or Free Speech: The Constitutionality of Public Accommodations Laws, 72 N.Y.U. L. REV. 1243, 1249 (1997). The First Amendment states, “Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the government for a redress of grievances.” U.S. CONST. amend. I.

but upon codification, many states have significantly expanded its scope by increasing both the types of businesses subject to the laws and the classes of people protected by them.\footnote{Griffin, supra note 4, at 1047-48.}

Public accommodations laws are generally designed to ensure equal access for all people to publicly available goods and services, even where privately owned businesses offer those goods and services.\footnote{Lerman & Sanderson, supra note 2, at 218.} The “public character” rationale—which states that owners of inns, restaurants, and other common carriers perform “quasi-public” services while operating for a profit—serves as a primary justification for abridging individual rights in this context.\footnote{Hurley, 515 U.S. at 571 (explaining that “innkeepers are a sort of public servants”) (quoting Rex v. Ivens, 7 Car. & P. 213, 219, 173 Eng. Rep. 94, 96 (N.P. 1835)); Griffin, supra note 4, at 1055.}

A predicament arises when a business offers inherently expressive goods or services, such as photography, music, or any other business that involves the commercialization of art.\footnote{Nat’l Endowment for the Arts v. Finley, 524 U.S. 569, 602 (1998) (Souter, J., dissenting) (“It goes without saying that artistic expression lies within this First Amendment protection.”); Griffin, supra note 4, at 1048.} If a customer wishes to hire an artist to provide artwork or other similarly expressive services for a cause with which the artist does not agree, the artist may be compelled by a state public accommodations law to express an idea, or associate himself with an idea, with which he does not agree on pain of

\footnote{Griffin, supra note 4, at 1055.}

\footnote{Lerman & Sanderson, supra note 2, at 218.}

\footnote{Lerman & Sanderson, supra note 2, at 218.}

\footnote{Griffin, supra note 4, at 1055.}

\footnote{Griffin, supra note 4, at 1048.}
civil sanctions. This compelled expression or association likely violates the artist’s First Amendment rights.12

A recent example is illustrative of the conflict. In the fall of 2006, Vanessa Willock and Misti Collinsworth sought a photographer for their same-sex commitment ceremony.13 Willock contacted Elaine Huguenin, co-owner and primary photographer of Elane Photography,14 through the company’s website seeking services.15 Huguenin declined to provide service.16 As an artist, Huguenin believed she expressed herself through her photographs and became part of the events that she photographed.17 For Huguenin to photograph a same-sex ceremony would express an idea contrary to her belief that marriage exists only between two individuals of the opposite sex.18 Willock filed a discrimination claim against Elane Photography with the New Mexico Human Rights Division. After an investigation, the Human Rights Commission found Elane Photography guilty of discrimination under the New Mexico public accommodations law, the Human Rights Act, § 28-1-7(F),19 and awarded Willock $6,637.94 in attorney’s fees.20 The District Court for the Second Judicial District of New Mexico affirmed on appeal.21 Neither the Human Rights Commission nor the district court gave significant consideration to Huguenin’s freedom of expression argument.22

12 Wooley v. Maynard, 430 U.S. 705, 714 (1977) (“[T]he right of freedom of thought protected by the First Amendment against state action includes both the right to speak freely and the right to refrain from speaking at all.”); see also W. Va. State Bd. of Educ. v. Barnette, 319 U.S. 624, 642 (1943) (“If there is any fixed star in our constitutional constellation, it is that no official, high or petty, can prescribe what shall be orthodox in politics, nationalism, religion, or other matters of opinion or force citizens to confess by word or act their faith therein.”); see also Hurley, 515 U.S. at 573 (“[A] fundamental rule of protection under the First Amendment [is] that a speaker has the autonomy to choose the content of his own message.”); Eugene Volokh, Compelling Speech by Commercial Photographers, Freelance Writers, Musicians, and So On, VOLOKH CONSPIRACY (Dec. 16, 2009, 4:01 PM), http://volokh.com/2009/12/16/compelling-speech-by-commercial-photographers-freelance-writers-musicians-and-so-on/.


14 Elaine Huguenin’s studio’s name is spelled without an “i.”

15 Willock, HRD No. 06-12-20-0685, slip op. at 4.

16 Id. at 5.

17 Id. at 6.

18 Id.

19 Id. at 14.

20 Id. at 19.


22 Willock, HRD No. 06-12-20-0685, slip op. at 16-17; Willock, CV-2008-06632, slip op. at 8-10. Just prior to going to press, the Court of Appeals for the State of New
The Elane Photography case is interesting for two reasons. First, when First Amendment principles are applied, it appears that Huguenin’s free speech defense was stronger than either the Commission or the district court acknowledged. Second, it calls into question the constitutionality of state public accommodations laws generally as applied to an enormous class of businesses.23

Part I of this note will summarize the Elane Photography case and its appeal to the New Mexico district court. Part II will demonstrate how the New Mexico Human Rights Act was unconstitutional as applied to Elane Photography under current First Amendment doctrine due to a gradual shift in jurisprudence toward greater First Amendment protection. In Part III, a brief discussion of public accommodations laws generally will show the seriousness of the potential conflict between these laws and the First Amendment freedom of expression24 and will discuss several suggested solutions to the problem.

I. WILLOCK V. ELANE PHOTOGRAPHY

A. The Facts

Elane Photography is a limited liability company co-owned by husband and wife Jonathan and Elaine Huguenin25 and operates in Albuquerque, New Mexico.26 Elaine Huguenin (Huguenin) was the studio’s primary photographer and Jonathan Huguenin was the business manager.27 The business provided photography services primarily for weddings and engagements.

Mexico handed down its decision on Elane Photography’s appeal from the district court. Like the district court, the Court of Appeals rejected the Huguenins’ First Amendment arguments. The Court of Appeals reasoned that Huguenin’s photography was not sufficiently expressive to warrant First Amendment protection and, therefore, the State could constitutionally apply the New Mexico public accommodations statute to Huguenin’s conduct. Elane Photography v. Willock, No. 30,203, at ¶ 29 (N.M. Ct. App. May 31, 2012). Further, to the extent Huguenin did produce expression, the Court of Appeals agreed with the district court that Huguenin was a mere conduit of her clients’ messages. Id. Because the Court of Appeals affirmed the district court’s decision and reasoning on Huguenin’s First Amendment argument, the analysis and reasoning in this note remain relevant.

23 This note is limited to discussion of state public accommodations laws only; the federal public accommodations law, 42 U.S.C. § 2000a (2006), embodied in Title II of the Civil Rights Act of 1964, is beyond the scope of this discussion.

24 To the extent that Elane Photography implicates other First Amendment concerns, such as the free exercise of religion, or claims arising under state law other than public accommodations law, such as the New Mexico Constitution or the New Mexico Religious Freedom Reformation Act (N.M. Stat. Ann. § 28-22-1 (2000)), those issues are not addressed by this note.

25 Willock, HRD No. 06-12-20-0685, slip op. at 2.

26 Id.

27 Id.
and also for its customers’ significant life events. The business had a website which featured sample wedding pictures taken by Huguenin and advertised the studio’s services. Huguenin’s initial contact with most clients was via e-mail through the website. After meeting with a potential client and agreeing to go forward, Huguenin would provide a written contract in which the company explicitly retained all rights with regard to the prints and proofs of the photographs, including the right to use them for “advertising, display or any other purpose thought proper by [the Studio].”

In the fall of 2006, Vanessa Willock (Willock) and Misti Collinsworth (Collinsworth) were seeking a photographer to record their same-sex commitment ceremony. On September 21, 2006, Willock e-mailed Huguenin after contacting the Elane Photography website, and specified in her e-mail that she needed a photographer for a same-gender ceremony. Within a day, Huguenin replied with the ambiguous statement that “[a]s a company, we photograph traditional weddings, engagements, seniors, and several other things,” but did not give Willock a definite answer whether Elane Photography would take the job. Approximately two months later, Willock sent another e-mail in order to clarify whether Elane Photography would serve same-sex couples. In her second reply, Huguenin responded, “[W]e do not photograph same-sex weddings.”

Upset by what appeared to be discrimination, the couple decided to confirm that the studio refused to serve them because of their sexual orientation. To that end, Collinsworth sent an e-mail requesting Elane Photography’s services without disclosing that she was having a same-sex ceremony or that she was Willock’s partner. Huguenin responded affirmatively with all the information Collinsworth requested and offered to set up a meeting to discuss the job in person. As a result of these events, Willock filed a discrimination claim with the Human

---

28 Id.
29 Id. at 3.
30 Id.
31 Id. at 4 (quoting testimony of Elaine Huguenin and Jonathan Huguenin; Ex. A).
32 Known at the time of the events as Misti Pascottini. Id. at 4, 7.
33 Id. at 4.
34 Id. at 5.
35 Id.
36 Id.
37 Id.
38 Id. at 6.
39 Id. at 6, 7.
40 Id. at 7.
Rights Division\textsuperscript{41} of the New Mexico Department of Labor\textsuperscript{42} against Elane Photography on December 20, 2006.

\textbf{B. The Human Rights Commission}

In its Decision and Final Order, the New Mexico Human Rights Commission (the Commission)\textsuperscript{43} made a number of significant findings of fact based on the testimony of the parties. Among other facts, the Commission found that “Elane Photography also had an unwritten company policy, which was shared between its co-owners, [the Huguenins], that Elane Photography would not photograph any image or event which was contrary to the religious beliefs of its co-owners.”\textsuperscript{44} Huguenin held the religious belief that marriage could only be between individuals of the opposite sex.\textsuperscript{45} The photographer also believed that “as an artist, [Huguenin] became a part of the events which she photographed and an owner of the images or messages conveyed in her photographs.”\textsuperscript{46} She therefore declined to provide her services to Willock because to do so would help convey a message that was contrary to her religious beliefs.\textsuperscript{47}

Willock’s claim asserted that Elane Photography’s denial of service violated section 28-1-7(F) of the New Mexico Human Rights Act.\textsuperscript{48} The statute states that it is unlawful discrimination for “any person in any public accommodation to make a distinction, directly or indirectly, in offering or refusing to offer its services, facilities, accommodation or goods to any person because of race, religion, color, national origin, ancestry, sex, sexual orientation, gender identity, spousal affiliation or

\textsuperscript{41} Currently the Human Rights Bureau. \textit{Id.} at 8.
\textsuperscript{42} Currently the New Mexico Department of Workforce Solutions. \textit{Id.} at 8.
\textsuperscript{43} On its Frequently Asked Questions page, the New Mexico Department of Workforce Solutions explains:

The Human Rights Commission is comprised of eleven citizens appointed by the governor to conduct hearings involving discrimination complaints. The eleven members volunteer their services and are not employees of the state. A commission hearing may be conducted by a single hearing officer or a three-member panel. The final decision in every case is made by a three-member panel either on cases the panel has heard or recommendations form [sic] the hearing officer.

\textsuperscript{44} \textit{Willock}, HRD No. 06-12-20-0685, slip op. at 4.
\textsuperscript{45} \textit{Id.} at 6.
\textsuperscript{46} \textit{Id.}
\textsuperscript{47} \textit{Id.}
\textsuperscript{48} \textit{Id.} at 10.
physical or mental handicap.”\textsuperscript{49} The Commission found that the e-mail correspondence between Willock and Huguenin established a prima facie case of discrimination in violation of the statute because Huguenin made a distinction in offering the services provided by Elane Photography based on Willock’s sexual orientation.\textsuperscript{50} Elane Photography submitted two defenses to the charge. First, Elane Photography challenged the application of section 28-1-7 to the business on the grounds that Elane Photography was not a “public accommodation.”\textsuperscript{51} Specifically, Elane Photography asserted that it was exempt from application of the statute because “a business entity of an expressive or artistic nature . . . [does] not meet the statutory definition of a ‘public accommodation’ under the [New Mexico Human Rights Act].”\textsuperscript{52} Second, Elane Photography argued that, even if it were a public accommodation subject to the New Mexico Human Rights Act, the Act was preempted by the First Amendment, which protected Elane Photography’s rights to free exercise of religion and free speech, including its right to refuse photographic service for those reasons.\textsuperscript{53} Elane Photography failed to assert any of the various exemptions to the statute expressly contained in section 28-1-9.\textsuperscript{54}

Based on its investigation, the Commission found that Elane Photography qualified as a public accommodation, because it was registered as a limited liability company, held itself open to the public by soliciting business through its website, and openly sold its services to the public.\textsuperscript{55} The Commission also rejected Elane Photography’s contention that expressive and artistic businesses are exempt from liability under section 28-1-7(F), because the statute provides no express exemption for such businesses.\textsuperscript{56}

In addressing Elane Photography’s First Amendment defenses, the Commission relied generally on Supreme Court precedent upholding the constitutionality of provisions similar

\textsuperscript{49} N.M. STAT. ANN. § 28-1-7(F) (2011). “[P]ublic accommodation’ means any establishment that provides or offers its services, facilities, accommodations or goods to the public, but does not include a bona fide private club or other place or establishment that is by its nature and use distinctly private.” \textit{Id.} § 28-1-2(H). \textit{Willock}, HRD No. 06-12-20-0685, slip op. at 10.

\textsuperscript{50} \textit{Willock}, HRD No. 06-12-20-0685, slip op. at 14.

\textsuperscript{51} \textit{Id.}

\textsuperscript{52} \textit{Id.} at 15.

\textsuperscript{53} \textit{Id.} at 14.

\textsuperscript{54} \textit{Id.}

\textsuperscript{55} \textit{Id.} at 15, 16.

\textsuperscript{56} \textit{Id.} at 15.
to section 28-1-7(F). The Commission explained that such provisions are justified because the State has a compelling interest in preventing “acts of invidious discrimination in the distribution of publicly available goods [and] services.” However, the Commission clarified that two important issues were not before the Commission for determination and were beyond the scope of the opinion. Those issues included the constitutionality of the New Mexico Human Rights Act and the preemption of the Act by the United States Constitution or other state law, including the New Mexico Constitution and the New Mexico Religious Freedom Restoration Act. Therefore, the decision of the Commission was limited to a finding that Willock had made out a prima facie antidiscrimination claim under section 28-1-7(F) and that Elane Photography failed to assert a valid exemption to the statute or otherwise rebut Willock’s showing. The Commission awarded Willock $6,637.94 in attorney’s fees and costs. Under section 28-1-11(E), Willock was entitled to actual damages under the statute as well as attorney’s fees. However, Willock declined to seek actual damages, even though she was given the opportunity to show proof of such damages at the hearing.

C. The Appeal

Elane Photography appealed the Human Rights Commission’s decision to the Second Judicial District Court of the State of New Mexico (the district court). In its appeal, Elane Photography asked the district court to reverse the Commission’s judgment because the judgment violated Elane Photography’s First Amendment rights of free exercise of religion and freedom of expression (including freedom from

57 Id. at 17.
58 Id. (quoting Roberts v. United States Jaycees, 468 U.S. 609, 628 (1984)).
59 See id. at 18.
60 See id. at 14.
61 See id.
62 See id. at 19.
63

Upon the conclusion of a hearing conducted by a hearing officer, the hearing officer shall prepare a written report setting forth proposed findings of fact and conclusions of law and recommending the action to be taken by the commission. . . . As part of its order, the commission may require the respondent to pay actual damages to the complainant and to pay reasonable attorneys’ fees.

64 Willock, HRD No. 06-12-20-0685, slip op. at 18.
Elane Photography also revived its argument that the business was not a public accommodation and therefore not subject to section 28-1-7, and it contended that its conduct was not discriminatory. 66

The district court issued its opinion on the parties’ cross motions for summary judgment on December 11, 2009. Finding no issue of material fact, the court denied Elane Photography’s motion for summary judgment and granted Willock’s motion. 67 The district court found, as a matter of law, that Elane Photography was a public accommodation. 68 Similarly, the district court affirmed the Commission’s finding that there was direct evidence of discrimination on Huguenin’s part, because Elane Photography had a policy to distinguish between opposite-sex couples and same-sex couples in providing wedding photography services. 69

Next, the district court found that the Commission’s application of the New Mexico Human Rights Act did not violate Elane Photography’s freedom of expression. 70 The district court began its analysis by distinguishing Supreme Court precedent that supported Elane Photography’s position, which stood for the proposition that various art forms, including “pictures, films, paintings, drawings, and engravings[,]” enjoyed full First Amendment protection because of their communicative nature. 71 By characterizing Elane Photography as a case dealing only with restrictions on who can buy artwork after the artist has offered it for sale, rather than with restrictions on the artwork’s dissemination, the district court found the precedent’s reasoning to be inapposite. 72 Further distinguishing the precedent, the district court pointed out the fact that Huguenin, as a hired photographer, did not choose the content of her own work, whereas the precedential cases all involved an artist whose works contained content of their own choosing or creation. 73

Like the Commission, the district court made a point of noting that nondiscrimination laws, such as New Mexico’s public

66 Id.
68 Id. at 6.
69 See id. at 8.
70 See id. at 11.
71 Id. at 8.
72 See id.
73 Id.
accommodations law, are generally constitutional.\textsuperscript{74} Further analyzing Supreme Court precedent finding state compulsion of speech unconstitutional,\textsuperscript{75} the district court found that, as a commercial photographer, Huguenin’s only message was “fine photography of special moments.”\textsuperscript{76} Thus, application of state public accommodations law was not an impermissible compulsion of an individual to affirm or disseminate the state’s ideological message, as a requirement to salute the flag or to carry the state motto on a license plate would be.\textsuperscript{77} Furthermore, the district court asserted that Huguenin was far from a communicator of artistic expression.\textsuperscript{78} Rather, the photographer was merely a conduit for her clients’ messages and thus was not afforded any constitutional protection from compelled speech, since no message of her own was affected by application of the statute.\textsuperscript{79}

II. A FIRST AMENDMENT VIOLATION

Two distinct, but related, lines of First Amendment doctrine are applicable to \textit{Elane Photography}: the compelled expression doctrine and the compelled expressive association doctrine. These doctrines independently demonstrate that the application of New Mexico’s public accommodations statute to \textit{Elane Photography}\textsuperscript{80} was an unconstitutional compulsion of speech.

A. The Compelled Speech Doctrine

Applying section 28-1-7(F) to \textit{Elane Photography} is a violation of the Huguenins’ First Amendment right to be free from compelled speech.\textsuperscript{81} It is a well-established principle of

\textsuperscript{74} Id. at 9 (quoting Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp., 515 U.S. 557, 572 (1995)).
\textsuperscript{75} See id. at 10 (citing Hurley, 515 U.S. at 566; Pac. Gas & Elec. Co. v. Pub. Util. Comm’n of Cal., 475 U.S. 1, 20-21 (1986)).
\textsuperscript{76} Id. at 11.
\textsuperscript{77} Id. at 10.
\textsuperscript{78} Id. at 11.
\textsuperscript{79} Id.
\textsuperscript{80} In the interest of clarity, for the duration of the paper, no distinction will be made between “Elane Photography” the business and “Elaine Huguenin” the individual. Because corporations have the same First Amendment rights as individuals, Citizens United v. FEC, 130 S. Ct. 876, 913 (2010), this will not affect the integrity of the legal analysis.
\textsuperscript{81} Indeed, this was Professor Eugene Volokh’s first reaction to the Human Rights Commission’s decision on April 9, 2008, even before he had read the opinion. In a post on his eponymous blog, Professor Volokh posited that, by applying the public accommodation statute to photography-as-art, the State of New Mexico may have run afoul of the First Amendment prohibition against compelled speech as articulated in \textit{Wooley v. Maynard}, 430
constitutional law that the First Amendment freedom of speech includes the right to choose what not to say. The right not to speak is most famously set forth in two Supreme Court cases, West Virginia State Board of Education v. Barnette and Wooley v. Maynard.

1. West Virginia State Board of Education v. Barnette

In Barnette, the Court struck down a West Virginia Board of Education resolution requiring all students to salute the flag and recite the Pledge of Allegiance. The first step in the Court’s analysis was to find that a flag salute and recital of the Pledge constituted expression for the purposes of the First Amendment. There was no question that the flag salute and pledge together constituted expression, because the ceremony was both “a compulsion of students to declare a belief” and a “require[d] affirmation of a belief and an attitude of mind.” That students were actually compelled to participate in the recital was necessary, and within suitably defined areas, a concomitant freedom not to speak publicly, one which serves the same ultimate end as freedom of speech in its affirmative aspect” (quoting Estate of Hemingway v. Random House, Inc., 244 N.E.2d 250, 255 (1968)).

Barnette, 319 U.S. at 626 n.2. Wooley, 430 U.S. at 714.

The text of the resolution, in pertinent part, stated:

Therefore, be it RESOLVED, That the West Virginia Board of Education does hereby recognize and order that the commonly accepted salute to the Flag of the United States is now becomes a regular part of the program of activities in the public schools, supported in whole or in part by public funds, and that all teachers as defined by law in West Virginia and pupils in such schools shall be required to participate in the salute, honoring the Nation represented by the Flag, provided, however, that refusal to salute the Flag be regarded as an act of insubordination, and shall be dealt with accordingly.
equally clear since refusal to comply was treated as insubordination and punished with expulsion pending compliance. Because the notion that the government could compel an individual to affirm an opinion or belief was anathema to the Court, the Court indicated that such a compulsion would be constitutional only if it passed an even higher standard than that applicable to government restrictions of speech.

The opinion then went on to analyze and overrule *Minersville School District v. Gobitis*, decided just three years earlier, which held that schools could condition access to public schools on participation in the flag pledge and salute. The Court rejected what it termed “the heart” of the *Gobitis* decision—the false premise that because “[n]ational unity is the basis of national security,” the government had authority to institute compulsory measures to achieve that goal. The First Amendment’s purpose is to guard against any such coercion of thought. By setting this clear boundary, the First Amendment prevents the slippery slope that begins with persuasion toward national unity and quickly devolves into compulsion of thought and then extermination of dissenters. Thus, the Court reasoned, although the state’s interest in promoting national unity was legitimate, that interest did not justify a compulsion of speech.

The *Barnette* decision therefore established the method for analyzing instances of government-compelled speech under the First Amendment. First analyze whether a law has the effect of eliciting some sort of expression, then decide whether the expression amounts to a “declaration” or “affirmation” of belief. If there are sanctions for noncompliance with the statute, an impermissible compulsion will be found and will

---

88 In fact, not only were the children punished, but parents of non-complying children were also sanctioned. The children’s absence for insubordination was treated as unlawful delinquency, for which parents were subject to a fine and jail time, if convicted. *Id.* at 629.

89 *Id.* at 633-34 (“It is now a commonplace that censorship or suppression of expression of opinion is tolerated by our Constitution only when the expression presents a clear and present danger of action of a kind the State is empowered to prevent and punish. It would seem that involuntary affirmation could be commanded only on even more immediate and urgent grounds than silence. But here the power of compulsion is invoked without any allegation that remaining passive during a flag salute ritual creates a clear and present danger that would justify an effort even to muffle expression.”).

90 *Id.* at 642; see *Minersville Sch. Dist. v. Gobitis*, 310 U.S. 586, 595 (1940).

91 *Gobitis*, 310 U.S. at 595; *Barnette*, 319 U.S. at 640.

92 *Barnette*, 319 U.S. at 642.

93 *Id.* at 641.

94 *Id.* at 640.
possibly be an even greater First Amendment harm than a restriction of speech.\textsuperscript{95}

2. \textit{Wooley v. Maynard}

In \textit{Wooley}, the Supreme Court confronted a very different set of facts than those of \textit{Barnette} but struck down a state law on much the same reasoning. As Jehovah’s Witnesses, George and Maxine Maynard felt the New Hampshire state motto, “Live Free or Die,” directly contravened their religious and moral beliefs.\textsuperscript{96} In an attempt not to disseminate the objectionable message, the Maynards began to cover the portion of their license plate where the motto was displayed.\textsuperscript{97} After being found guilty three times for violating a New Hampshire statute prohibiting the covering up of any letter or number on a state-issued license plate, Maynard brought a civil rights action under § 1983 for declaratory relief and to enjoin enforcement of this and another statute requiring that the license plates for all noncommercial vehicles bear the New Hampshire motto.\textsuperscript{98}

At the outset of its analysis, the Court framed the issue as “whether the State may constitutionally require an individual to participate in the dissemination of an ideological message by displaying it on his private property in a manner and for the express purpose that it be observed and read by the public.”\textsuperscript{99} Thus, much of the \textit{Barnette} analysis was already satisfied, since the court implicitly found the New Hampshire statute required individuals to express a message. By directly analogizing the license plate statute to the requirement in \textit{Barnette} to salute and pledge the flag, the Court reasoned that the New Hampshire statute co-opted the Maynards’ private property as a “mobile billboard” for the state’s own message.\textsuperscript{100} Even if passively carrying the license plate on one’s car was not as great a First Amendment harm as requiring active affirmation of a belief through speech and conduct, the Court nevertheless found the requirement was not constitutional.\textsuperscript{101}

As in \textit{Barnette}, the statutes at issue carried a penalty for

\textsuperscript{95} \textit{Id.} at 633.
\textsuperscript{97} \textit{Id.} at 707-08.
\textsuperscript{98} The statute prohibiting the covering up of any letters or numbers was interpreted to include the State motto. \textit{Id.} at 708-09.
\textsuperscript{99} \textit{Id.} at 713.
\textsuperscript{100} \textit{Id.} at 715.
\textsuperscript{101} \textit{Id.} at 717.
noncompliance.\footnote{Appellee George Maynard was issued a citation for cutting the words "or Die" off his plate and taping over the resulting hole as well as the words "Live Free." At a hearing, a trial judge fined him $25, but suspended the fine so long as he complied with the statute going forward. Maynard was fined $50, ordered to pay the original $25 fine, and sentenced to fifteen days in jail upon violating the statute a second time. Id. at 708.} Also as in Barnette, the Court rounded out its reasoning by inquiring into “whether the State’s countervailing interest is sufficiently compelling to justify requiring appellees to display the state motto on their license plates.”\footnote{Id. at 716.} The Court determined the state’s first interest—easily identifying passenger vehicles—did not justify the infringement on drivers’ rights because such an interest could be achieved by more narrowly tailored means that did not so “broadly stifle fundamental personal liberties.”\footnote{Id.} More importantly, the state’s second interest—fostering state pride—did not justify an infringement of rights because, “where the State’s interest is to disseminate an ideology, no matter how acceptable to some, such interest cannot outweigh an individual’s \textit{First Amendment} right to avoid becoming the courier for such message.”\footnote{Id. at 717.}

3. \textit{Elane Photography} Is a Case of Compelled Speech

As Barnette and Wooley explain, the analysis for the compelled speech doctrine has two steps. First, analyze whether a state law, regulation, or policy compels citizens to express or affirm a belief that they do not themselves hold. Then, ask whether the state’s interest in enforcing that law is compelling so that it justifies such a great constitutional harm. Under this analysis, it is likely that forcing Huguenin to take photographs of ceremonies that she believes are inherently wrong is a form of compelled speech. The Supreme Court recognizes artistic expression, including photography, as protected by the First Amendment, and it is unclear whether any legitimate state interest would justify this compulsion.

\textit{a. Photography as Speech}

As a threshold matter, it must be determined whether there is expression compelled by a statute that compels expression in a way prohibited by the First Amendment.\footnote{See W. Va. State Bd. of Educ. v. Barnette, 319 U.S. 624, 632 (1943).} The
protection of speech on political issues or issues of public concern is at the core of the First Amendment.107 The fight over marriage is not just over sincere religious and moral beliefs but also about entitlement to “legal, financial, and social benefits” the government affords married couples.108 The issue of marriage itself has been a public issue since before the United States existed.109 Further, the growing national Defense of Marriage movement opposing same-sex marriage in recent years demonstrates that same-sex marriage specifically is an issue of political concern.110 In 2003, Massachusetts became the first state to allow marriages between individuals of the same sex.111 Currently, a majority of states, including New Mexico, do not allow same-sex marriage.112 In the past decade, there have been several high-profile attempts to overturn bans, and to pass new legislation achieving marriage equality for same-sex couples, several of which were successful.113 Therefore,

107 Frisby v. Schultz, 487 U.S. 474, 479 (1988); Mills v. Alabama, 384 U.S. 214, 218 (1966) (“Whatever differences may exist about interpretations of the First Amendment, there is practically universal agreement that a major purpose of that Amendment was to protect the free discussion of governmental affairs.”); Roth v. United States, 354 U.S. 476, 484 (1957) (The First Amendment “was fashioned to assure unfettered interchange of ideas for the bringing about of political and social changes desired by the people.”).


111 Goodridge, 798 N.E.2d at 948 (finding that the Massachusetts Constitution prohibits treating same-sex couples differently than opposite-sex couples for the purposes of the state’s marriage statute).

112 New Mexico currently has no provision addressing same-sex marriage. See N.M. STAT. ANN. §§ 40-1-1 to -4 (2010). On January 4, 2011, the Office of New Mexico Attorney General Gary King released an opinion letter concluding that, “While we cannot predict how a New Mexico court would rule on this issue . . . it is our opinion that a same-sex marriage that is valid under the laws of the country or state where it was consummated would likewise be found valid in New Mexico.” N.M. Validity for Same-Sex Marriages Performed in Other Jurisdictions, Op. N.M. Att’y Gen., Gary K. King, No. 11-01 (2011), available at http://www.nmag.gov/pdf/4%20Jan%202011-Rep.%20Al%20Park-Opinion%2011-01%5B1%5D.pdf.

113 As of March 2012, eight states and the District of Columbia allow same-sex marriage: Connecticut, Iowa, New Hampshire, Maryland, Massachusetts, New York, Vermont, and Washington. Other states prohibit same-sex unions, but provide all or some legal benefits of marriage to same-sex couples. In February 2012, the United States Court of Appeals for the 9th Circuit declared California’s ban on same-sex marriage unconstitutional. Perry v. Brown, 2012 U.S. App. LEXIS 2328 (9th Cir. Feb. 7, 2012); see
expression on the issue of same-sex marriage deserves full First Amendment protection.\textsuperscript{114}

That the expression is in the form of photography, or, rather, a decision not to photograph, does not lessen the expression’s degree of protection. It has long been recognized that the Constitution protects various forms of expression, including most art forms.\textsuperscript{115} Photography specifically has been identified as art that falls within the protection of the First Amendment.\textsuperscript{116} Further, the Supreme Court has decided that, in addition to its protected status as a medium for the content it expresses, art is protected for its own sake because of its inherent expressive character.\textsuperscript{117} As the Court has explained,

\begin{quotation}
[C]onstitutional protection of artistic works turns not on the political significance that may be attributable to such productions, though they may indeed comment on the political, but simply on their expressive character, which falls within a spectrum of protected “speech” extending outward from the core of overtly political declarations. Put differently, art is entitled to full protection because our “cultural life,” just like our native politics, “rest[s] upon [the] ideal” of governmental viewpoint neutrality.\textsuperscript{118}
\end{quotation}

Thus, Elaine Huguenin’s photographs should be protected as expression under the First Amendment for conveying a message on a prominent social and political issue and as artistic works.\textsuperscript{119}

Further, the mere fact that clients may commission and pay for Huguenin’s photography does not diminish its expressive ability or First Amendment protection.\textsuperscript{120} There is a

\begin{quotation}
\end{quotation}

\begin{quotation}
\textsuperscript{114} Morse v. Frederick, 551 U.S. 393, 403 (2007) (“Political speech, of course, is ‘at the core of what the First Amendment is designed to protect.’” (quoting Virginia v. Black, 538 U.S. 343, 365 (2003) (plurality opinion))).
\end{quotation}

\begin{quotation}
\textsuperscript{115} Nat’l Endowment for the Arts v. Finley, 524 U.S. 569, 602-03 (1998) (Souter, J., dissenting) (“It goes without saying that artistic expression lies within this First Amendment protection.”).
\end{quotation}

\begin{quotation}
\textsuperscript{116} Kaplan v. California, 413 U.S. 115, 119-20 (1973) (“Pictures, films, paintings, drawings, and engravings . . . have First Amendment protection.”).
\end{quotation}

\begin{quotation}
\textsuperscript{117} Finley, 524 U.S. at 602-03 (Souter, J., dissenting).
\end{quotation}

\begin{quotation}
\textsuperscript{118} Id. at 602-03 (quoting Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 641 (1994)).
\end{quotation}

\begin{quotation}
\textsuperscript{119} Id.
\end{quotation}

\begin{quotation}
\textsuperscript{120} N.Y. Times Co. v. Sullivan, 376 U.S. 254, 266 (1964) (holding that otherwise constitutionally protected expression does not lose its First Amendment protection merely because it was bought and paid for); Griffin, supra note 4, at 1062 (“The Supreme Court clearly has rejected the significance of profit motive to [First Amendment] claims, indicating that this basis for regulation would be incompatible
long tradition of patronage in the arts and some of the most lauded classical artwork was produced on commission. Nor are the photographs subject to lesser constitutional protection as mere commercial speech. Commercial speech is speech that proposes a commercial transaction, such as flyers advertising the sale of goods, not speech that was commissioned from a provider by a client.

Because art, even if bought and commissioned, has protection as expression under the First Amendment, the facts of Elane Photography likely lead to a compelled speech problem under Barnette and Wooley. To be sure, one could argue that commercial photographers merely capture a memorable moment and contain little, if any, actual expression. But that argument contradicts both case law and reason. As discussed, the Supreme Court has emphasized that art is an inherently expressive medium, whatever the purpose behind its creation. The fact with the first amendment. That conclusion requires only the recognition that newspapers and books are sold for profit...
that there is comparatively more expression in an Ansel Adams photograph than in one taken by Elaine Huguenin, in any case debatable, is a question of degree that does not change the fact that expression exists, even if it is just the expression of a celebratory moment from a particular perspective.  

One could also argue that the expression is not the photographer’s own message, but rather that of the clients or subjects of the photograph. It is certainly the case that clients have their own views of the celebrated event, and it is possible that the photographer incorporates those views into her photographs as she takes them. But that is not necessarily the case, nor does that preclude the photographer’s own expression from being simultaneously produced. The difficulty in commercialized-art-public-accommodations cases is that the main purpose of the expression is to provide a service to clients. However, to raise the question is not to answer it. Providing a service and creating expression are not mutually exclusive activities. It may even be the case that clients seek particular service providers specifically because the messages they project through their services is one the clients desire to support. This is a common phenomenon, observable in every neighborhood where environmentally friendly dry cleaners and free trade coffee shops flourish. Whether it is the case here that Huguenin’s photography contained expression that was her own depends on the extent of the artistic involvement in creating the finished image. Huguenin asserts in her appeal to the district court that “it takes great skill, planning and aesthetic judgment to create a photograph,” which is probably accurate to a greater or lesser extent, depending on the circumstances. It stands to reason that the photographer exercises the skill, planning, and aesthetic judgment when a photograph is taken, rather than the commissioner of the

---

125 Volokh, supra note 81.

126 See Richard A. Epstein, Articles and Essays: The Constitutional Perils of Moderation: The Case of the Boy Scouts, 74 S. CALIF. L. REV. 119, 140 (2000) (arguing that it is nearly impossible to draw the line between expressive and non-expressive corporations today); id. ("It is sheer fantasy to assume that any successful organization fits this odd caricature of the firm [whose sole goal is profit], and is wholly indifferent to how it is perceived in the external world or by its own staff. It is commonplace to speak of 'corporate cultures' and to understand that these refer to the way in which particular firms position themselves in the many markets, internal and external, in which they do business.").

photograph or its subject. If Huguenin actively formed the content of the photograph through her artistic manipulation of the medium, then the photograph likely contained her own expressive interpretation of the scene. Thus, Huguenin was probably not merely a conduit for her clients’ messages, as the district court found.\footnote{Elane Photography, LLC v. Willock, No. CV-2008-06632, 11 ¶ 25 (N.M.2d Jud. Dist. Ct. 2009).}

Finally, the fact that Huguenin’s expressive photography does not itself contain a specific message is not an issue under the compelled speech doctrine. The violation occurs when an individual is forced to utter or affirm a belief not her own.\footnote{See generally Wooley v. Maynard, 430 U.S. 705 (1977); W. Va. State Bd. of Educ. v. Barnette, 319 U.S. 624 (1943).} There is no requirement that the individual has already expressed a view to the contrary.

Other facts of the case further support the conclusion that the expression in the photographs was Huguenin’s rather than her clients’. Not only did Huguenin produce the physical photos from the negatives she shot, Huguenin also retained ownership rights to all the images in the photographs.\footnote{Willock, HRD No. 06-12-20-0685, slip op. at 4 (quoting testimony of Elaine Huguenin and Jonathan Huguenin; Ex. A), available at http://volokh.com/files/willockopinion.pdf.} The images themselves were like her personal property, similar to George Maynard’s car in \textit{Wooley}. To force Huguenin to include images that conflict with her beliefs in her photographs is similar to forcing a driver to carry an unwanted ideological message on his license plate.\footnote{Wooley, 430 U.S. at 713.} The analogy is not perfect because, unlike in \textit{Wooley}, the expression itself is being co-opted, rather than an object with no inherently expressive nature. Therefore, the facts are closer to \textit{Barnette} where the violation was based on active expression of a contrary belief, rather than \textit{Wooley}’s more passive, forced dissemination. Requiring a photographer, on pain of civil sanction, to portray certain events in a positive light that she believes should not be so portrayed is essentially forcing her to support or adopt an idea not her own. This is similar to obliging a child to affirm a belief through recitation of a pledge, on pain of punishment, with which she does not agree.\footnote{Even if taking photographs involved no artistic skill or choices of light effect, perspective, angle, speed or layout, generic retail photography may still be protected by the First Amendment. Professor Eugene Volokh explains:} Because of this compelled
action, applying the public accommodations law may be an even greater infringement on Huguenin’s First Amendment rights than that found in *Wooley*.133

**b. State’s Interest in Compelling Speech**

The second prong of the compelled speech analysis investigates whether the state interest served by the infringing action justifies such a grave First Amendment violation.134 The Court has held that a state’s legitimate interest in promoting national unity,135 or in facilitating state enforcement of its laws,136 cannot justify compelling expression. Indeed, “where the State’s interest is to disseminate an ideology, no matter how acceptable to some, such interest cannot outweigh an individual’s First Amendment right to avoid becoming the courier for such message.”137 Because enforcing the New Mexico public accommodations law against Elane Photography would coerce individual promotion of the state’s ideological message, it would be absurd if the state’s interest in disseminating that ideological message could justify the constitutional violation that coercive dissemination would cause.138 To hold otherwise would undermine the primary purpose behind the First

---


133 *Wooley*, 430 U.S. at 713; id. at 715 (“Compelling the affirmative act of a flag salute involved a more serious infringement upon personal liberties than the passive act of carrying the state motto on a license plate, but the difference is essentially one of degree.”).

134 Id. at 716 (After finding the petitioner’s First Amendment rights are implicated, the Court “must also determine whether the State’s countervailing interest is sufficiently compelling to justify requiring [the compelled expression].”).


136 *Wooley*, 430 U.S. at 716.

137 Id. at 717.

Amendment—to protect expression of individual belief, irrespective of content, against government interference.139 Nevertheless, in the closely analogous compelled association cases, the Supreme Court has found that the interest of eliminating discrimination is sufficient to justify infringement on First Amendment rights.140 The elimination of discrimination is a primary purpose of public accommodations laws,141 and it was the compelling justification cited by the New Mexico Human Rights Commission for rejecting Huguenin’s First Amendment defenses.142 Although never explicitly overruled, the current status of the compelling interest test is in question following several important cases on the closely related expressive association doctrine.143

B. The Compelled Association Doctrine

Although not an enumerated right, the Supreme Court has found “implicit in the right to engage in activities protected by the First Amendment a corresponding right to associate with others in pursuit of a wide variety of political, social, economic, educational, religious, and cultural ends.”144 Without such protection for group effort, the Court reasoned, an individual’s other First Amendment freedoms would be significantly diminished.145 Just as the Court found that freedom

139 Boy Scouts of Am. v. Dale, 530 U.S. 640, 658-59 (2000); Whitney v. California, 274 U.S. 357, 375 (1927) (Brandeis, J., concurring) (the purpose of the First Amendment is to preserve a democratic form of government by ensuring government does not have the power to repress the public by coercing acceptance of a government approved message), overruled in part on other grounds, Brandenburg v. Ohio, 274 U.S. 357 (1927).
140 Bd. of Dirs. of Rotary Int’l v. Rotary Club of Duarte, 481 U.S. 537, 549 (1987) (“Even if the [public accommodations statute] does work some slight infringement on [the association’s] right of expressive association, that infringement is justified because it serves the State’s compelling interest in eliminating discrimination against women.”); Roberts v. United States Jaycees, 468 U.S. 609, 628 (1984) (“[A]cts of invidious discrimination in the distribution of publicly available goods, services, and other advantages cause unique evils that government has a compelling interest to prevent—wholly apart from the point of view such conduct may transmit. Accordingly, like violence or other types of potentially expressive activities that produce special harms distinct from their communicative impact, such practices are entitled to no constitutional protection.”).
141 Lerman & Sanderson, supra note 2, at 238-40.
143 Bernstein, supra note 138, at 116.
144 Roberts, 468 U.S. at 622.
145 Id. (“According protection to collective effort on behalf of shared goals is especially important in preserving political and cultural diversity and in shielding dissident expression from suppression by the majority.”); Boy Scouts of Am. v. Dale, 530
of expression corresponded with a right to be free from compelled expression, so has the Court found a corresponding right of freedom from compelled association. Thus, the Court views the right of freedom of association as a necessary corollary to the enumerated First Amendment rights, and the compelled association doctrine is directly analogous to the compelled speech doctrine because the Court derived both from the same rationale.

The Supreme Court’s most significant freedom of association cases involve the application of state public accommodations laws to expressive activities that result in compelled association problems. The three most instructive cases in this area are Roberts v. United States Jaycees, Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston, and Boy Scouts of America v. Dale.

1. Roberts v. United States Jaycees

The United State Jaycees, a nonprofit educational and charitable membership organization, limited their regular membership to men ages eighteen to thirty-five. Women were admitted as nonvoting associate members only. In the mid-1970s, two Minnesota chapters, St. Paul and Minneapolis,

---

U.S. 640, 647-48 (2000) (“This right is crucial in preventing the majority from imposing its views on groups that would rather express other, perhaps unpopular, ideas.”).


147 Dale, 530 U.S. at 648; Roberts, 468 U.S. at 623.


149 See Barnette, 319 U.S. at 641 (explaining that the First Amendment was designed to protect against attempts of the government or a majority from imposing on a minority any “uniformity of sentiment”).

150 See Hurley v. Irish-Am. Gay, Lesbian & Bisexual Grp., 515 U.S. 557, 572-73 (1995) (holding that the Massachusetts public accommodations law was unconstitutional as applied to the sponsor of Boston’s annual St. Patrick’s Day parade because the parade’s sponsor could not be compelled to include any group whose message would alter the expression of the parade as a whole); Dale, 530 U.S. at 648 (finding unconstitutional the application of the New Jersey public accommodations law to a private non-profit civic and educational association because “[t]he forced inclusion of an unwanted person in a group infringes the group’s freedom of expressive association if the presence of that person affects in a significant way the group’s ability to advocate public or private viewpoints”); Roberts, 468 U.S. at 623 (holding that the Minnesota public accommodations law infringed upon the civic organization’s First Amendment freedoms when it compelled the organization to admit women as members, but the infringement was justified because the burden on the members’ speech was incidental when compared with the State’s compelling interest in abolishing sex discrimination.).


152 Roberts, 468 U.S. at 613.

153 Id.
began to admit women as regular voting members in violation of the organization’s bylaws. After the national organization imposed various sanctions and threatened to revoke the offending chapters’ charters, the chapters filed discrimination suits with the Minnesota Department of Human Rights. The local chapters alleged a violation of the Minnesota public accommodations law, which prohibited the denial to “any person the full and equal enjoyment of the goods, services, facilities, privileges, advantages, and accommodations of a place of public accommodation because of race, color, creed, religion, disability, national origin or sex.”

Reversing a decision by the Court of Appeals for the Eight Circuit, the Supreme Court found that the Minnesota public accommodations law was constitutionally applied to the national organization to require the admission of women as full voting members. To reach that conclusion, the Court inquired whether applying this statute to the Jaycees would infringe upon the group’s First Amendment rights, and found there was no infringement because admitting women would not in any way impede the Jaycees from disseminating their views. However, the Court clarified that a finding of infringement does not end the analysis; infringement on a group’s right to expressive association may be “justified by regulations adopted to serve compelling state interests, unrelated to the suppression of ideas, that cannot be achieved through means significantly less restrictive of associational freedoms.” After this pronouncement, the Court quickly concluded that the compelling interest of eradicating sex discrimination served by Minnesota’s public accommodations law justified any imposition on the Jaycees.


In Hurley, the Supreme Court unanimously held that the Massachusetts public accommodations law was unconstitutional.

---

154 Id. at 614.
155 Id. at 621-22, 627.
156 Id. at 621-22, 627.
157 Id. at 623.
158 Id. at 623.
159 Id. at 623.
160 Id.
as applied to a private, unincorporated group that organized Boston’s yearly St. Patrick’s Day-Evacuation Day Parade.161 The private council of veterans was authorized to organize the yearly parade by Boston’s mayor in 1947.162 In 1992, the respondents formed the Irish-American Gay, Lesbian and Bisexual group of Boston (GLIB) in order to join the parade “to express pride in their Irish heritage as openly gay, lesbian, and bisexual individuals, to demonstrate that there are such men and women among those so descended, and to express their solidarity with like individuals who sought to march in New York’s St. Patrick’s Day Parade.”163 The council denied GLIB’s application to join the parade, but GLIB marched anyway pursuant to a state court order granting the group the right to participate.164 When GLIB’s application was denied again in 1993, the group brought suit against the council under Massachusetts’s public accommodations law.165

Like the analyses in the compelled speech cases, the Court first identified that there was expression involved entitled to First Amendment protection.166 It is important to note that although the parade did not have any specific message to convey, the Court held its expression was nevertheless protected.167 The fact that the organizers did not originate the message that each contingent expressed was not a reason for denying constitutional protection to the parade as a whole, because the mere act of selecting each piece for inclusion was sufficient to impart constitutional protection.168 A key part of the Court’s analysis involved a discussion of the history of public accommodations laws, and concluded that “[p]rovisions like these are well within the State’s usual power to enact when a legislature has reason to believe that a given group is the target of discrimination, and they do not, as a general matter, violate the First or Fourteenth Amendments.”169 Additionally, the Court

---

162 Id. at 560.
163 Id. at 561.
164 Id.
165 Id.
166 Id. at 568-69.
167 Id. at 569.
168 Id. at 570 ("Nor, under our precedent, does First Amendment protection require a speaker to generate, as an original matter, each item featured in the communication. Cable operators, for example, are engaged in protected speech activities even when they only select programming originally produced by others.").
169 Id. at 572.
found the statute at issue unproblematic, because it was neither directed at speech nor a content-based restriction of speech.\footnote{170}

However, the Court found the statute was nevertheless unconstitutional as applied since it would compel the parade organizers to change the content of their expression.\footnote{171} The Court explained that, “[s]ince every participating unit affects the message conveyed by the private organizers, the state courts’ application of the statute produced an order essentially requiring petitioners to alter the expressive content of their parade.”\footnote{172} The state improperly converted the organizers’ expression into the public accommodation, rather than apply the statute properly.\footnote{173} A different result would have obtained had GLIB shown that the organizers excluded its members or other people from participating in the approved parade units on the basis of a protected classification (sexual orientation).\footnote{174} Rather, the issue was whether they could be excluded as a group because of their message.\footnote{175} To apply a public accommodations statute in this manner infringes on an essential First Amendment principle, “that a speaker has the autonomy to choose the content of his own message.”\footnote{176}

The Hurley Court also explicitly rejected GLIB’s argument that the parade was merely a conduit for the participants’ messages, which does not have a right to free expression, rather than expression in and of itself.\footnote{177} Finally, and most importantly, the Court conspicuously avoided all mention of the Roberts compelling interest analysis.\footnote{178} In fact,

\footnote{170} Id. (“Nor is this statute unusual in any obvious way, since it does not, on its face, target speech or discriminate on the basis of its content, the focal point of its prohibition being rather on the act of discriminating against individuals in the provision of publicly available goods, privileges, and services on the proscribed grounds.”).

\footnote{171} Id. at 573.

\footnote{172} Id. at 572-73.

\footnote{173} Id. at 573.

\footnote{174} Id. at 572.

\footnote{175} Id.

\footnote{176} Id. at 573.

\footnote{177} Id. at 575.

\footnote{178} Roberts v. United States Jaycees is cited by the Hurley Court only four times in the entire opinion, never once relying on the Roberts analysis: twice in the Court’s summary of the case’s previous history in the Massachusetts trial court and Supreme Judicial court (at 563 and 565); once for the proposition that public accommodations laws are generally constitutional (at 572); and a last time to support the Court’s distinguishing of a different compelled association case, New York State Club Ass’n v. City of New York, 487 U.S. 1 (1988). Hurley, 515 U.S. at 580. The lack of the Court’s reliance on Roberts left open two possible interpretations of the continued validity of Roberts. One possibility was that the Court felt expressive association deserved the same First Amendment protection as pure expression, in which case Roberts and its “compelling interest” test were essentially overruled. Alternatively, it could merely have signified that the Court saw Hurley as a pure expression case and,}
the Court explicitly rejected the idea that a public accommodations law could be constitutionally applied to restrict expression and declared that, “While the law is free to promote all sorts of conduct in place of harmful behavior, it is not free to interfere with speech for no better reason than promoting an approved message or discouraging a disfavored one, however enlightened either purpose may strike the government.”

3. Boy Scouts of America v. Dale

In Dale, the Supreme Court again considered whether a state public accommodations law could be constitutionally applied to compel association. This time the issue was whether New Jersey could apply its public accommodations law to compel a national nonprofit educational association to keep on as a member an individual whose message was at odds with the views the organization wished to express. After ten years as a model Boy Scout, James Dale’s application to become an adult member and assistant scout-master was approved. Shortly thereafter, Dale openly acknowledged his homosexuality, became involved with the on-campus Lesbian/Gay Alliance at his university, and gave an interview in a local newspaper discussing his efforts to address “homosexual teenagers’ need for gay role models.” Following publication of this interview, Dale’s adult membership in the Scouts was revoked. Dale sued under New Jersey’s public accommodations statute, which prohibits discrimination in places of public accommodation on the basis of sexual orientation.

therefore, Roberts simply did not apply. In Dale, the Court resolved the confusion in favor of the first interpretation by clarifying that, where a First Amendment violation has been found in the expressive association context, Hurley, not Roberts, controls. Boy Scouts of Am. v. Dale, 530 U.S. 640, 659 (2000); Bernstein, supra note 138, at 118.  

179 Hurley, 515 U.S. at 579. This statement is not irreconcilable with Roberts. The Roberts analysis permitted application of a public accommodations law so long as the resulting infringement was “justified by regulations adopted to serve compelling state interests, unrelated to the suppression of ideas, that cannot be achieved through means significantly less restrictive of associational freedoms.” Roberts v. United States Jaycees, 468 U.S. 609, 623 (1984). Thus, if a law interferes with speech for the sole purpose of promoting a state-sponsored message, the Court seems to imply by the above statement in Hurley, the law would not satisfy the “unrelated to the suppression of ideas” requirement of the Roberts standard. Id. Nevertheless, the Court did not apply the Roberts reasoning to Hurley.  

180 Dale, 530 U.S. at 644.  
181 Id.  
182 Id. at 645.  
183 Id.  
184 Id.
The Court began its analysis with the established principle that “forced inclusion of an unwanted person in a group infringes the group’s freedom of expressive association if the presence of that person affects in a significant way the group’s ability to advocate public or private viewpoints.”\textsuperscript{185} Finding that the Boy Scouts of America was engaged in expression protected by the First Amendment, the Court then considered whether inclusion of an individual with contrary views would burden the organization’s expression.\textsuperscript{186} The Court found that it did.\textsuperscript{187}

At this point, rather than engaging in the \textit{Roberts} compelling interest balancing analysis, the Court distinguished \textit{Roberts} by deemphasizing the extent to which its holding relied upon the compelling interest test.\textsuperscript{188} The Court insisted the \textit{Roberts} decision rested on the fact that no “serious burden on the male members’ freedom of expressive association” was demonstrated.\textsuperscript{189} Because the burden on the group’s First Amendment rights was not serious, it was proper to take the state’s interest in enforcing the statute into account.\textsuperscript{190} However, in this case, applying the public accommodations statute to the Boy Scouts did infringe the group’s associational rights, and thus \textit{Hurley}’s “traditional First Amendment analysis” controlled.\textsuperscript{191} Just as in \textit{Hurley}, the state’s interest could not overcome the “severe intrusion on the Boy Scouts’ right to freedom of expressive association.”\textsuperscript{192}

Finally, in responding to an argument made by the dissenters, the Court took care to emphasize that the content of the Boy Scouts’ message did not influence the majority’s conclusion.\textsuperscript{193} The Court echoed \textit{Hurley}, explaining that,

\textquote{Public or judicial disapproval of a tenet of an organization’s expression does not justify the State’s effort to compel the organization to accept members where such acceptance would derogate from the organization’s expressive message. “While the law is free to promote all sorts of conduct in place of harmful behavior, it is not free to interfere with speech for no better reason than

\begin{itemize}
\item \textsuperscript{185} Id. at 648.
\item \textsuperscript{186} Id.
\item \textsuperscript{187} Id. at 656.
\item \textsuperscript{188} Id. at 657-58; Bernstein, supra note 138, at 111.
\item \textsuperscript{189} Roberts v. United States Jaycees, 468 U.S. 609, 626 (1984).
\item \textsuperscript{190} Dale, 530 U.S. at 658-59 (the organization’s “interest in freedom of expression has been set on one side of the scale, and the State’s interest on the other”).
\item \textsuperscript{191} Id.
\item \textsuperscript{192} Id.
\item \textsuperscript{193} Id. at 661.
promoting an approved message or discouraging a disfavored one, however enlightened either purpose may strike the government.” 194

Despite the fact that there were four dissenters, none of them appear to disagree with the majority over dropping the compelling interest analysis for cases where public accommodations law present First Amendment problems. 195 Rather, the dissenters disagreed with the majority over whether the association expressed a message, its rights were infringed, and its message impaired by compelling the group to keep Dale on as an employee. 196

4. The Compelled Association Doctrine Further Demonstrates a First Amendment Violation in Elane Photography

By applying the principles that evolved in Roberts, Hurley, and Dale, it is evident that the New Mexico public accommodations statute cannot be constitutionally applied to Elane Photography. 197

a. Roberts Distinguished

Even if Roberts remains good law, which has been questioned after Hurley and Dale, 198 its reasoning cannot be applied to Elane Photography. Although the Court found the Jaycees engaged in protected expression, there was “no basis in the record for concluding that admission of women as full voting members will impede the organization’s ability to engage in these protected activities or to disseminate its preferred views.” 199 In Elane Photography, however, application of the New Mexico statute would directly impede Huguenin’s ability to disseminate her preferred views; it would force the photographer to affirm and possibly even endorse an ideology 200

195 Id. at 665; Bernstein, supra note 138, at 125-26.
198 Bernstein, supra note 138, at 124.
199 Roberts v. United States Jaycees, 468 U.S. 609, 627 (1984). Many commentators take issue with the Court’s assertion that admitting women to an all male organization would not materially alter that organization’s stance on social or political issues. Bernstein, supra note 138, at 97.
200 See supra Part II.A.3.a.

\textbf{b. Hurley and Dale Control the Outcome of Elane Photography}

In \textit{Hurley}, the Court began its investigation into whether application of the Massachusetts public accommodations law violated the parade organizers’ First Amendment rights with the observation that the statute had been applied to the parade “in a peculiar way.”\footnote{\textit{Hurley}, 515 U.S. at 572.} Rather than seeking access to the parade as participants in the organizers’ message, the group sought to have their own message included as part of the parade over the organizers’ objections.\footnote{Id.} This application of the statute “had the effect of declaring the sponsors’ speech itself to be the public accommodation.”\footnote{Id. at 573.} To condone such an application of the statute would eviscerate the axiomatic First Amendment principle “that a speaker has the autonomy to choose the content of his own message.”\footnote{Id.} To apply the New Mexico statute to Huguenin’s photographs would have the same unconstitutional effect of co-opting expression itself as the public accommodation, rather than the retail photography service Huguenin provides.

Also in \textit{Hurley}, the Court explicitly found that no individual petitioner was excluded from participation in the parade; the petitioners were only excluded to the extent they sought to alter the parade sponsors’ message.\footnote{Id.} Similarly, there was no showing the Huguenins denied service to individuals on discriminatory grounds.\footnote{Id.} Rather, the Huguenins asserted, they purposefully selected the content of their photographs so that their expression did not promote activities that conflict with their beliefs.\footnote{Appeal from the Decision and Final Order of the N.M. Human Rights Comm’n at 2, Elane Photography, LLC.} For example, the
Huguenins claim they refuse to take any “photographs that present abortion or horror movies or pornography in a favorable light.” Thus the Huguenins, like the parade organizers in Hurley, were merely exercising their First Amendment right to select the content of their expression and not discriminating in providing service on an impermissible basis.

The district court attempted to distinguish Hurley by arguing that Huguenin’s free speech rights were not implicated because her own message was not being co-opted. Rather, the district court claimed that Huguenin was merely “a conduit or an agent for its clients.” This counterargument was rejected by the court in Hurley and should be rejected here as well. Huguenin asserts that she does not simply point and shoot her camera to convey her “client’s message of a day well spent.” The court need not take her word for it; it stands to reason that this assertion must be at least partially true. If anyone could arrange the composition, lighting, and angles of a photograph, and if there was only one possible way to take a picture such that only one possible picture of any give scene existed, no one would ever hire a professional photographer. But arranging composition, lighting, and angles is not always a simple matter, and there are an infinite number of possible pictures to take. Thus, there must be at least some skill and selection to the process that makes the district court’s conduit analogy inappropriate. The selection of what to include and how to include it is more like the parade organizers exercise of editorial control over the content of the parade “upon which the State can not intrude.” In fact, Huguenin’s control over her photographs may be even greater than the organizer’s control over the parade because Elane Photography explicitly retains the rights to all photographs taken. Thus, the “conduit” analogy is inappropriately applied to Elane Photography.

---

210 Id.
211 Hurley, 515 U.S. at 570.
213 Willock, CV-2008-06632, at 11.
214 Hurley, 515 U.S. at 575.
215 Appeal from the Decision and Final Order of the N.M. Human Rights Comm’n, at 1-2, Elane Photography, LLC; see also Willock, CV-2008-06632, at 11.
216 Hurley, 515 U.S. at 575.
218 Hurley, 515 U.S. at 575.
Because both speakers exercise control over the content of their messages, coercing inclusion of a message not their own would cause the resulting expression to be perceived as “worthy of presentation and quite possibly of support.” As already explained, such a consequence would run afoul of the First Amendment in the most dangerous way.

Similarly, in Dale, the Court confirmed that where application of a public accommodations law would significantly burden First Amendment rights, the Court should apply traditional First Amendment analysis rather than the compelling interest balancing of Roberts. Thus, once a “severe intrusion” on First Amendment rights was identified, the Court found the infringement could not be justified, even by the state’s interest in enforcing its public accommodations law. Because the Boy Scouts believed Dale’s views were “inconsistent with the values it seeks to instill in its youth members,” forcing the organization to include him would “surely interfere with the Boy Scout’s choice not to propound a point of view contrary to its beliefs.” The Huguenins asserted that they held a similar belief, which if sincerely held, leads to a direct First Amendment infringement since they would be forced to promote or at least portray in a positive light ideas inconsistent with those beliefs.

The Court’s opinion in Dale is interesting for deferring to the Boy Scouts views regarding its own expression, declining to investigate into either “the nature of its expression” or the group’s “view of what would impair its expression.” Deferring to the organization claiming infringement certainly makes the analysis in difficult cases simpler. There would have been two opinions in favor of Elane Photography had the Human Rights Commission and the district court deferred to Huguenin’s assertions that she “believes that she implicitly endorses the

219 Id.
220 Id. at 576 (“[W]hen dissemination of a view contrary to one’s own is forced upon a speaker intimately connected with the communication advanced, the speaker’s right to autonomy over the message is compromised.”).
222 Id.
223 Id. at 654.
225 Dale, 530 U.S. at 653.
viewpoints communicated by her photography,” and that it “compels Elane Photography to participate in and advance a viewpoint it would not do so absent government coercion.” Justice Stevens, in his dissent in *Dale*, criticized this deference as “an astounding view of the law” that will lead to nothing less than “a free pass out of antidiscrimination laws.” Yet, it is probably necessary to sufficiently protect First Amendment rights. As the Supreme Court unanimously stated in *Hurley*, “a narrow, succinctly articulable message is not a condition of constitutional protection.” Requiring an individual to have developed beliefs and already expressed them, either in writing or through other expressive means, may not be sufficiently protective of Free Speech. Nevertheless, the facts of *Elane Photography* do not require so much deference that plausibility is strained. It is not unreasonable to infer that a religious individual sincerely holds beliefs at odds with those that would be promoted by participating in and helping to celebrate a same-sex wedding.

Finally, neither *Hurley* nor *Dale* can be distinguished on the grounds that the cases involved First Amendment infringements on nonprofit organizations, while Elane Photography is a for-profit business. First of all, and most obviously, the United States Jaycees is also a nonprofit organization. Thus, even if *Elane Photography* could be distinguished from *Hurley* and *Dale* on this basis, *Roberts* would be similarly distinguished. Moreover, the organizations in all three cases were either expressly or impliedly deemed to be public accommodations. Thus, since Elane Photography was also a public accommodation, the fact that it was a for-profit business is irrelevant. Lastly, the *Dale* Court clearly explains that the distinguishing factor between applying *Roberts*, on one hand, and *Hurley* on the other, is whether the

---


227 *Id.* at 5.

228 *Dale*, 530 U.S. at 686 (Stevens, J., dissenting).

229 *Id.* at 688.


231 For an excellent argument of this point see Epstein, *supra* note 126, at 126-31.


233 *Roberts*, 468 U.S. at 626; *Boy Scouts of Am. v. Dale*, 530 U.S. 640, 657 (2000) (majority opinion). In *Hurley*, the Court specified that the parade’s expression was not a public accommodation, but assumed that the parade itself was a public accommodation. *Hurley*, 515 U.S. at 579.
First Amendment has been infringed. As demonstrated, applying the New Mexico public accommodations law did infringe the Huguenins’ First Amendment rights. Thus, the Hurley and Dale First Amendment analysis applies, under which the state’s interest in eliminating discrimination through enforcing public accommodations laws cannot justify the violation of the Huguenins’ rights.

III. PROPOSALS

_Elane Photography_ is a difficult case because it raises sensitive issues that strike at the heart of the moral and religious beliefs of a great many people. It is important to remember that the substance of the expression in cases like these is legally irrelevant. As the Court’s admonition at the end of Dale makes clear, “[t]he First Amendment protects expression, be it of the popular variety or not.” Indeed, a main purpose of the First Amendment is to protect the nation from government efforts to coerce unity of thought, because to do so is the first step toward destroying our democratic system of government. However, the significant purpose of the states in promulgating public accommodations laws—protecting the dignity and rights of access of all citizens—cannot be abandoned. The following suggestions are three possible ways to preserve public accommodations laws without significantly impairing their ability to curb discrimination while at the same time reducing the possibility of First Amendment violations on expressive businesses.

One obvious solution is to simply revert to a more narrow definition of “public accommodation” to exclude expressive businesses. A good definition that would help to limit the conflict between such laws and the Constitution comes from the “public character” rationale behind public accommodations laws. As Pamela Griffon explains,

The public character idea reflects the rationale for common law regulation of inns and common carriers. The owner of such facilities was

---

235 _Id._ at 659.
236 _Bernstein, supra_ note 138, at 125 n.222.
237 _Dale_, 530 U.S. at 660.
238 _W. Va. State Bd. of Educ. v. Barnette_, 319 U.S. 624, 641 (1943) (“[W]e set up government by consent of the governed, and the Bill of Rights denies those in power any legal opportunity to coerce that consent. Authority here is to be controlled by public opinion, not public opinion by authority.”).
239 _Griffin, supra_ note 4, at 1054.
deemed to be engaged in “quasi-public” service, because the property was put to a use in which the public had an interest. . . . Under this view, the purpose of restaurants, theatres and hotels is the public purpose of making a profit, which indicates that all paying customers will be accepted. In this context, racial or other discrimination among customers is unreasonable because such differences are irrelevant to the purpose for which the facilities operate.240

This rationale supports defining public accommodation as “an establishment in which minimal association exists between proprietor and customers, and in which the service relation is brief, casual and routine. In addition, the establishment provides a service necessary to the public, and a high degree of competition exists among establishments of the same kind.”241 Alaska’s public accommodation definition is not a bad example of one that strikes a nice balance between the competing interests at stake. That statute limits its scope to a place that caters or offers its services, goods, or facilities to the general public and includes a public inn, restaurant, eating house, hotel, motel, soda fountain, soft drink parlor, tavern, night club, roadhouse, place where food or spiritous or malt liquors are sold for consumption, trailer park, resort, campground, barber shop, beauty parlor, bathroom, resthouse, theater, swimming pool, skating rink, golf course, cafe, ice cream parlor, transportation company, and all other public amusement and business establishments, subject only to the conditions and limitations established by law and applicable alike to all persons.242

The list of enumerated businesses is inclusive enough to cover most situations and to promote inclusion of all people in most places. It is true that some of the places listed could be expressive in nature, especially because the list is left open by the clause on the end extending the definition to “all other amusement and business establishments.” However, the entire definition is conveniently qualified by any “conditions and limitations established by law,” which thereby automatically limits the statute to its constitutionally permitted scope, as interpreted by Hurley and Dale, which prohibits application of these statutes where First Amendment rights are infringed. Although the Constitution automatically limits all state statutes, adding an explicit qualifying clause allows an interpreting court to avoid having to strike down the entire statute on a facial challenge, thereby permitting a fluid

240 Id. at 1054-55.
241 Id. at 1055.
definition of expressive business to develop slowly and thoughtfully over time without undermining the important equal access goals the statutes further.

Constructions to be avoided are those like California’s, which evades the problem of definition altogether by prohibiting discrimination “in all business establishments of every kind whatsoever;”243 and those like New Jersey’s statute, which lists, but does not limit its scope to, more than sixty types of businesses, public places and institutions, including “any producer, manufacturer, wholesaler, distributor, retail shop, store, establishment, or concession dealing with goods or services of any kind.”244

This solution is neither perfect nor original. It is not perfect because it leaves the dirty job of line drawing to the courts on a case-by-case basis. Although, as mentioned earlier, this buys the courts time to develop a workable test for when a business is expressive and for achieving a better balance between the important goals of public accommodations law and the foundational right of free speech, it also leaves room for inconsistent application of the law.

The proposal is not original because a similar solution was recently suggested by James Gottry, as part of a more comprehensive two-pronged approach to alleviating the conflict between the First Amendment and public accommodations laws.245 The first prong of Gottry’s solution suggested that public accommodations laws should be legislatively narrowed in scope, both in term of the types of businesses they cover and the classes of people they protect.246 The second prong suggested that courts take an active role in avoiding constitutional conflicts by interpreting statutes narrowly, and by engaging in a more robust analysis by 1) inquiring whether there in fact is discrimination of a class denying access to the business or service, rather than a decision not to express an antithetical point of view; 2) investigating whether there in fact is expression involved; 3) investigating the quality of

---

243 CAL. CIV. CODE § 51(b) (West 2010).
244 N.J. STAT. ANN. § 10:5-5(l) (West 2010).
246 Id. at 997.
expression;247 and 4) applying Hurley to balance states’ interests against any First Amendment infringement found.248

This proposal raises several issues that require comment. First, as suggested above, narrowing the scope of public accommodations laws to more traditional common carriers and purveyors of necessary services is supported by the historical purpose of the law, and it helps avoid major First Amendment conflicts. However, limiting the suspect classes protected by public accommodations does not serve both these goals. The historical purpose of the law is not served because a blind adherence to the text of the original civil rights statutes, which protected only race, fails to do justice to the fundamental principle of equality upon which this nation is built and which has been recognized by many states as evidenced by the addition of gender, national origin, marital status and sexual orientation as protected classes in their public accommodations statutes.249 Second, restricting the scope of the protected classes does not help avoid First Amendment conflicts. National debates on race and gender are far from resolved. Even if public accommodations laws were restricted to protect the two least controversial “suspect classes,” the risk of First Amendment violations for expressive business would remain huge. Restricting the coverage of suspect classes would therefore both undermine the principles of equality on which historical public accommodations laws were based and fail to avoid the essential conflict between those laws and the First Amendment.

The second prong of the proposal is vulnerable to the same criticism as the first proposed solution above, which restricts only business types—it would require endless litigation and factual analysis of what types of businesses are expressive.250 This is a daunting proposition considering the vast array of artistic businesses that are potentially expressive—for example, custom baked goods,251 theme party

247 It is universally understood that certain types of speech are entitled to more First Amendment protection than others. See supra notes 107, 114 (explaining that political speech is the essence of speech afforded First Amendment protection).
248 See Gottry, supra note 245, at 1000-02.
249 See supra note 6.
251 Baked goods may already be a problem. For several years in a row, a ShopRite in New Jersey declined to provide a birthday cake inscribed with the name of a young boy called Adolf Hitler Campbell. See 3-Year-Old Hitler Can’t Get Name on Cake, MSNBC (Dec. 17, 2008, 6:40 PM), http://www.msnbc.msn.com/id/28269290/ns/us_news-weird_news/t/-year-old-hitler-cant-get-name-cake/.
planning services, or calligraphers for hire. If sorting through these, one business at a time, is not difficult enough, the issue becomes further complicated by the fact that it is very difficult to identify instances of expressive activity without a preexisting identifiable message.\textsuperscript{252} Elane Photography exemplifies this issue. After determining that Huguenin herself expressed no particular message of her own through her photographs, the district court found her First Amendment rights were not burdened by having to express someone else’s message.\textsuperscript{253} But the court has stated and reaffirmed that there need not be an identifiable message in order to constitute expression.\textsuperscript{254} All that is required is expressive activity,\textsuperscript{255} which could be anything.

The second proposal’s guidelines for determining whether expression exists do not help make this analysis any easier. Those guidelines suggest that courts follow the standard set out in \textit{Texas v. Johnson}\textsuperscript{256} for identifying expression. The \textit{Johnson} test asks whether there is an intent to convey a particularized message, and whether it is likely that viewers would understand the message being conveyed.\textsuperscript{257} That standard will not aid the courts because it was developed to identify expression conveyed via conduct—for example, the symbolic act of burning a flag, which is distinct from the more conventional modes of expression through utterance or modes akin to utterance.\textsuperscript{258} Indeed, the Court’s failure to apply that standard in \textit{Dale} or \textit{Hurley} emphasizes its inapplicability to cases involving utterance or analogous expression. Elane Photography demonstrates the difference—Huguenin does not intend to make a statement by the act of taking or refusing to take a photograph; rather, the photographs themselves are the expression. There is no magic formula for identifying an expressive business, and at the same time, a case-by-case development will necessarily entail a long and bumpy road.

\textsuperscript{252} See Epstein, \textit{supra} note 126, at 126-27.
\textsuperscript{255} \textit{Dale}, 530 U.S. at 655.
\textsuperscript{257} \textit{Id.} at 404.
\textsuperscript{258} \textit{Id.} at 402-03.
A third possible solution avoids the difficulties of distinguishing between expressive and nonexpressive businesses altogether. Professor Richard Epstein believes it is nearly impossible to draw the line between expressive and nonexpressive businesses, but after Dale, there is no need to.\textsuperscript{259} All that is required for constitutional protection is that a business “engages in expressive activity that could be impaired.”\textsuperscript{260} Epstein argues that the epitome of a nonexpressive business—“the profit-making corporation that ships goods, provides services, and cares only for its bottom line”—is a caricature that no real business fits.\textsuperscript{261} Corporations today have corporate cultures that help win the loyalty and approval of both employees and clients.\textsuperscript{262} These corporate identities are built through expressive activity, whether by donating to causes, participating in community service projects, taking voluntary environmental protection measures, or conducting employee health initiatives.\textsuperscript{263} This means that the Dale standard for expressive activity is always met, because all organizations engage in expressive activity by building their corporate identity.\textsuperscript{264}

In addition to his argument that it is meaningless to draw lines between expressive and nonexpressive businesses, Epstein contends that no legal basis exists for government control over decisions of private individuals to discriminate in providing goods or services.\textsuperscript{265} Further, he makes the case that in developed private markets there is no need for it, since “voluntary segmentation of population groups” may be beneficial for their members, and people may prefer it.\textsuperscript{266}

This does not, however, mean that discrimination by anyone, anywhere is okay; there are still compelling justifications for upholding public accommodations laws in at least one class of business.\textsuperscript{267} Where there is no possibility for voluntary segmentation antidiscrimination laws may be justified.\textsuperscript{268} That is, wherever a business holds a monopoly position in the market, and people cannot benefit from the social and economic efficiencies that result from voluntary

\textsuperscript{259} Epstein, supra note 126, at 139-40.
\textsuperscript{260} Dale, 530 U.S. at 655.
\textsuperscript{261} Epstein, supra note 126, at 139.
\textsuperscript{262} Id.
\textsuperscript{263} Id. at 140.
\textsuperscript{264} Id.
\textsuperscript{265} Id. at 133.
\textsuperscript{266} Id. at 134.
\textsuperscript{267} Id. at 121.
\textsuperscript{268} Id. at 136.
organizing into internally coherent groups, the government should apply public accommodations laws to protect against the economic and social harms that monopolies have the disproportionate power to cause.\footnote{Id.} Thus, public accommodations laws should only apply to government institutions, which have a monopoly as a matter of law,\footnote{Id. at 121.} and other essential businesses that “typically supply standard commodities—electrical power, telephone service, railroad transportation—and only work because the firm is largely indifferent to the identity and personal characteristics of its customers.” \footnote{Id. at 137.} Recognizing the major weakness in this theory, Epstein points out that the main problem with his suggestion will be defining what exactly constitutes a monopoly.\footnote{Id. at 121.} This is complicated not just by the haziness of characteristics that imply monopoly status—stable, long term, and having minimal competition—but also the difficulty of identifying the geographic scope of the business, and the relevant market that should underlie the analysis.

None of the theories discussed in this section provide a magic antidote to the conflicts raised by *Elane Photography*. Nevertheless, identifying the common ground among them may help illuminate potential areas of future reform. All of the theories discussed agree with the basic proposition that the application of public accommodations laws to businesses that provide standard or essential services and have no reason to distinguish between their customers on the basis of any personal characteristics would not violate the First Amendment.\footnote{Griffin, supra note 4, at 1055; Gottry, supra note 245, at 997, 965-68 (recommending a return to the historical contours of public accommodations laws, which limited their scope to “essential goods and services”); Epstein, supra note 126, at 137.} That, at least, is a start. All three theories would probably also agree that the application of public accommodations laws to businesses where the provider-consumer interaction is “brief, casual and routine” are similarly unlikely to be sites of frequent First Amendment infractions. In all other types of businesses, the presence of expression will serve as a threshold issue that will likely do most of the “work” in the First Amendment analysis. This is because once expression is found, it is comparatively simple to determine whether the expression has been coerced or would be altered by the application of a public accommodations
statute; and if there is no expression, there is no problem. Perhaps the best (but likely not ideal) solution then is to focus on when a business is engaging in expression in a particular circumstance, rather than whether it is an expressive business as a matter of law. The precise standard will have to be developed over time, likely through some sort of imprecise totality of the circumstances analysis. This instance-focused approach (in contrast to the approaches that focus on definitions of expressive businesses above) gives more consideration to the high values of equality and access that public accommodations laws serve, while protecting First Amendment values as well. It also takes into account the fact that similar behavior may be found to be expressive in some circumstances and not in others, even if done by the same person. Although this method will also result in case by case doctrinal development, perhaps a factual approach will eventually lead to more clarity in the law. In the meantime, statutory definitions of public accommodations need not be completely rewritten. A simple qualification, like Alaska’ s, that the definition is subject “to the conditions and limitations established by law” will allow the doctrine to unfold in the courts.

Elane Photography may be one of the first cases to highlight the tension between public accommodations laws and the First Amendment right to be free from compelled speech, but it will not be the last. One can easily imagine, for example, that the development of new technology and the rise of social media as a commercialize-able form of expression will lead to an infinite number of possible First Amendment violations. But these new issues reflect old problems, and it is appropriate to look to old wisdom for guidance. Justice Brandeis once opined that the Framers of the U.S. Constitution “believed that freedom to think as you will and to speak as you think are means indispensable to the discovery and spread of political truth; that without free speech and assembly discussion would be futile; that with them, discussion affords ordinarily adequate protection against the dissemination of noxious doctrine.” The best way to fight “invidious discrimination” in expressive

---

274 For example, photography may not always be automatically determined to be expressive. A photographer who takes private jobs, plays an active part in staging her photographs and exercises judgment in choosing lighting, perspective, and composition is not the same as an employee of a Sears photography studio, which is a large national chain, characterized by general openness to the public whose employees likely have little discretion in discharging their responsibilities.

275 See supra note 242.

businesses is, therefore, through means that encourage more speech and discussion and not through speech restrictions.

Susan Nabet†

† J.D., Brooklyn Law School, 2012; A.B., Princeton University, 2005. Professor William Araiza deserves special thanks for suggesting this topic and for being so generous with his time and assistance throughout the writing process. Thanks, too, to the staff of the Brooklyn Law Review for their extraordinary efforts on my behalf and to Robert Sobelman, whose excellent counsel I will continue to seek for many years to come. Finally, I am deeply grateful for the love and support of my husband Benjamin Nabet, and my parents Ann and Herb Berezin.
Protecting the Border, One Passenger Interrogation at a Time

INTRODUCTION

The terrorist attacks that occurred on September 11, 2001, represented the ultimate intersection between criminal and immigration law. Because many of the terrorists had entered the United States legally with visas issued by the Immigration and Naturalization Service (INS), the tragedy revealed the deficiencies in the administration of laws that provided for entry into the United States. Thus, in the years following September 11, immigration policy has been transformed to ensure that persons who have not been properly screened and verified are not allowed to remain in the country. That transformation has included a greater criminalization of immigration violations as “illegal immigrants have come to be seen as synonymous with terrorists.” The new priorities of immigration agencies and authorities have become to restrict admission and increase deportations with the purported goal of rooting out terrorists and increasing the security of the nation.

2 See 9/11 Commission Report, supra note 1, at 80-82.
3 See Staff Statement No. 1, supra note 1, at 3 (“Our immigration system before 9/11 focused primarily on keeping individuals intending to immigrate from improperly entering the United States.”); see also Nora V. Demleitner, Misguided Prevention: The War on Terror as a War on Immigrant Offenders and Immigration Violators, 40 CRIM. L. BULL., no. 6, at 5 (2004) (noting that in mid-2001, fighting terrorism was not INS’s main concern).
4 See generally Demleitner, supra note 3.
5 Id. at 2.
7 Demleitner, supra note 3, at 1. “Because of the focus on foreign terrorism, immigration law has become a major investigatory and enforcement tool on the frontline in the fight against terrorism.” Id.
Charged with fighting terrorism and increasing homeland security, immigration agencies and officials have seen their power strengthened and expanded. For years, immigration agents have been routinely boarding domestic trains and buses traveling within one hundred miles of the border and interrogating passengers about their citizenship status and requesting their immigration paperwork. Agents have arrested and detained thousands of passengers that did not have their immigration paperwork. Individuals who were detained were sent either to detention facilities or local prisons and county jails, and most were held there until they were able to post a bond.

This purportedly legal practice—the Immigration and Nationality Act (INA) allows immigration officials to interrogate anyone they suspect of being an illegal alien—actually goes far beyond what Congress intended. Congress gave officials the authority to interrogate individuals in this way at the border, or close to the border, because of the greater need to investigate entrants into the United States. Instead, immigration officials have been using the authority of the INA to question and arrest passengers on domestic vessels that have not crossed a border and will not be crossing a border. Further, the explanation of fighting terrorism and arresting recently entered illegal immigrants cannot be justified by this practice, as “the vast majority of those arrested . . . had been in


9 Bernstein, supra note 8; N.Y. CIVIL LIBERTIES UNION, FAMILIES FOR FREEDOM & NYU SCH. OF LAW IMMIGRANT RTS. CLINIC, JUSTICE DERAILED: WHAT RAIDS ON NEW YORK'S TRAINS AND BUSES REVEAL ABOUT BORDER PATROL'S INTERIOR ENFORCEMENT PRACTICES 1, 4 (2011) [hereinafter NYCLU REPORT], available at http://www.nyclu.org/publications/report-justice-derailed-what-raids-trains-and-buses-reveal/about-border-patrols-interi. “[N]early all individuals arrested during transportation raids are detained by CBP without being screened for risk of flight, threat to the community, or other considerations . . . regardless of whether they are recent entrants apprehended at the border or have resided in the United States for years.” Id. at 14.

10 NYCLU REPORT, supra note 9, at 14-15.


12 See infra Part II.A (discussing the “border-search exception,” which allows officials greater leeway in investigating individuals when they arrive from outside the U.S. specifically because they are coming from outside the country).

13 Bernstein, supra note 8.
the country for more than one year,” and many had been in the country for more than three years.14

Immigration officials violate the Fourth Amendment when they interrogate random passengers in this way. In the seminal Fourth Amendment case Terry v. Ohio, the Supreme Court laid out the parameters of police officers’ authority.15 The Court specifically stated that a seizure occurs when a police officer “restraints” an individual’s ability to “walk away.”16 An officer who has seized an individual must have reasonable suspicion to justify the intrusion.17 The practice at issue here is a coercive display of police authority that constitutes a seizure, because the passengers do not feel free to refuse to respond.18 Under Terry and the Court’s subsequent cases further defining a “seizure,”19 immigration officers must have reasonable suspicion or probable cause to carry out their investigations.

This practice directly results from the post-September 11 expansion in immigration officials’ powers, which gives agents unprecedented authority under the official purpose of fighting terrorism.20 However, the post-September 11 policies have resulted in almost no arrests for the actual crime of terrorism.21 On the other hand, the lives of individuals who have been in the United States for years have been unjustifiably disturbed; in many instances, the consequences for individuals questioned during these transportation checks are dire, since the INA permits mandatory detention for

---

14 NYCLU REPORT, supra note 9, at 8-9; see also Bernstein, supra note 8.
16 See id. at 16.
17 See id. at 20-21.
18 See infra Part III.A; see also United States v. Mendenhall, 446 U.S. 544, 552 (1980) (“Only when the officer, by means of physical force or show of authority, has in some way restrained the liberty of a citizen may we conclude that a ‘seizure’ has occurred.”).
19 See infra Part III.A.
20 See infra Part I.
21 DAVID COLE & JAMES X. DEMPSEY, TERRORISM AND THE CONSTITUTION 232-34 (3d ed. 2006) (The authors note that “the war on terrorism, at least at home, has netted almost no actual terrorists.” Specifically, the authors write that although “[t]he Justice Department boasts that its terrorism investigations have led to more than 300 criminal indictments [and] over 100 convictions,” the vast majority of those convictions have been “for minor charges, not terrorism.” Further, “few of the government’s indictments charge actual terrorism.”); see also Demleitner, supra note 3, at 1 (“[S]o far these special measures have yielded few tangible results.”); CONSTITUTION PROJECT, THE USE AND ABUSE OF IMMIGRATION AUTHORITY AS A COUNTERTERRORISM TOOL: CONSTITUTIONAL AND POLICY CONSIDERATIONS 1 (2008), available at http://www.constitutionproject.org/pdf/Immigration_Authority_As_A_Counterterrorism_Tool.pdf (“As the bipartisan 9/11 Commission’s staff found, there is no evidence that the post-September 11 immigration initiatives targeted at Arabs and Muslims succeeded in identifying any actual terrorists.”).
individuals with questionable status and possibly immediate deportation without judicial review. These actions are troubling because the courts, the INA, and the current administration recognize that some immigrants, even those in the country illegally, deserve certain protections. Recently, the Obama administration has stated that it will focus on deporting convicted criminals and individuals who pose national security risks rather than illegal immigrants with no criminal records. Additionally, Senator Richard J. Durbin has sponsored legislation called the Development, Relief, and Education for Alien Minors Act of 2010 (the DREAM Act) that would provide a path to citizenship for certain young illegal immigrants who came to the United States as children. Thus, the Obama administration and even members of Congress have recognized that although individuals may be in the country illegally, their ties to the United States may afford them greater protection from intrusion and seizure than first-time entrants or suspected terrorists.

24 Historically, the Supreme Court has recognized that individuals who have resided in the United States for some time are entitled to more procedural protections than first-time entrants. Landon v. Plasencia, 459 U.S. 21, 32 (1982) (“[O]nce an alien gains admission to our country and begins to develop the ties that go with permanent residence his constitutional status changes accordingly.”); Yamataya v. Fisher, 189 U.S. 86, 100-01 (1903) (holding that an illegal alien, who “has become subject in all respects to” the jurisdiction of the United States and has become “a part of its population,” is entitled to some due process protections).
25 The INA’s cancellation of removal procedures are arguably a recognition by Congress that individuals who have lived in the U.S. for many years and have established ties to the country may, in certain instances, gain permanent resident status and repose. Cf. 8 U.S.C. § 1229b.
27 David M. Herszenhorn, Senate Blocks Bill for Young Illegal Immigrants, N.Y. TIMES (Dec. 18, 2010), http://www.nytimes.com/2010/12/19/us/politics/19immig.html?sq=DREAM%20act&st=cse&adxnnl=1&scp=1&adxnnlx=1293375627-zGDKC2XQgm3tzHzoOO5g. The Senate voted down the bill, in a vote by 55-41 in favor of the bill. Id.
In November 2011, three immigrant rights groups—the New York Civil Liberties Union, Families for Freedom, and the Immigrant Rights Clinic at New York University School of Law—released a report examining data of arrests that occurred during these “transportation raids” in upstate New York. The report found that these document checks “do little to protect the border, but they threaten constitutional protections that apply to citizens and non-citizens alike.” In particular, the authors concluded that the majority of those arrested and detained were not recent border-crossers, agents violated “established arrest procedures,” and anecdotal reports indicated that officers used racial profiling to pick out the individuals stopped and questioned. The report advocated ending this practice and putting in place more constitutional and procedural protections.

In fact, the Border Patrol has taken some action to scale back the amount of random transportation checks that occur. In October 2011, the agency ordered field offices that were not near the southwest border to conduct checks in train and bus stations and airports only when “they have specific ‘actionable intelligence’ that there is an illegal immigrant there who recently entered the country.” However, this order “has not been made public,” and, as stated by a spokesman for the U.S. Customs and Border Protection, “does not amount to a change in policy.” Additionally, Border Patrol Agents and their union, the National Border Patrol Council, have criticized the move. The union stated that the reduction in the number of transportation checks has “handcuff[ed] the effectiveness of Border Patrol agents.” The union also alluded ominously to the September 11 attacks and stated, “A decade ago nineteen illegal aliens overstayed visas . . . which resulted in nearly 3,000 Americans losing their lives. This lesson must be lost on

---

28 NYCLU REPORT, supra note 9.
29 Id. at 1.
30 Id. at 2.
31 Id. at 25.
33 Id.
35 NBPC Press Release, supra note 34.
those running the Border Patrol in Washington.” Thus, although there has been a shift away from using transportation checks on domestic bus and train routes, no formal or official order has ended the practice, and there has been backlash.

As evident from the National Border Patrol Council’s statement, illegal immigrants are still closely associated with terrorists, and until there is permanent action to curb the practice of transportation checks, they may continue to be used at any time. Therefore, Congress and the Department of Homeland Security (DHS), the agency responsible for border security and the training of Border Patrol agents, must take official action to curtail this practice. Congress should amend the INA to permit Border Patrol agents to question individuals only when they have reasonable, particularized suspicion that the passenger is either in the country illegally or may be a terrorist, and the DHS should provide ongoing training to its agents on administering this standard.

This note will analyze the Border Patrol’s interrogation of passengers on domestic vessels under the Fourth Amendment and border search jurisprudence, and it will argue that the practice is unconstitutional under the Fourth Amendment. Therefore, this practice must be curtailed. Part I will discuss the structure of immigration authority and immigration law in the United States, including the changes in that structure after the September 11 attacks. Part II will discuss the law that defines the parameters of searches at the border and analyze whether this practice is illegal under that framework. Part III will discuss the general search and seizure law after Terry and argue that random requests for immigration documents of passengers on domestic vessels is both unconstitutional and based on unsound policy. Part IV will propose an amendment to the INA to curb this practice and a refinement of DHS training policy that would permit officers to question passengers only when they have individualized, reasonable suspicion to believe that the passenger is either an illegal alien or a terrorist.

Although protecting the nation from terrorism is an imperative objective, the Border Patrol’s interrogation of domestic passengers does nothing to further that goal. On the other hand, this practice violates passengers’ personal liberties in contravention of the Fourth Amendment. Because it is

36 Id.
unconstitutional and ineffective, Congress and the DHS should limit Border Patrol agents’ authority.

I. CHANGES IN IMMIGRATION LAW AND POLICY AS A RESPONSE TO TERRORISM

Even before the attacks on September 11, Congress made changes to immigration law to deter and penalize illegal immigration. After September 11, however, the connection between immigration and terrorism became explicit and Congress acted to strengthen the police powers of immigration officials by increasing funds, personnel, and the jurisdiction in which they could interrogate and detain individuals suspected of being illegal immigrants. The changes in immigration law and policy after September 11 have led to the current policy of randomly stopping, interrogating, and detaining passengers on common carriers travelling on routes exclusively within the United States.

A. Immigration Law Prior to September 11, 2001

Before the massive changes wrought by the terrorist attacks on September 11, immigration officials were responsible for regulating “travel, entry, and immigration” into the United States. Congress charged the INS, the primary agency overseeing immigration, principally with preventing individuals from entering the country illegally and working in the United States without authorization. Although the INS had a staff of about “9,000 Border patrol agents, 4,500 inspectors, and 2,000 immigration special agents,” the job function of these individuals was not framed in the context of national security. Instead, the INS had responsibility for the “controlled entry” of temporary visitors and the administration of programs that allowed non-citizens to become naturalized or to gain

37 See infra Part I.A.
38 See infra Part I.B.
39 See infra Part I.C.
40 9/11 COMMISSION REPORT, supra note 1, at 383-84.
42 Id. (The INS was charged with investigating and sanctioning corporate employers who hired illegal immigrants, as well as the deportation of those illegal immigrants); Donald Kerwin & Margaret D. Stock, The Role of Immigration in a Coordinated National Security Policy, 21 GEO. IMMIGR. L.J. 383, 386 (2007).
43 9/11 COMMISSION REPORT, supra note 1, at 80.
44 Id. at 383-84.
permanent resident status. The INS’s main enforcement function was to detect and remove aliens who had entered the country illegally or stayed past the expiration of their legal documents. Therefore, searching for terrorists was not the priority of the INS or its employees.

The powers of the INS were defined by the INA. Since its enactment in 1952, Congress has amended the INA numerous times. The history of the INA—specifically the amendments enacted in the last twenty years—demonstrates the role that the ideals of fighting terrorism and national security have played in transforming immigration law to expand the power of immigration officials to find and remove certain individuals.

For instance, in 1996, following the 1993 bombing of the World Trade Center and the 1995 Oklahoma City bombing, Congress passed the Anti-Terrorism and Effective Death Penalty Act (AEDPA). A reaction to the threat of terrorism, the AEDPA broadened the law under which individuals could be denied entry based on their suspected connections to terrorism. Previously, the government had the burden of proving that the individual denied entry or facing deportation had “personally engaged” in terrorist activity. Under the new Act, admissibility would be denied to any individual who was

---

45 Smith, supra note 41, at 308.
47 9/11 Commission Report, supra note 1, at 81.
48 See Regina Germain, AILA’s Asylum Primer 24 (6th ed. 2010).
49 See Smith, supra note 41, at 305-08 (summarizing enactment of immigration legislation from the late 1800s to the present); Public Laws Amending the INA, U.S. Citizenship & Immigr. Servs., http://www.uscis.gov/portal/site/uscis/menuitem.f6da51a2342135be7e9d7a10e0dc91a0/?vgnextoid=fa7e539dc4bed010VgnVCMI000000eced190aRCRD&vgnextchannel=fa7e539dc4bed010VgnVCM1000000eced190aRCRD&CH=publaw (last visited Jan. 24, 2012).
50 Farnam, supra note 6, at 22-23.
51 Perhaps foreshadowing the legislative response to the September 11 attacks, AEDPA was enacted primarily in the shadow of the Oklahoma City Bombing. Cole & Dempsey, supra note 21, at 132. Immediately following the attack, “Members of Congress . . . felt tremendous pressure to pass antiterrorism legislation. It did not matter that the proposals in the president’s initial bill were directed largely against international terrorism, while the Oklahoma bombing was the work of homegrown criminals.” Id.
52 Id. at 143.
found to be “a representative” or a “member” of a terrorist organization.\textsuperscript{53} Thus, the new law took away the requirement of individual responsibility and instead denied entry to individuals based merely on their suspected associations with terrorist groups.\textsuperscript{54}

Following the passage of the AEDPA, Congress passed the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (IIRIRA)\textsuperscript{55} to more forcefully address the problem of illegal immigration into the United States,\textsuperscript{56} which was linked to terrorism. “IIRIRA was a major overhaul of the entire INA”\textsuperscript{57} that “drastically changed the landscape of immigration law.”\textsuperscript{58} The IIRIRA made significant changes to immigration enforcement both at the borders and in the country’s interior. In particular, the Act authorized the Attorney General to


\textsuperscript{54} COLE & DEMPSEY, supra note 21, at 143 (arguing that AEDPA substituted the requirement of a “personal connection to terrorist activity” for “guilt by association,” which is otherwise prohibited by the First Amendment).


\textsuperscript{56} See H.R. REP. No. 104-879, at 95 (1997). In its report on the history of the enactment of the IIRIRA, the Judiciary Committee characterized the state of illegal immigration, at the time, as such:

[M]ore than 4 million illegal aliens resided in the United States at the start of the 104th Congress, with an average net increase each year of 300,000; approximately half of these illegal residents had arrived with legal temporary visas and had overstayed; each year, tens of thousands of illegal aliens were ordered deported but were not removed from the United States due to lack of resources and legal loopholes; and the legal immigration system failed to unite nuclear families promptly, encouraged the “chain migration” of extended families, and admitted the vast majority of immigrants without regard to their level of education, job skills, or language preparedness.

\textsuperscript{57} Scott Aldworth, Note, Terror Firma: The Unyielding Terrorism Bar in the Immigration and Nationality Act, 14 LEWIS & CLARK L. REV. 1159, 1167 (2010).

\textsuperscript{58} FARNAM, supra note 6, at 30.
substantially increase the number of border patrol agents, install additional physical barriers and roads in the vicinity of the U.S. border, and buy any additional equipment necessary to stop illegal immigration. Further, the IIRIRA included changes to the procedures for inspecting, detaining, and removing aliens. For example, the Act established “expedited removal,” under which an individual in certain circumstances could be deported without judicial review, and mandated the detention of individuals facing expedited removal.

Thus, even before September 11, Congress responded to the fear of terrorism by targeting illegal immigration—it expanded the money and resources given to the INS, broadened the definition of which individuals could be deported and denied entry, and removed judicial review for deportation proceedings in certain cases. After September 11, the perceived tie between immigration and terrorism became even stronger, and Congress and the President responded accordingly.

B. Immigration Law after September 11, 2001

In 2005, “The 9/11 Commission Report” was released to the public. The National Commission on Terrorist Attacks

59 IIRIRA § 101(a). Specifically, IIRIRA states that “[t]he Attorney General in each of fiscal years 1997 . . . [–] 2001 shall increase by not less than 1,000 the number of positions for full-time, active-duty border patrol agents within the Immigration and Naturalization Service.” Id.

60 Id. § 102(a).

61 Id. tit. III; see also FARNAM, supra note 6, at 30-31 (“IIRIRA . . . implement[ed] changes in border control, document fraud dealings, admissibility procedures, removal processes, asylum and refugee law, with implications for international students, and visas and consular procedures in general. This law is perhaps the biggest overhaul of the U.S.’s immigration system since the INA . . . .”).


63 8 U.S.C. § 1225(b)(1)(B)(iii)(IV). The mandatory detention provision of the IIRIRA has proven to be one of its most controversial aspects. See generally Nancy Morawetz, Understanding the Impact of the 1996 Deportation Laws and the Limited Scope of Proposed Reforms, 113 HARV. L. REV. 1936 (2000); see also Lenni B. Benson, As Old as the Hills: Detention and Immigration, 5 INTERCULTURAL HUM. RTS. L. REV. 11, 12 (2010) (“We use immigration law and detention as a weapon in the law’s enforcement because we seek to control our border. . . . [O]ur ‘border control’ is person control and containment.”).

upon the United States (the Commission), which had been created by Congress and the President to investigate the attacks, was highly critical of the INS. The Commission found that in the 1990s, the INS was “seriously hampered by outdated technology and insufficient human resources.” As a result, the Commission concluded, the INS could barely handle its actual primary functions, much less be prepared to fight terrorism.

In fact, before September 11, there was no U.S. government agency whose primary responsibility was analyzing the travel of foreign nationals to determine any potential terrorist threat. The Commission found this particularly troubling and noted, “For terrorists, travel documents are as important as weapons.” In his testimony before the House Committee on the Judiciary shortly after the attacks, Attorney General John Ashcroft emphasized this point and stated, “The ability of alien terrorists to move freely across our borders and operate within the United States is critical to their capacity to inflict damage on our citizens and facilities.” His statement made clear that immigration law would soon become a tool for detecting criminal terrorist activity.

The Commission proposed the establishment of a new agency—the Department of Homeland Security (DHS)—to address the deficiencies in the U.S. immigration system that September 11 made painfully clear. The new department would be a consolidation of the multitude of government agencies that had previously been tasked, separately, with homeland security and immigration oversight. The hope was

---

65 9/11 COMMISSION REPORT, supra note 1, at xv.
66 See id. at 383-84. The Commission pointedly noted that, “[T]he routine operations of our immigration laws—that is, aspects of those laws not specifically aimed at protecting against terrorism—inevitably shaped al Qaeda’s planning and opportunities.” Id. at 384.
67 Id. at 80.
68 Id. at 384.
69 Id.
70 Id.
72 See id. at 4. Ashcroft stated that a goal of the DOJ would be to increase the authority of the INS, as part of broader changes in “law enforcement.” He thus placed the changes to immigration policy within the larger context of improving criminal law enforcement. See id. at 4-7.
that one department “that has no question about either its mission or its authority,” along with one “comprehensive national strategy” for fighting terrorism, would be better able to secure the nation.

Thus, in November of 2002, President George W. Bush signed into law the Homeland Security Act of 2002, which established the DHS and directed that the Secretary of Homeland Security would run the department. Unlike the INS, the DHS had the primary purpose of fighting terrorism. The U.S. Government manual states that the DHS “leads the unified national effort to secure America. It will prevent and deter terrorist attacks and protect against and respond to threats and hazards to the Nation. The Department will ensure safe and secure borders, welcome lawful immigrants and visitors, and promote the free-flow of commerce.” By its own mandate, the goal of the country’s main immigration authority is to fight terrorism first and to welcome immigrants second.

The DHS was created to unify the various immigration and homeland security functions that had previously been spread across different agencies. Thus, the immigration departments of the DHS each have a specific function that combines immigration and law enforcement. These departments include the United States Citizenship and Immigration Services (USCIS), which establishes and administers immigration and

(“Before the establishment of the Department of Homeland Security, homeland security activities were spread across more than 40 federal agencies and an estimated 2,000 separate Congressional appropriations accounts.”).

75 DHS PROPOSAL, supra note 73, at 5.
76 DHS BRIEF HISTORY, supra note 74, at 4.
77 Id. at 7.
79 DEPT OF HOMELAND SEC., U.S. GOVERNMENT MANUAL 228 (2009-10), available at http://www.gpoaccess.gov/gmanual/browse-gm-09.html; FARNAM, supra note 6, at 53 (“Since 9/11 . . . the priority of the United States government and the agencies that control immigration in this country . . . is protection.”).
80 U.S. GOVERNMENT MANUAL, supra note 79, at 226.
81 In the section defining the mission of DHS, the Homeland Security Act of 2002 lists first and foremost that “[t]he primary mission of the Department is to . . . prevent terrorist attacks within the United States.” Pub. L. No. 107-296, § 101(b)(1); see also ABA REPORT, supra note 46, at 1-5 (“DHS serves both an enforcement function . . . and a service function . . . .”).
naturalization policy;\textsuperscript{83} the United States Customs and Border Protection (CBP), which is responsible for securing the borders;\textsuperscript{84} and the United States Immigration and Customs Enforcement (ICE), which enforces immigration laws in the country’s interior.\textsuperscript{85} Specifically, the CBP conducts inspections of arriving people and goods at ports of entry and is charged with “deterrence or apprehension of illegal immigrations between ports of entry.”\textsuperscript{86} Further, the Border Patrol is responsible for securing both the country’s international land borders with Canada and Mexico, as well as the United States’s coastal borders.\textsuperscript{87} The ICE, on the other hand, is charged with conducting investigations in the country’s interior, as well as with the detention and removal of noncitizens.\textsuperscript{88} The three agencies work together to conduct the removal proceedings for noncitizens: the CBP and the ICE initially determine which individuals should be subject to removal proceedings, while the USCIS conducts the proceedings to determine whether these individuals should be granted legal status to remain in the United States.\textsuperscript{89}

Additionally, after the September 11 attacks, Congress significantly increased the number of personnel and funds available to immigration authorities in every state along the northern border.\textsuperscript{90} In the Uniting and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (PATRIOT) Act of 2001, Congress appropriated “such sums as may be necessary” to triple the number of Border Patrol personnel, Customs Service personnel, and INS inspectors.\textsuperscript{91} Further, the PATRIOT Act appropriates an additional $50,000,000 to both the INS and the U.S.
Customs Service to improve and acquire any technology and equipment necessary to monitor the northern border.92 Since its creation in 2003,93 the CBP has grown into a major government operation. It currently employs more than 20,000 Border Patrol agents and has officers at more than 330 ports of entry.94 Further, the agency received $10.1 billion in funding for fiscal year 2010.95 More than $2 billion has been appropriated to border security, including personnel, infrastructure, and technology.96 These figures indicate that as the perceived link between immigration and terrorism has grown, Congress and the President have increased funds, personnel, and support to the DHS. That support has directly led to the expanded use of random document checks of passengers traveling domestically on public transportation.

C. Internal Document Checks

Policing immigration has become one of the government’s primary methods for fighting terrorism and providing national security.97 Border Patrol agents have broad statutory authority to conduct investigations of individuals they suspect to be in violation of immigration laws. Currently, under the INA, immigration officials do not need a warrant to “interrogate any alien or person believed to be an alien as to his right to be or to remain in the United States,”98 arrest any alien they see illegally entering the United States or who they believe is in the country illegally and “is likely to escape before a warrant

92 Id. § 402(4).
96 Id.
97 See Stephen H. Legomsky, The New Path of Immigration Law: Asymmetric Incorporation of Criminal Justice Norms, 64 WASH. & LEE L. REV. 469, 509 (2007) (“Perhaps no single development better exemplifies the public association of immigration and terrorism than the transfer of immigration functions to a Department whose defining mission is counter-terrorism.”); Demleitner, supra note 3, at 9-12 (“[I]mmigration-related activities have been prosecuted as criminal violations but justified as anti-terrorism measures.... It was 9/11 that made immigration.... a national security issue.”).
can be obtained for his arrest,” \footnote{Id. § 1357(a)(2).} and “to board and search for aliens any vessel . . . and any railway car, aircraft, conveyance, or vehicle” within a reasonable distance of a U.S. border.\footnote{Id. § 1357(a)(3). See infra Part II.A (discussing how the INS has defined “reasonable distance”).} Beyond the border or its “functional equivalent,” however, immigration officials must have at least reasonable suspicion to warrant an inquiry or a search.\footnote{See infra Part II.A (discussing the Supreme Court’s interpretation of this statute as it applies to roving patrols at the border and beyond).} Nonetheless, following the general expansion of authority after the September 11 attacks, CBP agents have utilized this statutory grant of power to conduct random interrogations of passengers traveling domestically within the United States—that is, not at the border or its functional equivalent.\footnote{See infra Part II.B (arguing that the area where transportation checks occur does not fall under the definition of the “functional equivalent” of the border).}

CBP agents conduct inspections by boarding domestic trains, buses, and ferries traveling along domestic routes (meaning the vessels never cross the border) and inquiring about passengers’ immigration status.\footnote{Bernstein, supra note 8; Emily Bazar, Some Travelers Criticize Border Patrol Inspection Methods, USA TODAY (Oct. 2, 2008, 1:08 AM), http://www.usatoday.com/news/nation/2008-09-30-border-patrol-inside_N.htm; Branch-Briso, supra note 8.} At any given time, at least six uniformed officers, all armed, may board the vessel and—with no preface and no outright indication that the passenger may refuse to give their consent—ask passengers whether they are U.S. citizens.\footnote{Bernstein, supra note 8.} As one journalist observed, passengers “startled from sleep, simply stared, and the agents prompted them: ‘State your citizenship for me, please, sir. What country were you born in?’”\footnote{Id.; see also NYCLU REPORT, supra note 9, at 7.} Border Patrol agents request the immigration documents of any passengers who are not U.S. citizens.\footnote{Bernstein, supra note 8.} Passengers without their documents are removed from the train and detained for further investigation and questioning, in full view of the other passengers.\footnote{Bernstein II, supra note 104.}

Though this practice is not widely publicized,\footnote{NYCLU REPORT, supra note 9, at 7; see also Bernstein, supra note 8 (noting that “[d]omestic transportation checks are not mentioned in a report on” CBP’s “northern border strategy”).} it is actually quite large in scope. In August 2010, Nina Bernstein of
the New York Times reported that each year, as a result of these transportation checks, hundreds of passengers on trains and buses traveling along the northern border are taken to detention—and placed in removal proceedings—because they do not have “satisfactory immigration papers” with them. Bernstein further suggested that such stops account for about 3,000 arrests based on immigration violations each year. In 2008, in a similar report regarding transportation checks on domestic trains, buses, and ferries, USA Today reported that transportation checks in CBP’s Buffalo sector resulted in the arrest of 1,786 illegal immigrants. In the agency’s New Orleans sector, 1,754 illegal aliens were arrested on bus checks.

These internal transportation checks have been “fueled by . . . an expanding definition of border jurisdiction” and justified primarily as a security measure to prevent terrorism. Further, immigration officials argue such checks are necessary on transportation near the border, because illegal immigrants would flee deeper into the interior after entering. The problem with this argument, however, is that CBP agents are boarding domestic carriers that have not crossed the border. The vast majority of those arrested and detained had resided in the United States for years—some had overstayed their visa status, were in the process of changing their status, or had actually been granted legal status already. Nancy Morawetz, a leading immigration scholar who

---

109 Bernstein, supra note 8.
110 Id.
112 Id. The New Orleans sector covers seven states. Id.
113 Bernstein, supra note 8; see also Shachar, supra note 62, at 811-19 (arguing that expedited removal is an example of how the definition of the “border” is changing—becoming “malleable and movable”—and “blurring the line between the perimeter and the interior”).
114 “Our mission is to defend the homeland, primarily against terrorists and terrorist weapons,” said the immigration official in charge of the Border Patrol station in Rochester, NY, where 1040 people were arrested in 2008, “95% of them from buses and trains.” Bernstein, supra note 8.
115 A supervisory agent in Washington, D.C. explained CBP’s policy by stating that, “If you have someone attempting to illegally enter the United States, it’s very unlikely that they’re going to stay 15 yards from the international border . . . . We want to take a layered approach.” Bazar, supra note 103 (internal quotation marks omitted). Similarly, an immigration official in Washington State argued that, “The first line of defense is on the immediate border . . . . We have to have a second line of defense.” Bazar, supra note 111 (internal quotation marks omitted).
116 NYCLU REPORT, supra note 9, at 6-11. The report further states that “Less than one percent of those arrested had entered the United States within the last 72 hours.” Id. at 10; see also Nadja Drost, Heighted Security at U.S.-Canada Border Catching Few Terror
supervises the NYU Immigrant Rights Clinic and is one of the authors of the NYCLU report, argues that while the expansion of the CBP’s authority was meant to deal with “border security,” it has actually become “interior enforcement to sweep up farmworkers and students.”

Such document checks have questionable utility for fighting terrorism as well. As a CBP public affairs officer stated, “If you look at our apprehensions, a small percentage have anything to do with terrorism . . . .” Ninety percent of the CBP’s prosecutions in 2008 were for immigration—not criminal—violations. Thus, the government cannot justify transportation checks on domestic carriers either on the ground of detecting terrorism or finding illegal immigrants attempting to flee the border.

II. SEARCHES OF INDIVIDUALS AT THE BORDER

Immigration officials have the power to conduct searches and seizures at the border through a long-standing exception to the Fourth Amendment known as the “border-search exception.” Under this doctrine, officials may search anyone crossing the border or its “functional equivalent” without a warrant, and in some cases without reasonable suspicion or probable cause, for the sole reason that those individuals have entered the country from elsewhere. The Supreme Court has held that interrogations conducted by immigration officials beyond the border or its “functional equivalent” are subject to normal Fourth Amendment requirements: they must be supported by reasonable suspicion or probable cause, since the reasons for the “border-search exception” no longer apply. In the INA, Congress granted much broader discretion to immigration officials to stop and inquire about individuals’ immigration status. However, as the Supreme Court noted in Almeida-Sanchez v. United States,
a case that addresses immigration officials’ investigative power under the INA, “[N]o Act of Congress can authorize a violation of the Constitution.”

A. The INA and Exceptions to the Fourth Amendment at the Border

Since the First Act of Congress, legislators and the Supreme Court have granted broad leeway to officers searching people, cars, or objects that have just entered the country. In 1789, Congress stated that nothing beyond “suspicion” was necessary to examine items crossing the border. This leeway has been termed the “border-search exception” to the Fourth Amendment bar against “unreasonable searches and seizures.” As the Supreme Court explained while officially upholding the exception in United States v. Ramsey, a border search is “reasonable” within the meaning of the Fourth Amendment for “the single fact that the person or item in question had entered into our country from outside.” Thus, an individual entering the country from outside should expect to be searched and should therefore have a lesser expectation of privacy. Further, vehicles may be searched without a warrant when they cross “an international boundary because of national self-protection reasonably requiring one entering the country to

125 Act of July 31, 1789, ch. 5, 1 Stat. 29, 43 (1789) (“[I]t shall be lawful for the . . . officer of the customs, after entry made of any goods, wares or merchandise, on suspicion of fraud, to open and examine . . . any package or packages thereof . . . If any of the packages so examined be found to differ in their contents from the entry, and it shall appear that such difference hath been made with intention to defraud the revenue, then all the goods, wares or merchandise contained in such package or packages, shall be forfeited . . . .”).
126 Id.
128 Ramsey, 431 U.S. at 619.
129 United States v. Montoya de Hernandez, 473 U.S. 531, 539 (1985) (citations omitted) (“T]he expectation of privacy [is] less at the border than in the interior . . . .”); see also Ramsey, 431 U.S. at 618 (The Court affirmed that there is no warrant requirement for a border search because “a port of entry is not a traveler’s home. . . . Customs officials characteristically inspect luggage and their power to do so is not questioned . . . ; it is an old practice and is intimately associated with excluding illegal articles from the country.” (quoting United States v. Thirty-seven Photographs, 402 U.S. 363, 376 (1971) (internal quotation marks omitted))); United States v. Espericueta-Reyes, 631 F.2d 616, 622 (1980) (noting that “[t]he power to detain persons at the border while their possessions are searched derives from the nation’s right to regulate who and what may enter the Country”); Border Searches and the Fourth Amendment, 77 YALE L.J. 1007, 1012 (1968) (explaining that one of the justifications for the border search exception is that because “the individual crossing a border is on notice that certain types of searches are likely to be made, his privacy is arguably less invaded by those searches”).
identify himself as entitled to come in.”130 Probable cause (much less a warrant) is therefore not a requirement for a search conducted at the border.131

Searches that occur at the “functional equivalent” of the border also fall under the border-search exception.132 The functional equivalent of the border is an area where most individuals have just crossed the border so the justifications for the exception still apply.133 The functional equivalent could include, for example, “an established station near the border, at a point marking the confluence of two or more roads that extend from the border,” or an airplane arriving in the United States on a nonstop flight from Mexico.134 The exception to the Fourth Amendment applies at the “functional equivalent” of the border for the same reason as it applies at the border: for the “single fact” that an individual has entered the country from outside.135

Additionally, at a traffic checkpoint located reasonably close to the border, police officers may conduct a “stop[] and questioning” without reasonable, individualized suspicion or probable cause.136 Although this is looser than typical Fourth Amendment requirements, the rule is again based on the closeness of the search to the border; in upholding it, the Supreme Court argued that the government interest in preventing illegal aliens from using highways to quickly get away from the border outweighed the private interest in privacy.137 Further, the Court reasoned that defendants have a decreased expectation of privacy at a traffic checkpoint as compared to their homes,138 and government agents subject them to less fright or annoyance than when a roving patrol stops them, because the government agents ask individuals only a few questions, and they can see that other vehicles are also getting stopped and questioned.139

When searches are conducted beyond the border, however, the Supreme Court has held that more information is needed to stop, question, and search individuals. Officers in roving patrols beyond the border or its functional equivalent

131 Ramsey, 431 U.S. at 619.
134 Almeida-Sanchez, 413 U.S. at 273.
135 United States v. Jackson, 825 F.2d 853, 858 (5th Cir. 1987) (citations omitted).
137 See id. at 556-57.
138 Id. at 558-62 (citations omitted).
139 Id. at 557-58 (citations omitted).
can only stop and question vehicles when they have “specific articulable facts . . . that reasonably warrant suspicion that the vehicle[] contain[s] [illegal] aliens.”140 Further, police officers in roving patrols or checkpoints beyond the border or its functional equivalent may not search a vehicle without probable cause or consent.141

The INA, however, does not have these same requirements of reasonable suspicion or probable cause. It eliminates the warrant requirement and authorizes immigration officials to “interrogate” anyone believed to be an alien about his or her immigration status.142 Additionally, “within a reasonable distance from” a United States border, officials may board and search “any railway car, aircraft, conveyance, or vehicle” for illegal immigrants.143 The INS had further defined reasonable distance as “within 100 air miles from any external boundary of the United States.”144 Despite this language, the Supreme Court has held that when immigration officials conduct searches beyond the border or its functional equivalent, the investigation must be supported by reasonable suspicion, probable cause, or consent.145

For example, when a roving patrol stopped and questioned passengers in a car “on a California road that lies at all points at least 20 miles north of the Mexican border,” the Court held that this stop was of a “wholly different sort” than an investigation that occurs at the border or its functional equivalent.146 Even though the government was backed by the

---

140 United States v. Brignoni-Ponce, 422 U.S. 873, 884 (1975). Reviewing courts, in determining whether there was reasonable suspicion, look at the “totality of the circumstances” of each case. See United States v. Arvizu, 534 U.S. 266, 273 (2002) (citations omitted). Under this analysis, courts look at whether the officer had a “particularized and objective basis’ for suspecting legal wrongdoing,” id., taking into account the officer’s experience and training, including an immigration official’s experience with particular routes used by illegal aliens and the official’s “experience as a border patrol agent.” See id. at 277 (citations omitted).

Further, an individual’s Mexican ancestry alone does not provide such facts and is therefore an illegal reason to stop and question an individual, although Mexican appearance is a “relevant factor.” Brignoni-Ponce, 422 U.S. at 886-87.141

141 See United States v. Ortiz, 422 U.S. 891, 896-97 (1975) (holding that, because of the significant invasion of privacy in searching private cars, “at traffic checkpoints removed from the border and its functional equivalents, officers may not search private vehicles without consent or probable cause”); Brignoni-Ponce, 422 U.S. at 881-82; Almeida-Sanchez v. United States, 413 U.S. 266, 267-68, 274-75 (1973) (holding that a warrantless search conducted by roving patrol on a road that never intersects the Mexican border, with no probable cause justification, violates the Fourth Amendment).


143 Id. § 1357(a)(3).


145 See Almeida-Sanchez, 413 U.S. at 272-73; Brignoni-Ponce, 422 U.S. at 884.

146 Almeida-Sanchez, 413 U.S. at 273.
INA, the Court stated, specifically in response to their use of the statute, “[N]o Act of Congress can authorize a violation of the Constitution.” The Court held the search of passengers in a vehicle that never intersected the border unconstitutional without probable cause, even though the INA, on its face, authorized the stop. Similarly, transportation checks conducted on domestic common carriers are unconstitutional because the vessels never cross the border and are therefore not subject to the “border-search exception.”

B. Domestic Document Checks Do Not Fall Under the Border-Search Exception

Like the Supreme Court, Congress and the DHS have argued that certain measures, inapplicable in the country’s interior, are necessary at the border. In passing the expedited removal and mandatory detention provisions of the IIRIRA, the House Judiciary Committee argued, “If not detained, the aliens would most often disappear and become long-term illegal residents.” Further, in explaining the expansion of the expedited removal program, the DHS stated that its focus was on “unlawful entries that have a close spatial and temporal nexus to the border”—meaning that this expanded authority is necessary, “because many aliens will arrive in vehicles that speedily depart the border area, and because other recent arrivals will find their way to near-border locales seeking transportation to other locations within the interior of the U.S.” These policies arose because of the belief that more stringent requirements are necessary at the border to prevent aliens who have just entered from fleeing. The government’s explanation of its actions reflects an understanding that there is something different about individuals who have just entered the country from outside.

The government has justified the document checks on public transportation on similar grounds. However, the justification is incongruent with the actual practice. The passengers are questioned while traveling on trains and buses

147 Id. at 272.
148 Id. at 272-73.
151 See supra Part I.C.
that have only domestic routes.\textsuperscript{152} Though they are traveling within one hundred miles of the border, they have not recently crossed the border. Moreover, they are not at the “functional equivalent” of the border, because they have not entered the country from outside—and there is no reason to suspect that they have, since a train or bus traveling domestically never intersects with another country. When the Court gave examples of what may constitute the “functional equivalent” of the border, it provided areas where the majority of the individuals have recently entered the country from elsewhere.\textsuperscript{153} Thus, the individuals stopped during the transportation checks should have the same expectation of privacy as any individual in the interior of the United States. Moreover, the argument that these vessels may contain illegal aliens attempting to flee into the country’s interior is weak, again, because these vehicles have not come from across the border.

III. POLICE QUESTIONING OF INDIVIDUALS

The random questioning has also been justified by the general Fourth Amendment principle that a police officer may approach an individual in a public place and ask that person questions—without any requirement of a warrant, probable cause, or reasonable suspicion—as long as “a reasonable person would understand that he or she could refuse to cooperate.”\textsuperscript{154} The encounter becomes a “seizure” that implicates the Fourth Amendment if the individual questioned believes that he or she is not free to walk away.\textsuperscript{155} The conduct of immigration officials on domestic carriers is coercive and therefore a “seizure” under the Fourth Amendment. As long as CBP agents continue their interrogations without reasonable suspicion or probable cause, they are conducting illegal seizures in violation of the Fourth Amendment.

\textsuperscript{152} See Bernstein, supra note 8.

\textsuperscript{153} See Almeida-Sanchez, 413 U.S. at 273.


\textsuperscript{155} Delgado, 466 U.S. at 215, see Terry v. Ohio, 392 U.S. 1, 16 (1968) (stating that “whenever a police officer accosts an individual and restrains his freedom to walk away, he has ‘seized’ that person”).
A. Defining a “Seizure”

As the Supreme Court noted in Florida v. Royer,156 no “litmus-paper test” determines whether a consensual encounter or a seizure has occurred.157 Instead, courts look at the totality of the circumstances and decide whether, based on all the circumstances, a “reasonable person would have believed he was not free to leave.”158 This is not a question of the subjective experience of the individual, but of the objective factors and what a reasonable person would have concluded from them.159 If the encounter was “so intimidating,”160 or the officer “by means of physical force or show of authority”161 has indicated that the individual is not free to leave, then a “seizure” has occurred within the meaning of the Fourth Amendment.162

Courts have looked at a variety of factors in determining whether a “seizure” has occurred. In United States v. Mendenhall, the Supreme Court articulated several specific factors that may indicate officers’ behavior is coercive and constitutes a seizure.163 These factors include “the threatening presence of several officers, the display of a weapon by an officer . . . or the use of language or tone of voice indicating that” an individual’s compliance is required.164 Further, the Court has held that “[t]he cramped confines of a bus are one relevant factor that” could be used to determine whether the individual being questioned felt that he was free to leave.165

If, in fact, a seizure has occurred, the officers must have “reasonable, articulable suspicion” to justify it166—even if the

---

157 Royer, 460 U.S. at 506-07; see also Ohio v. Robinette, 519 U.S. 33, 39 (1996) (noting that the Court has “consistently eschewed bright-line rules” in determining whether a search is reasonable under the Fourth Amendment, “instead emphasizing the fact-specific nature of the reasonableness inquiry”).
158 Delgado, 466 U.S. at 216; see also United States v. Mendenhall, 446 U.S. 544, 554 (1980).
159 See Delgado, 466 U.S. at 228 (Brennan, J., dissenting).
160 Id. at 216 (majority opinion).
161 Terry v. Ohio, 392 U.S. 1, 19 n.16 (1968).
162 See id.; accord Mendenhall, 446 U.S. at 554 (“As long as the person to whom questions are put remains free to disregard the questions and walk away, there has been no intrusion upon that person’s liberty or privacy as would under the Constitution require some particularized and objective justification.”).
163 446 U.S. 544 (1980).
164 Id. at 554.
165 Florida v. Bostick, 501 U.S. 429, 439 (1991) (The Court noted, however, the fact that the questioning took place on a bus, by itself, is not dispositive.); see also United States v. Drayton, 536 U.S. 194, 204 (2002).
166 Florida v. Royer, 460 U.S. 491, 502 (1983). In his concurrence to Royer, Justice Brennan argued that any time a uniformed police officer approaches an
seizure at issue is “only a brief detention.” This reasonable suspicion standard requires the officer to show specific facts, “which, taken together with rational inferences,” would “reasonably warrant” the seizure. Further, even if the officer has specific facts to justify the seizure, the officer’s conduct is still limited in scope: the officer’s actions must be “reasonably related in scope to the justification” for the seizure. With these rules, the Court has tried to limit police intrusion into individual privacy in an effort to balance important government interests with the individual interests in privacy protected by the Fourth Amendment.

In a few illustrative cases, the Court has applied this totality of the circumstances test to determine whether a “seizure” had occurred, with varying results that indicate changes in policy rather than a consistent application of the Court’s own principles. For example, in INS v. Delgado, INS agents entered a workplace to look for illegal aliens. Agents were positioned near all of the building’s exits and others moved throughout the factory asking employees questions. If employees answered “unsatisfactorily” or stated that they were aliens, the agents asked for their immigration documents. The Court held that in this situation, a seizure had not occurred; even though agents were stationed at all the exits, the individual and requests to see his documents (in this case, the defendant’s airplane ticket), a seizure has occurred, because “[i]t is simply wrong to suggest that a traveler feels free to walk away when he has been approached by individuals who have identified themselves as police officers and asked for, and received, his airline ticket and driver’s license.”

---

167 Mendenhall, 446 U.S. at 551 (citations omitted); see also INS v. Delgado, 466 U.S. 210, 227 (1984) (Brennan, J., dissenting) (“The Fourth Amendment protects an individual’s personal security and privacy from unreasonable interference by the police, even when that interference amounts to no more than a brief stop and questioning concerning one’s identity.”).

168 Terry v. Ohio, 392 U.S. 1, 21 (1968); see also Mendenhall, 446 U.S. at 551.

169 Terry, 392 U.S. at 29. For example, in Terry, the Supreme Court upheld the inclusion at trial of a gun that a police officer obtained by frisking the defendant on the street. The Court held that the personal intrusion was permissible because the officer’s justification for conducting the search—his and bystanders’ protection—was limited in scope to a search aimed at discovering instruments that could be used to assault an officer.

170 See id. at 20-21 (“In order to assess the reasonableness of [the police officer’s] conduct as a general proposition, it is necessary first to focus upon the governmental interest which allegedly justifies official intrusion upon the constitutionally protected interest of the private citizen, for there is no ready test for determining reasonableness other than by balancing the need to search [or seize] against the invasion which the search [or seizure] entails.” (citations omitted)).


172 Id. at 212.

173 Id.

174 Id. at 212-13.
employees could have “simply refused to answer.” The agents were only questioning the employees; this conduct “could hardly result in a reasonable fear that [employees] were not free to continue working or to move about the factory.” Further, while employees may have believed they would be questioned if they tried to leave the building, they should not have reasonably believed “they would be seized or detained in any meaningful way” for doing so. Therefore, looking at the totality of the circumstances, the Court concluded that at all times, a reasonable employee—in a building where INS agents were walking around and standing near all the exits—should have felt free to walk away or refuse to answer, and thus, a seizure had not occurred. No reasonable, individualized suspicion was required to question the individual employees here.

Justice Brennan, in his dissent, argued instead that, based on all the objective factors, a seizure had in fact occurred because “a reasonable person could not help but feel compelled to stop and provide answers to the INS agents’ questions.” Contrary to the majority’s description, the encounter from the viewpoint of the employees was an intimidating show of authority that required individualized suspicion. The inspection was a surprise one, carried out by fifteen to twenty-five agents “who moved systematically through the rows of workers” while showing their badges and asking questions. Employees suspected of being illegal aliens were handcuffed and led away to vans waiting outside, while agents stationed at the exits could prevent others from escaping questioning. Under these “tactics,” the employees could not possibly have felt they could refuse to answer the questions and leave. The agents

175 Id. at 218. The fact that the inspection occurred at a workplace was also a factor in the Court’s decision: Justice Rehnquist argued that, at work, individuals have an obligation to their employers to remain at work—therefore, their freedom of movement is already somewhat restricted. See id. Additionally, Justice Powell in concurrence argued that an employee’s “expectation of privacy in the plant setting . . . certainly is far less than the traditional expectation of privacy in one’s residence.” See id. at 224 (Powell, J., concurring).

176 Id. at 220-21 (majority opinion).
177 Id. at 219.
178 Id. at 212, 220-21.
179 See id. at 221.
180 Id. at 229 (Brennan, J., dissenting).
181 Id. at 229-30.
182 Id. at 230.
183 Id.
184 Id. at 230-31.
therefore needed to have reasonable and particularized suspicion to question each individual.\footnote{Id. at 232-34. Here, there was no individualized suspicion to question anyone—the INS agents were instructed to interrogate “virtually all persons” in the factory. Id. at 233. In response, Justice Brennan argued, “To say that such an indiscriminate policy of mass interrogation is constitutional makes a mockery of the words of the Fourth Amendment.” Id. at 234.}

Finally, Justice Brennan articulated an important distinction for purposes of immigration law—the difference between a legal and an illegal alien.\footnote{Id. at 235.} He argued that “the mere fact that a person is believed to be an alien provides no immediate grounds for suspecting any illegal activity.”\footnote{Id.} In the context of a large-scale inspection like the one at issue, “it is virtually impossible to distinguish fairly between citizens and aliens,” and therefore, the INS needed particularized, reasonable factors to question individuals suspected of being in the country illegally, rather than just to question anyone who may be an alien.\footnote{Id. at 235-36.}

As hinted at by the dissent, the INS v. Delgado decision was motivated by the strong public interest in curbing illegal immigration.\footnote{See César Cuauhtémoc García Hernández, La Migra in the Mirror: Immigration Enforcement and Racial Profiling on the Texas Border, 23 NOTRE DAME J.L. ETHICS & PUB. POL'Y 167, 176-77 n.40 (2009) (“Since a person might not feel free to leave when the only avenue for leaving a location is blocked by law enforcement officials, the Supreme Court’s decision in Delgado that individuals caught in the midst of a workplace inspection by immigration officials were not seized for purposes of the Fourth Amendment suggests that the only distinguishing criterion is that the Delgado incident occurred in the immigration context rather than the traditional criminal law context.”).} The show of force was so strong that it is not reconcilable with the Court’s previously articulated factors in Mendenhall; no reasonable employee in the situation could have felt free to refuse the agents’ questions or to leave the building when officers entered and essentially blocked the exits. Thus, the public policy interest weighed more strongly than the individual interest in privacy. While such a balancing of factors is standard in the Court’s Fourth Amendment cases,\footnote{Terry v. Ohio, 392 U.S. 1, 21 (1968). In Terry, the particular interest at issue was the necessity of the police officer to ensure that the individual he was questioning was not dangerous to himself or the surrounding public. Id. at 23.} it also has a ratcheting effect in that the Court describes more police encounters as “consensual” rather than “seizures”—even if the individual feels they have been detained—as long as the policy interest is strong enough to justify them.
Cases where individuals were stopped during drug searches illustrate the ratcheting effect. In *Florida v. Bostick*, the police officers conducting a drug search on a bus questioned the defendant without any articulable suspicion and then searched his belongings with his consent. The defendant argued that in the “cramped confines” of the bus, where police officers essentially blocked the exits and displayed their badges and firearms, he did not feel free to refuse consent either to the questioning or the search of his belongings. The Court held that the mere fact that the encounter took place on a bus did not constitute a seizure per se.

This practice was again upheld in *United States v. Drayton*, where the Court held that police officers conducting a drug search on a bus are not required to inform passengers of their right to refuse cooperation. In *Drayton*, because the police officers did not “brandish” the weapons they were clearly carrying or “make any intimidating movements,” a reasonable passenger would have felt free not to cooperate with the police. In particular, the Court held that because the public knows most law enforcement officers are armed, “[t]he presence of a holstered firearm thus is unlikely to contribute to the coerciveness of the encounter absent active brandishing of the weapon.”

As the dissent pointed out in *Bostick*, the practice at issue in these cases occurred within the broader context of the “War on Drugs.” By placing the decision within this framework, Justice Marshall hinted that the public policy was an important factor in the majority’s decision and argued that based on the mere facts alone, a seizure had in fact occurred. The dissent asserted that the practice was “intimidating” and coercive, and the presence of the officers was “threatening.”

When armed and uniformed police officers question passengers

192 Id. at 431-32 (The defendant was carrying cocaine in his luggage and was therefore arrested.).
193 Id. at 435.
194 Id. at 438-39. The case was remanded to the lower court for a determination of whether the conduct of the officers in this particular case, based on all the factors (not just that the encounter occurred on a bus), “communicated to a reasonable person that” he was free not to consent. Id. at 439-40.
196 Id. at 206-07.
197 Id. at 203-04.
198 Id. at 205.
200 See id. at 440.
201 Id. at 446-47.
202 *Drayton*, 536 U.S. at 210 (Souter, J., dissenting).
on a bus, effectively blocking the exit, they create “an atmosphere of obligatory participation” in which a reasonable person would not feel free to refuse cooperation.203 The dissent argued that such a seizure requires reasonable suspicion; without that, such bus sweeps “violate[] the core values of the Fourth Amendment.”204 Based on the public interest in finding illegal drugs, the Court in these cases found that a seizure had not occurred—despite factors that, objectively, appeared to be exactly those that did require reasonable suspicion in the Court’s prior opinions.

B. Internal Document Checks on Domestic Vessels Are an Illegal Seizure

The internal document checks violate the “core values” of the Fourth Amendment, because they are conducted without any reasonable or particularized suspicion.205 In Terry, the Supreme Court asserted that a seizure occurs when a police officer, “by means of physical force or show of authority,”206 has restrained an individual’s freedom to “walk away.”207 Under the factors articulated in Mendenhall,208 immigration officials inspecting individual passengers on domestic train and bus routes display a coercive amount of power and thus, their actions constitute a seizure. This practice must include reasonable, particularized suspicion, and it must be limited in scope.

The factors here amount to a coercive practice, where ordinary individuals do not feel free to refuse to answer.209 Fully uniformed and armed immigration officials board trains and buses and ask passengers at random whether they are U.S. citizens.210 In full view of other passengers, they check

203 Id. at 212.
204 Bostick, 501 U.S. at 440 (Marshall, J., dissenting).
205 See id.
206 Terry v. Ohio, 392 U.S. 1, 19 n.16 (1968) (emphasis added).
207 Id. at 16.
208 United States v. Mendenhall, 446 U.S. 544, 554 (1980) (citations omitted); see supra Part II.B.
209 In researching this issue for the New York Times, journalist Nina Bernstein specifically tested whether she could refuse. See Bernstein II, supra note 104. When asked whether she was a citizen, she replied, “I don’t want to answer that question.” Id. The officers moved on. Id. Bernstein, however, was fully aware of the law which allows individuals to refuse consent, and, as she herself pointed out, was “a white woman in jeans who had spoken American English with no accent.” Id. She had no further evidence of other individuals refusing consent, because they all answered the officers’ questions. Id.
210 Id. (“[H]alf a dozen men in green uniforms with pistols on their hips strode down the platform . . . .”).
individuals’ immigration documents and passports\textsuperscript{211} and remove from the train for further questioning those that do not have their papers.\textsuperscript{212} Several immigration officers essentially block the exits, by standing between the passenger and the doors.\textsuperscript{213} Moreover, as the dissent pointed out in \textit{Bostick}, “Because the bus is only temporarily stationed at a point short of its destination, the passengers are in no position to leave as a means of evading the officers’ questioning.”\textsuperscript{214} The officers also create an “intimidating atmosphere,” by walking among passengers and abruptly asking them questions while clearly armed.\textsuperscript{215} Taking away passengers in clear sight of others adds to the officers’ demonstration of authority and power, which would make it difficult for a reasonable person to believe they can refuse to answer the officers’ inquiries.\textsuperscript{216}

Further, to determine whether a police practice violates the Fourth Amendment, courts often balance the government interest at issue with the private interest in privacy.\textsuperscript{217} In \textit{INS v. Delgado} and the drug search cases, the public interest in favor of curbing illegal immigration or fighting the war on drugs, respectively, outweighed the private interest in privacy.\textsuperscript{218} The random document checks conducted on domestic buses and trains are justified primarily by the aim of fighting terrorism. As demonstrated by changes in immigration law and congressional spending after the September 11 attacks, fighting terrorism has become a national priority.\textsuperscript{219} The random passenger surveys conducted on domestic vessels are an outgrowth of those policy changes.

\textsuperscript{211} \textit{Id.}
\textsuperscript{212} \textit{Id.}
\textsuperscript{213} See United States v. Drayton, 536 U.S. 194, 211 (2002) (Souter, J., dissenting) (“The officers took control of the entire passenger compartment... The reasonable inference was that the ‘interdiction’ was not a consensual exercise...”).
\textsuperscript{215} See INS v. Delgado, 466 U.S. 210, 230, 234 (1984) (Brennan, J., dissenting) (“To say that such an indiscriminate policy of mass interrogation is constitutional makes a mockery of the words of the Fourth Amendment.”); see also NYCLU REPORT, \textit{supra} note 9, at 21 (arguing that “when an armed agent questions passengers on a train or bus, sometimes in the middle of the night with a flashlight glaring at the rider’s face, few individuals would feel that they have the right to refuse to answer the agent’s questions”).
\textsuperscript{216} See \textit{Delgado}, 466 U.S. at 230 (Brennan, J., dissenting).
\textsuperscript{217} See United States v. Martinez-Fuerte, 428 U.S. 543, 555 (1976) (“In delineating the constitutional safeguards applicable in particular contexts, the Court has weighed the public interest against the Fourth Amendment interest of the individual...” (citations omitted)); Terry v. Ohio, 392 U.S. 1, 20-21 (1968).
\textsuperscript{218} See \textit{supra} Part III.A.
\textsuperscript{219} See \textit{supra} Part I.B and C.
Although fighting terrorism is vital, it is questionable whether these types of tactics actually work towards that goal. Most of the arrests that occur as a result of these document checks are for immigration, not criminal, violations; a minuscule percentage are at all related to terrorism. Fighting terrorism is a critical goal—but it is not accomplished by these types of random searches.

On the other hand, the potential for private harm is huge. Individuals who cannot present proper immigration documents are removed from the train or bus and taken to detention facilities or local prisons for further questioning and investigation. According to the NYCLU’s Report, the majority of individuals who were arrested and detained could not be released without posting a bond, which could range from $1500 to $20,000. The Report postulated that as those arrested had just been traveling, they likely did not have such large amounts of cash on them and probably had to wait several days before being released. Because the Supreme Court has upheld the constitutionality of mandatory detention, once these individuals have been detained, they likely have few legal options that would allow them to be released quickly. Further, under the INA’s expedited removal provisions, once they are arrested, certain individuals may even face immediate deportation without judicial or administrative review of their case. Finally, the “exclusionary rule,” which holds that evidence found in violation of the Fourth Amendment must be excluded from actions by the government against an individual, does not apply to civil deportation hearings.
immigration agents act unlawfully, unlike in an ordinary criminal trial, any illegal immigrant arrested as a result of an illegal search will still be subject to deportation. Further, many commentators have raised concerns about whether this practice amounts to racial profiling. Although immigration agents assert that they randomly question people and do not base suspects on race, at present no limitations ensure that racial profiling is not occurring.

Therefore, although the justification for these transportation checks is, on its face, without fault, the actual results have not borne out this justification. Moreover, the risk of private harm far outweighs the public utility of this practice. Private individuals that are randomly questioned by this procedure, if they are noncitizens and not carrying proper documentation, may be subjected to detention and deportation—and the possible unlawfulness of the procedure will not protect them. This is unjust both because many of the individuals arrested do have lawful status to be in the United States, and because many have been here for long periods of time and have established ties to the country and the right to more constitutional protection than someone entering for the first time. Congress and the DHS should both institute procedures and amendments to ensure that immigration officials proceed carefully and are limited in their actions.

IV. PROPOSAL

Congress must act to restrict the practice of randomly questioning individuals traveling on domestic buses and trains.

---

228 Id. at 1050-51. This is because an immigration violation is considered an ongoing violation, and the Court has argued that,

Applying the exclusionary rule in proceedings that are intended not to punish past transgressions but to prevent their continuance or renewal would require the courts to close their eyes to ongoing violations of the law. This Court has never before accepted costs of this character in applying the exclusionary rule.

Id. at 1046.

229 See NYCLU REPORT, supra note 9, at 7, 16; Bazar, supra note 103.

230 See NYCLU REPORT, supra note 9, at 26 (“While CBP has refused to release its training materials on racial and ethnic profiling, accounts of its operations raise serious concerns that Border Patrol agents resort to racial and ethnic profiling techniques to determine who to stop, question or arrest. Such accounts indicate that even if CBP policy expressly forbids racial and ethnic profiling, additional guidance and training of Border Patrol officers is necessary to ensure appropriate compliance.”)

231 Id. at 6-7.

232 See supra notes 24-27 and accompanying text.
about their immigration status by amending the INA to define the parameters of a permissible stop and inquiry more narrowly. The Supreme Court in United States v. Brignoni-Ponce held that beyond the border or its functional equivalent, officers in roving patrols can only stop individuals if they have reasonable suspicion to believe they are illegal aliens. This same standard should be applied to the transportation checks at issue here. Border Patrol agents must have reasonable suspicion to stop and question someone on the train or bus about their immigration status.

Further, as in Brignoni-Ponce and Terry, the search must be limited in scope. There is a difference between believing that someone is an alien, an illegal alien, or a terrorist. If immigration agents believe that such transportation checks are necessary to stop terrorism and to curb illegal immigration, they must have reasonable suspicion—specifically, articulable facts—to believe that the individual they are questioning is either a terrorist or an illegal alien.

The INA itself provides a model for the framing of this amendment. The INA states that immigration officials have the authority to “conduct a search, without warrant, of the person, and of the personal effects in the possession of any person seeking admission to the United States, concerning whom such officer or employee may have reasonable cause to suspect that grounds exist for denial of admission to the United States.” The “reasonable cause” limitation should be added to the other parts of the statute as well, requiring that officers may interrogate individuals suspected of being illegal aliens, or board vessels to search for illegal aliens, only if they have “reasonable cause” to suspect that the individual is an illegal alien or a terrorist.

Finally, Congress should act pragmatically when enacting this amendment by requiring that the DHS train immigration officials on determining what constitutes reasonable suspicion and when they may question individuals on suspicion that they

---

233 422 U.S. 873 (1975).
234 Id. at 881-82. Specifically, officers must be “aware of specific articulable facts, together with rational inferences from those facts, that reasonably warrant suspicion that the vehicles contain aliens who may be illegally in the country.” Id. at 884.
235 See NYCLU REPORT, supra note 9, at 25-26.
236 See Brignoni-Ponce, 422 U.S. at 881; Terry v. Ohio, 392 U.S. 1, 29 (1968) (holding that a stop and search inquiry must be “reasonably related in scope to the justification for [the] initiation”).
are in the country illegally or are involved in terrorism. In particular, this training should address the issue of racial profiling to ensure that the officers’ determinations are based on more than race. This training program can be modeled on the USCIS Asylum Division Training Section, a “national training course that is specific to asylum adjudicators.” Asylum officers are required to attend periodic trainings both in national offices and in their regional offices that educate them on the relevant asylum law, how to properly adjudicate applications, and interviewing and writing skills, among other topics. Officers are also required to attend periodic trainings to update their skills and their knowledge of the field.

A similar training program can be instituted for Border Patrol agents who investigate domestic trains and buses. Like the asylum officers, Border Patrol officers can be required to learn about Fourth Amendment case law and its proper application in the field. Further, officers can learn the proper evidentiary standards required to question passengers and the factors that may have more relevance to making a decision about investigating passengers. Officers can also receive training on questioning passengers in less obtrusive ways.

Although the Border Patrol has recently taken steps to curb the practice of searching domestic trains and buses for illegal immigrants, and has stated that it will conduct searches only when it has information regarding a threat, these changes have not been codified in any formal agency policy, rule, or law. To ensure that long-term changes are made to the practice, Congress should enact formal changes to the INA, and the DHS should put in place ongoing training courses for its Border Patrol agents.

The Fourth Amendment protects individuals from unreasonable searches and seizures. When immigration officials, with no articulable reason, question passengers traveling on domestic train and bus routes about their immigration status, they violate this fundamental mandate of the U.S. Constitution.

---

239 See Brignoni-Ponce, 422 U.S. at 885-87 (stating that an individual’s Mexican ancestry, alone, does not provide such facts and is therefore an illegal reason to stop and question an individual, although Mexican appearance is a “relevant factor”).

240 Asylum Division Training Programs, U.S. CITIZENSHIP & IMMIGR. SERVS., http://www.uscis.gov/portal/site/uscis/menuitem.5af9bb59519e35e666f14176543f6d1a?vgnextoid=2a1d1a877b4be110VgnVCM1000004718190aRCRD&vgnextchannel=3a82ef4c766fd010VgnVCM100000ecd190aRCRD (last updated Apr. 12, 2011).

241 Id.

242 Id.

243 NBPC Press Release, supra note 34.

244 U.S. Relaxes Canadian Border Checks as Agents Are Told to Stop Searching Buses, Trains and Planes for Illegal Immigrants, supra note 32.
Congress and the DHS must work together to set clearer limitations on the practice and to train immigration officials in the proper ways to conduct investigations.

CONCLUSION

In the years following the September 11 terrorist attacks, immigration became linked with terrorist activity in the public mind. Congress and the President responded by enacting policies to curb illegal immigration with the principal aim of fighting terrorism. In particular, they established one agency, the DHS, to oversee both immigration and national security and made explicit the administration’s view that immigration and terrorism were intricately linked. Congress also increased funding to the newly-created government agencies whose purpose became to secure the national borders in order to fight terrorism.

As a result of these actions, the fear of terrorism and terrorists has become a fear of illegal immigration, and the overall climate of fighting illegal immigration has led to a policy whereby immigration agents board domestic trains and buses and freely question all passengers regarding their citizenship and immigration status. This practice has been justified as a terrorism-fighting measure, but it has not resulted in the capture of any terrorists. Instead, agents have arrested students and other individuals who have resided in the country for years.245

President Barack Obama has argued that individuals who have been in the United States since they were children should not be treated in the same way as other illegal immigrants.246 He supported passage of the Dream Act, legislation that would have enabled individuals that had been in the country since childhood and had completed college or military service in the United States to become citizens.247 Though the bill was defeated in the Senate,248 the official sanction for such a law demonstrates that not all aliens, who may technically be illegal, should be treated alike.

---

245 Bernstein, supra note 8.
246 Herszenhorn, supra note 27.
247 Id.
248 Id.
The random interrogation of passengers traveling on domestic routes unjustifiably subjects them to an unconstitutional intrusion on their personal privacy. Legislative and administrative action must limit this practice in its scope.

*Dina Kleyman†

† J.D., Brooklyn Law School, 2012; B.A., New York University, 2006. Thank you to the members of the *Brooklyn Law Review*, to Professor Maryellen Fullerton, who suggested this topic and provided invaluable guidance in thinking it through, and to Professor Dan Smulian, for his example in how to be an ethical, compassionate lawyer. Thank you also and always to my family and friends, for their constant love, support, and clear-headed advice.
Gods Behind Bars

PRISON GANGS, DUE PROCESS, AND THE FIRST AMENDMENT

I fear for the 85 who don’t got a clue.

-Method Man¹

INTRODUCTION

In 1963, Clarence Smith, at the time known as Clarence 13X, was kicked out of the Nation of Islam (NOI) for questioning the divinity of Wallace Fard Muhammad (or possibly for refusing to give up gambling, or any number of other violations).² The following year, after being shot in a gambling den in Harlem known as “The Hole,” he changed his name to Allah (an acronym for “Arm, Leg, Leg, Arm, Head”) and began gathering a group of teenagers around him and preaching his unique version of NOI theology.³ From these humble beginnings, Allah’s Nation would spread from its spiritual homeland in Harlem throughout the country by way of prisons and hip-hop, becoming a significant cultural force.⁴ Allah taught that 85 percent of the people are ignorant and are subjugated by the 10 percent, the governments and corporations.⁵

¹ METHOD MAN, Raw Hide, on RETURN TO THE 36 CHAMBERS (Ol’ Dirty Bastard, Elektra 1995).
³ Id. at xiii. The Nation of Islam, founded in roughly 1931 in Detroit by Wallace Fard Muhammad, as clarified and expanded by Elijah Muhammad, expounded a spirituality based on general Muslim principles, filtered through an African-America perspective. See generally A Brief History on the Origin of the Nation of Islam in America, NATION OF ISLAM, http://www.noi.org/about.shtml (last visited Jan. 16, 2011). While Allah borrowed from the NOI’s emphasis on the divinity of the black man, he deemphasized the supernatural elements of Islam (and denied Wallace Fard Muhammad’s divinity), while also expressing less concern than Elijah about clean lifestyles and abstaining from drugs and alcohol. KNIGHT, supra note 2, at 114 (discussing Allah’s conflicted views of the value of drugs). But see id. at 60 (describing young Five Percenters abstaining from pork). The relationship between the NOI and Allah’s Five Percenters is complex. See generally id. ch. 3 (exploring Clarence/Allah’s relationship with Malcolm X and the Harlem Temple of the NOI). For a more thorough discussion of Five-Percenter beliefs, see infra Part II.A.
⁴ KNIGHT, supra note 2, at xiii.
⁵ Id. at 37.
remaining 5 percent are the poor righteous teachers, who reject the 10 percent and their “mystery god,” and know that god exists in themselves, as “Asiatic” black men.\textsuperscript{6} They devote themselves to the search for universal knowledge and truth through Allah’s system of divine mathematics.\textsuperscript{7} Allah’s followers called themselves the Nation of Gods and Earths (as all male believers are “Gods” and women are “Earths”),\textsuperscript{8} or the Five-Percent Nation, generally shortened to Five Percenters.

Meanwhile, in 1966, George Lester Jackson was serving time in San Quentin State Prison for armed robbery when he became interested in the Black Panthers and Marxist Revolutionary ideology.\textsuperscript{9} He and several fellow revolutionary inmates formed a group called the Black Guerilla Family (BGF), a loosely organized black Marxist organization based in prison, with the goals of protecting the safety and dignity of black men in prison and, eventually, overthrowing the United States government.\textsuperscript{10} Allah was murdered in 1969 in a Harlem housing project,\textsuperscript{11} and Jackson was shot and killed by a prison guard during an alleged escape attempt in 1971.\textsuperscript{12} The two share little in common beyond being espousers of different varieties of Afrocentric philosophy; but, today, the organizations are linked by their shared large presence in the American prison system. Also, many states consider both organizations dangerous prison gangs.\textsuperscript{13}

During the 1990s, in light of high crime rates and growing public fear as urban street gangs spread across the country,\textsuperscript{14} prison officials began turning to the problem of gang violence among prisoners.\textsuperscript{15} In response to the nationalization of some street and prison gangs, the FBI developed a National Gang Strategy in 1993, encouraging the cooperation of law enforcement and corrections officials to control gang violence.\textsuperscript{16}

\begin{thebibliography}{9}
\bibitem{6} Id.
\bibitem{8} K NIGHT, supra note 2, at 215.
\bibitem{9} IMPRISONED INTELLECTUALS 84-85 (Joy James ed., 2003).
\bibitem{10} Id. at 85.
\bibitem{11} IMPRISONED INTELLECTUALS, supra note 2, at 120.
\bibitem{12} IMPRISONED INTELLECTUALS, supra note 9, at 85-86.
\bibitem{13} See infra Part II.
\bibitem{14} See, e.g., Gary Lee, Big-City-Style Gangs Find a New Frontier on Plains of Kansas, L.A. TIMES, Aug. 8, 1993, at 4.
\end{thebibliography}
The focus on controlling gang activity in prison was unsurprising, given the historical connection between prisons and street gangs; certain large street gangs, such as the Mexican Mafia, originated within the corrections system. While states varied in their approaches to gang management, by the mid-90s, many had adopted a comprehensive “Security Threat Group” management system to control and monitor gang activity inside prison walls.

The consequences of an inmate’s designation as a member of a security threat group (STG) vary from state to state. But at least fourteen states, using a system of administrative (as opposed to disciplinary) segregation, isolate suspected gang members from the general prison population. The process of assigning an inmate STG status also varies from state to state, but many have followed California’s “gang validation” model, which assigns point values to criteria ranging from the prisoners’ admission of gang membership or fraternization with known gang members, to possession of gang-affiliated tattoos, literature, or photographs. In many cases, the only way for a prisoner to remove gang affiliation is to go through a “debriefing” process, in which the prisoner repudiates his or her gang activity, gives a full recounting of his or her gang activities, and, generally, names other unvalidated gang members.

While STG systems inevitably restrict the constitutional rights of inmates, common sense dictates that upon being sentenced to prison, a person must expect to surrender a portion of the rights that are protected in outside society. The
Supreme Court, however, has not always approached prisoners’ constitutional rights through the prism of which rights prisoners retain. Rather, the Court has accepted that prisoners have constitutional rights and has considered the nature of a prisoner’s particular asserted right weighed against the strong state interest in managing a prison system.22 Although this approach seems protective of prisoners’ rights, in practice, the courts show extreme deference to the prison administration in the large majority of claims.23

The test for infringement of prisoners’ rights, based on the Due Process clauses of the Fifth and Fourteenth Amendments,24 was clearly laid out by the Supreme Court in 

**Turner v. Safley.**

Holding that a restriction on prisoners’ rights must be “related to legitimate penological interests,” the Court established a four-part test to be used in analyzing claims:26 the restriction must be reasonably related to the stated goal; it must leave alternate means for the prisoner to exercise the right in question; the risks caused by accommodation must be significant; and the prison cannot have any easily available alternatives that would achieve its goal.27 In practice, as long as the first prong—the rational relation test—is met, courts tend to find that the others are met as well.28

This note will begin by examining the *Turner* test in some detail in Part I to explore the likelihood of a suspected member of an STG prevailing in a challenge of his or her STG status.29 Because courts have uniformly found STG programs permissible under due process, prisoners can generally only prevail by showing that the prison’s application of its gang validation standards was improper.30 The discussion will then

---


23 *See, e.g.*, Bull v. City & Cnty. of S.F., 595 F.3d 964, 999 (9th Cir. 2010) (“[D]eference to prison administrators is instrumental in maintaining prison security.”); *see also infra* Part II.A.

24 U.S. Const. amend. V; id. amend. XIV, § 1. While the Eighth Amendment is implicated by decisions involving prisoners’ rights, this note will mostly focus on the due process aspects.

25 *Turner*, 482 U.S. 78.

26 *Id.* at 89.

27 *Id.* at 89-90.

28 *See infra* Part I.A.

29 A full analysis of the adequacy of the due process protections afforded by prisons is beyond the scope of this note, but has been written about extensively by other commentators. *See, e.g.*, Tachiki, *supra* note 21.

30 *See infra* Part I.B.
be narrowed, in Part II, to the more specific and unique cases of the Five Percenters and Black Guerilla Family. These groups should present a more difficult question for the courts when applying the Turner analysis because, as a pseudo-religious group and a political organization, respectively, the First Amendment interests at stake are considerably higher than those of traditional prison gang members. Furthermore, and warranting extra scrutiny of Five Percenters’ religious rights in prison, the War on Terror has already cast the Muslim population into an unfavorable light. Because associational rights are quite understandably the First Amendment rights most limited by incarceration,\(^{31}\) identification of First Amendment rights with stronger protections is essential to this analysis. Because the Five Percenters and BGF are religious and political organizations, respectively, those rights are present.

Parts II.A.1 and II.B.1 will discuss the history and ideology of the Five Percenters and Black Guerilla Family to provide a foundation for the First Amendment issues to follow. Parts II.A.2 and II.B.2 will consider the approach taken by the courts in cases involving Five Percenters and BGF members to date and will attempt to show that the Turner test, because it lacks any way to accommodate heightened First Amendment interests in religious and political speech, is unable to adequately protect the rights of these groups. When considered in light of the severity of the deprivations faced by inmates who find themselves tangled in the STG apparatus,\(^ {32}\) discussed in Part III, this flaw of the Turner test can lead to a severe failure of due process. Part IV will synthesize this information to illustrate the specific failings of the Turner test as applied to the Five Percenters and BGF and similar groups, and will suggest alternate approaches the courts could use to reach an outcome that better balances the legitimate penological interests of the prison with the legitimate First Amendment interests of the prisoners.


\(^{32}\) While a full examination of the implications of one such deprivation, indefinite solitary confinement, is well beyond the scope of this note, it has been the subject of much recent discussion by legal scholars. See, e.g., Jules Lobel, Prolonged Solitary Confinement and the Constitution, 11 U. PA. J. CONST. L. 115 (2008). For excellent ongoing analysis of the social and legal issues raised by the use of solitary confinement, see James Ridgeway & Jean Casella, How Many Prisoners Are in Solitary Confinement in the United States?, SOLITARY WATCH, http://www.solitarywatch.com (last visited Feb. 2, 2012). Part IV, infra, contains a brief discussion of the issue.
I. DUE PROCESS: PRISONERS’ CLAIMS OF DEPRIVATION OF CONSTITUTIONAL RIGHTS

A. The Turner Test

While “[p]rison walls do not form a barrier separating prison inmates from the protections of the Constitution,” courts have not been particularly friendly to claims brought by prisoners alleging a violation of their constitutional rights. *Turner v. Safley* laid out the framework for assessing prisoner claims. The U.S. Supreme Court held in *Turner* that “when a prison regulation impinges on inmates’ constitutional rights, the regulation is valid if it is reasonably related to legitimate penological interests.” To determine if the restriction in question is “reasonably related to legitimate penological interests,” *Turner* requires a four-prong analysis: first, the regulation must have a “valid, rational connection” to the penological interest used to justify it; second, the prison must show that there are “alternative means of exercising the right” available to the prisoner; third, the court weighs the potential impact on other prisoners and staff if the accommodation is allowed; and finally, the prison must not have easy alternatives available to achieve the same interest without restricting prisoners’ rights.

Although the *Turner* test seems protective (the right of prisoners to marry was upheld in that case), in application, it has been extremely difficult for prisoners to succeed on constitutional claims. The first prong (a rational connection to a valid penological goal) seems to have considerably more weight than the other three. The “alternative means of exercising” prong is subject to flexible interpretation; even in a case where there was clearly no alternative means, the Supreme Court held that while this was evidence of unreasonableness, it

---

34 *Turner*, 482 U.S. 78. *Turner v. Safley* involved two different deprivations within prison: Missouri prisoners challenged, in a class action suit, the constitutionality of strict limits on inmate-to-inmate correspondence, and the constitutionality of a prison policy requiring approval from the superintendent before an inmate can be married. *Id.* at 81-82. The correspondence restrictions were upheld by the Court in Justice O’Connor’s opinion, but the marriage restriction was held to be unreasonable. *Id.* at 81.
35 *Id.* at 89.
36 *Id.* at 89-90.
37 *Id.* at 78.
38 See Beard v. Banks, 548 U.S. 521, 532 (2006) (“In fact, the second, third, and fourth factors, being in a sense logically related to the Policy itself, here add little, one way or another, to the first factor’s basic logical rationale.”).
was “not conclusive of the reasonableness of the Policy.”\textsuperscript{39} The fourth prong was limited by the \textit{Turner} decision itself, which emphasized that the Court was not applying a “least restrictive alternative” test. Rather, the Court was applying a test in which the Court would only make sure the prison regulation was not an “exaggerated response to prison concerns.”\textsuperscript{40} Under the \textit{Turner} test, the Supreme Court has upheld prison regulations that denied access to any “newspapers, magazines, or personal photographs,”\textsuperscript{41} prevented Muslim prisoners from attending a religiously commanded Friday evening prayer service,\textsuperscript{42} severely restricted visitation rights,\textsuperscript{43} imposed up to sixteen-day delays in access to legal materials,\textsuperscript{44} and subjected an inmate to treatment with antipsychotic drugs against his will.\textsuperscript{45}

\textsuperscript{39} Id. (citations omitted).

\textsuperscript{40} \textit{Turner}, 482 U.S. at 90; see also Overton v. Bazzetta, 539 U.S. 126, 136 (2003).

\textsuperscript{41} \textit{Beard}, 548 U.S. at 526. In this case, prisoners in restricted confinement in Pennsylvania prisons challenged a prison regulation denying such prisoners access to any “newspapers, magazines, and photographs” on the grounds that it served no legitimate penal interest. Id. at 527. The Third Circuit agreed, id. at 528, but the Supreme Court reversed, finding persuasive the prison’s claim that the regulations served the goal of “motivating better behavior.” Id. at 531 (internal quotation marks omitted).

\textsuperscript{42} \textit{O’Lone} v. Estate of Shabazz, 482 U.S. 342 (1987). Muslim prisoners assigned to outside work duty sued on the grounds that the security policies of the prison prevented them from attending certain Muslim religious ceremonies. Id. at 347. While the Third Circuit found that prison officials should have the burden to show there were no other reasonable methods available to achieve their goals, the Supreme Court refused to endorse their reasoning. Id. at 350. The Court found the \textit{Turner} prongs to be met, and emphasized a refusal to allow First Amendment concerns to overcome the strong presumption against judicial involvement in prison policy decisions. Id. at 352-53.

\textsuperscript{43} \textit{Overton}, 539 U.S. 126. In 1995, Michigan revised their prison visitation policy to allow inmates to receive visits from only immediate family members and an approved list of ten other individuals. Id. at 129. Children were not allowed to be listed, unless they were “children, stepchildren, grandchildren, or siblings of the inmate.” Id. While this prevented inmates from seeing nieces, nephews, or children over which the inmate no longer had parental rights, the Court found the prison’s interests in maintaining security and protecting children to satisfy \textit{Turner}. Id. at 133. The Court also upheld a near-complete ban on visitation for inmates with substance-abuse violations. Id. at 134.

\textsuperscript{44} Lewis v. Casey, 518 U.S. 343, 362 (1996). With fairly little discussion, Justice Scalia struck down a District Court injunction. He wrote that restrictions on dangerous prisoners’ access to law libraries and long delays in receiving legal mail were not problematic “so long as they are the product of prison regulations reasonably related to legitimate penological interests.” Id. Evidencing a common train of thought in these cases, at least among a segment of the Court, Justice Scalia wrote that the “wildly-intrusive” injunction represented “the \textit{ne plus ultra} of what our opinions have lamented as a court’s in the name of the Constitution, becoming . . . enmeshed in the minutiae of prison operations.” Id. (citations omitted).

\textsuperscript{45} \textit{Washington} v. Harper, 494 U.S. 210, 225 (1990). Respondent, a mentally ill prisoner in Washington State Penitentiary, began refusing antipsychotic medication after having consented to treatment for several years. Id. at 214. After an administrative hearing, the prison found respondent to be dangerous and ordered treatment to continue against his will. Id. at 217. Respondent filed a § 1983 action alleging that the administrative hearing, without judicial review, violated his due process
Disciplinary, as opposed to regulatory, decisions are
governed by the even more lenient standard of \textit{Superintendent v. Hill}.\footnote{46} Under \textit{Hill}, courts will not overturn a prison’s decision under due process so long as “some evidence supports the decision by the prison disciplinary board.”\footnote{47} The Supreme Court emphasized that an analysis under the \textit{Superintendent v. Hill} formula “does not require examination of the entire record,” but can be met as long as there is “any evidence in the record that could support the conclusion reached by the disciplinary board.”\footnote{48} Depending on the court’s understanding of the STG program, the \textit{Superintendent v. Hill} standard is sometimes invoked instead of \textit{Turner}.ootnote{49} Because of the degree of deference given under \textit{Turner}, however, the outcome is generally substantially the same.

Under either test, the Supreme Court has emphasized
that prison administrators are due considerable deference. Writing for the majority in \textit{Overton v. Bazzetta}, Justice Kennedy stated, “We must accord substantial deference to the professional judgment of prison administrators, who bear a significant responsibility for defining the legitimate goals of a corrections system and for determining the most appropriate means to accomplish them.”\footnote{50} Construing \textit{Overton}, Justice Souter wrote for the plurality in \textit{Beard v. Banks} that, even on summary judgment, “disputed matters of professional judgment” should be considered with “deference to the views of prison authorities.”\footnote{51} Lower courts have been happy to comply.\footnote{52}

The application of \textit{Turner}, therefore, is not nearly as protective of prisoners’ rights as its plain language would
suggest. Because of the Supreme Court’s continued emphasis
on deference to the decisions of prison authorities, much of the
force of the decision has been neutered. The Court’s refusal to
consider seriously the strength of the First Amendment
interests implicated by prison policies, coupled with the test’s
lack of an explicit mechanism to do so, has left even strong
First Amendment claims with little legal precedent to hang on.

B. Turner and Security Threat Group Designation

Courts have largely approved STG systems under the
deferential test created by Turner. States use a variety of
systems to designate inmates as STG members. In California,
prisoners must be “validated” as gang members, a process
which requires corrections staff to show three independent
points of evidence demonstrative of gang membership, one of
which is a direct link to a current validated gang member, to
assign them STG status. This evidence can be written
material, tattoos, “hand signs,” “distinctive clothing,”
photographs, staff monitoring of communication with other
prisoners, possession of names or addresses of other gang
members, information from other agencies, evidence in the
prisoners’ trial transcripts of gang activity, or visitors with
gang ties. There seems to be no specific requirement of when

---

53 See supra notes 40-44, 50-52.
54 See supra note 39.
55 See, e.g., Bruce v. Ylst, 351 F.3d 1283, 1287 (9th Cir. 2003).
56 CAL. CODE REGS. tit. 15, § 3378(c)(3) (2009). New inmates are investigated by
a gang coordinator upon admission. Id. § 3378(c). Inmates are entitled to an interview, and
all validation materials are disclosed to the inmate before the hearing. Id. § 3378(c)(6). Any
information given by the inmate is added to the gang validation package, and the inmate is
given a written report within fourteen days of the hearing. Id. § 3378(c)(6)(D). The
validation package is reviewed by the chief of the Office of Correctional Safety, id.
§ 3378(c)(6), and is subsequently reviewed annually. Id. § 3378(c)(7).
57 Id. § 3378(c)(8). Recent cases have dealt with validations based on
possession of newspapers belonging to a validated gang member and a
diary/autobiography mentioning the founder of a prison gang, Ellis v. Cambra, No.
1:02-cv-05646-AWI-SMS PC, 2010 WL 4137158, at *5-6 (E.D. Cal. Oct. 19, 2010); a
signature on a birthday card for a validated gang member and a Meso-American
symbol, allegedly used by the Mexican Mafia, drawn on a piece of paper, Treglia v. Dir.
24, 2010); and perhaps more typically, a combination of anonymous informants and
correspondence with validated gang members, Gray v. Woodford, Civ. No. 05-CV-1475
MMA (CAB), 2010 WL 2231805, at *8 (S.D. Cal. Feb. 10, 2010). While at least one court
expressed concern that the validation hearings were perfunctory, “hollow gestures,” it
refused to say they are meaningless. Madrid v. Gomez, 889 F. Supp. 1146, 1276 (N.D.
Cal. 1995); see also Aviña v. Medellin, No. CIV S-02-2661-FCD KJM P, 2010 WL
3516345, at *7 (E.D. Cal. Sept. 3, 2010) (finding that plaintiff had not received due
process during the validation process).
validation proceedings may be held; claims of retaliatory validation proceedings are common.\textsuperscript{58} Inmates may be declassified after demonstrating no gang involvement for two years, in the case of prisoners housed in the general population, or six years in the case of prisoners in Secure Housing Units.\textsuperscript{58} While California’s rules are quite developed, and New Jersey uses a similar approach,\textsuperscript{60} other states leave prison staff essentially unfettered discretion in determining who and what constitutes a security threat group.\textsuperscript{61}

Unsurprisingly, given the deference courts show to prisons’ administrative decisions, challenging the deprivations inherent in STG designation is not often successful. The Ninth Circuit has taken a hands-off approach to California’s “gang validation” process, which often leads to indefinite solitary confinement in “Secure Housing Units.”\textsuperscript{62} Concluding that the decision to assign suspected gang members to solitary confinement concerned administrative—and not disciplinary—segregation, the court has held that “the assignment of inmates within the California prisons is essentially a matter of administrative discretion.”\textsuperscript{63}

Not all challenges are unsuccessful, however. After years of litigation, a former inmate, who spent eight years in solitary confinement for gang affiliation, convinced a district court that his due process rights were violated: he never received an initial hearing before the validation process began, and the evidence used to validate him failed to meet the “some evidence” standard dictated in \textit{Superintendent v. Hill}.\textsuperscript{64} Where courts find for the prisoner, they generally do so based on insufficient procedural due process, not substantive grounds.\textsuperscript{65}


\textsuperscript{59} \textit{CAL. CODE REGS. tit. 15, § 3378(d)-(e)}.

\textsuperscript{60} \textit{See} \textit{Blyther v. N.J. Dep’t of Corr.}, 730 A.2d 396, 398 (1999).

\textsuperscript{61} \textit{See}, e.g., \textit{COLO. REV. STAT. § 17-1-109(2) (2006)}; \textit{TEX. GOV'T CODE ANN. § 508.1141 (West 2010)}.

\textsuperscript{62} \textit{Bruce}, 351 F.3d at 1287; \textit{see also} Michael Montgomery, \textit{Ex-Prisoner Sues California over Years in Solitary}, NAT'L PUB. RADI0: ALL THINGS CONSIDERED (Mar. 8, 2009), http://www.npr.org/templates/story/story.php?storyId=101501841.

\textsuperscript{63} \textit{Bruce}, 351 F.3d at 1287 (quoting Munoz v. Rowland, 104 F.3d 1096, 1098 (9th Cit. 1997) (internal quotation marks omitted)).

\textsuperscript{64} \textit{Lira v. Cate}, No. C 00-0905 SI, 2010 WL 727979, at *2 (N.D. Cal. Sept. 30, 2009).

\textsuperscript{65} \textit{See}, e.g., Aviña v. Medellín, No. CIV S-02-2661-FCD KJM P, 2010 WL 3516343, at *8 (E.D. Cal. Sept. 3, 2010) (holding that, in the absence of prisoner receiving a timely hearing, the prison’s showing of “some evidence” to support gang validation was “moot”); \textit{Lira}, 2010 WL 727979, at *2.
Cases like those are the exception rather than the rule. The Third Circuit applied *Turner* to STG classification in New Jersey in *Fraise v. Terhune* and found that the first three factors “weigh[ed] strongly in favor” of the prison policy, while the fourth was close enough to pass the test.66 The Fourth Circuit did the same in South Carolina.67 Having found that in the case of an “individualized” decision a full adjudicative hearing may be required, New Jersey state courts apply a slightly higher level of due process protection.68 Nonetheless, The New Jersey courts have treated STGs similarly to the Ninth Circuit in California, finding that, because the designations are administrative and nondisciplinary in nature, the prison’s assignment procedures satisfy due process.69

C. Other Legal Approaches to STGs

Prisoners have tried—generally unsuccessfully—to challenge the STG system via other legal routes as well. Eighth Amendment claims, for example, are occasionally attempted.70 The standard for cruel and unusual punishment claims based on prison conditions is quite high: the Supreme Court has held that, short of a showing of “deprivations denying ‘the minimal civilized measure of life’s necessities,’” prisoners’ claims must fail.71 In the absence of any deprivations rising to that level, claims that administrative segregation violates the Eighth Amendment have not survived summary judgment.72 Accordingly, STG classification does not seem to raise any serious Eighth Amendment concerns, notwithstanding some creatively framed arguments to the contrary.73

Other prisoners have attempted to use the Fifth Amendment privilege from self-incrimination to challenge the

---

66 *Fraise v. Terhune*, 283 F.3d 506, 521 (3d Cir. 2002); see also infra notes 103-05 and accompanying text.
72 *See Jenkins*, 2009 WL 3963638, at *2.
debriefing process. Although prisoners must acknowledge and repudiate gang activity to be eligible for debriefing (which is often the only way out of STG classification), the California courts have held that there is no Fifth Amendment issue because the debriefing process cannot be used in subsequent criminal proceedings.

A last possible approach, which may be slightly more effective than Eighth Amendment or Fifth Amendment self-incrimination claims, is a Fourteenth Amendment equal protection challenge. In *Johnson v. California*, the Supreme Court, applying strict scrutiny, found a prison procedure that segregated all new prisoners by race to be unconstitutional. Though STG members are not a protected class, prisons may not use race as a proxy for gang membership. This is somewhat complicated by the reality of prison gangs; six of the largest national prison gangs have members of only one race.

*Turner*, at least as applied to date, seems to offer little hope for prisoners challenging STG designations per se. In the absence of much willingness to find that STG procedures run

---

74 COLUM. HUM. RTS. L. REV., supra note 19, at 15.
76 *Johnson v. California*, 543 U.S. 499, 505 (2005). In a purported attempt to prevent gang violence among new inmates and transferees housed in reception centers, California adopted a de facto policy of complete racial segregation. *Id.* at 502. In striking down the regulation under strict scrutiny, Justice O’Connor cited cases that integrated prison systems. Justice O’Connor noted that, “by insisting that inmates be housed only with other inmates of the same race, it is possible that prison officials will breed further hostility among prisoners and reinforce racial and ethnic divisions.” *Id.* at 507. Justice O’Connor further noted that *Turner* was not applicable to cases involving racial segregation: “The right not to be discriminated against based on one’s race . . . is not a right that need necessarily be compromised for the sake of proper prison administration.” *Id.* at 510. An equal protection challenge asserting that STG apparatuses are using race as a proxy for gang membership could very well be successful under *Johnson*. Such an approach is beyond the scope of this note, however.
77 See United States v. Taveras, 585 F. Supp. 2d 327, 338 (E.D.N.Y. 2008) (holding that government’s use of Dominican heritage as evidence of membership in a Dominican prison gang was impermissible at the sentencing stage of a capital trial); Jimenez v. Cox, No. 3:05-CV-00638-LRH, 2008 WL 4525581, at *8 (D. Nev. Jan. 15, 2008) (denying government’s summary judgment motion where government failed to present evidence showing they were not using race as a proxy for STG membership); see also *Johnson*, 543 U.S. at 511 (“When government officials are permitted to use race as a proxy for gang membership and violence without demonstrating a compelling government interest and proving that their means are narrowly tailored, society as a whole suffers.”).
afoul of the *Turner* formula,\(^79\) and a similar unwillingness to adjust the formula to accommodate more significant prisoner interests,\(^80\) inmates hoping to challenge their deprivations under such systems are at a dead end. While other legal approaches are possible, none offers great possibilities, as evidenced by the paucity of attempts to use them.\(^81\) The remainder of this note will examine the First Amendment interests at stake in the cases of the Five Percenters and the Black Guerilla Family. In turn, this note will suggest some possible approaches to better reflect the actual balance between deprivations and prison interests.

II. **FIRST AMENDMENT**

To understand the unique First Amendment concerns raised by the Five Percenters and Black Guerilla Family, a brief discussion of the groups’ backgrounds is necessary. The ideological, religious (in the case of The Five Percenters), and political (in the case of the BGF) underpinnings of the groups are indispensable to understanding the legal implications of the group members’ status in prison. The background information will be followed by a discussion of the First Amendment implications relevant to each of these groups in the prison setting to better understand how the *Turner* test fails to fully accommodate the interests of these group members in prison.

A. **Five Percenters**

1. Background

Founded in Harlem in the 1960s by Clarence 13X, the Five Percenters (also known as the Nation of Gods and Earths, or NGE) and the closely related Moorish Science Temple of America\(^82\) practice a pseudo-religious, but nontheistic,

\(^79\) *See supra* notes 61-64 and accompanying text.

\(^80\) *See supra* Part I.A.

\(^81\) *See supra* Part I.C.

\(^82\) The Moorish Science Temple of America, founded by Noble Drew Ali in Chicago in 1928, is a precursor to both the Nation of Islam and the NGE. *Moorish American History, Moorish Sci. Temple of Am.*, http://www.moorishsciencetempleofamericainc.com/MoorishHistory.html (last modified Aug. 8, 2008); *see also* KNIGHT, *supra* note 2, at 19-22. It is also a legally recognized religion in some prisons that consider Five Percenters a gang. *Id.* at 167. However, because of their affiliation with street gangs including the Vice Lords, El Rukns, and Black Gangster Disciples, their members are subject to some restrictions in
philosophy. Clarence 13X taught that black people are the original people and the “fathers and mothers of civilization,” and that the black man is God. Five Percener beliefs emphasize the teaching of knowledge via a system of Supreme Mathematics. As would be expected from its membership of occasionally troubled, inner-city young men, the NGE’s connection to violence and crime goes back to its very founding. Shortly after Clarence 13X became Allah, he was shot twice in the group’s headquarters, possibly over a gambling debt. The Five Percenters first entered law enforcement consciousness in the disturbances after the assassination of Malcolm X, when six Five Percenters, including Allah, were arrested for assaulting a police officer. Within a day, the FBI had notified director J. Edgar Hoover of the group’s existence. Contemporary newspaper accounts referred to the Nation as a “terror group” and a “hate group.” Nonetheless, in spite of,

other states. See Morris-Bey v. Debruyn, No. 96-3027, 1997 WL 527857, at *1 (7th Cir. Aug. 21, 1997).

The NGE emphasizes that it is not a racist or anti-white organization, however. DISCLAIMER, 5\% NETWORK, http://www.bazzworks.com/an/disclaimer.html (last visited Jan. 13, 2012). There is even a Milwaukee-based group that asserts that white men can be “Gods,” though this stance is generally rejected by the rest of the Nation. KNIGHT, supra note 2, at 237.

What We Teach, 5\% NETWORK, http://www.allahsnation.net/What.html (last visited Jan. 13, 2012). The Supreme Mathematics is a system of assigning symbolic meaning to the numbers 0 through 9. The NGE also use a system of Supreme Alphabets, which similarly applies a specific meaning to each letter. The systems are used as a basis for the practice of “breaking down,” or reducing words and concepts to their most basic meanings, as provided by the Supreme Mathematics. KNIGHT, supra note 2, at 49-55.

KNIGHT, supra note 2, at 57.

Id. at 70.

Id. at 71. This also marks the beginning of law enforcement’s fundamental misunderstanding of the group; in the FBI teletype prepared and sent to J. Edgar Hoover after the New York disturbance, the group was identified as “the five percent of Muslims who smoke and drink.” Memorandum from FBI on Disturbance by Group Called “Five Percenters” to Dir., FBI 3 (June 6, 1965) (from FOIA Request: Five Percenters, Part 1 of 2), available at http://vault.fbi.gov/5percent/Five%20Percenters%20Part%201%20of%202.


Homer Bigart, Wingate Warns of Negro Revolt if Haryou’s Program Is Curbed, N.Y. TIMES, Oct. 15, 1965, at 1, (from FOIA Request: Five Percenters, Part 1 of 2), available at http://vault.fbi.gov/5percent/Five%20Percenters%20Part%201%20of%202 at 84. This article contains another unique misstatement of the meaning behind the group’s name, citing police reports purportedly quoting group members as saying, “85 per cent of all of the Negroes are like cattle, 10 per cent are Uncle Toms, and we are the 5 per cent who know what belongs to us.” The media would struggle with the meaning of the Five Percenters’ name, even after Mayor John Lindsay began working with the group, with the Times writing in 1967: “The Five Percenters take their name from their contention that only 5 per cent of all Negroes are militant enough to redress
and in fact because of, the earlier attempt on his life, Allah preached a philosophy of nonviolence.90

After Allah’s murder in 1969 (by which time the group’s reputation had improved to the point that Mayor John Lindsay personally expressed his condolences to the Nation),91 the Five Percenter movement seemed on the verge of death as well.92 Two different but connected threads would keep the NGE alive, however. In the 1980s, Allah’s lessons would reach their widest audience yet through the music of New York-based hip hop artists like Rakim, Big Daddy Kane, Brand Nubian, and Poor Righteous Teachers, all of whom were practicing Five Percenters and injected NGE language and philosophy into their music.93 In the 1990s, superstars like Nas, Digable Planets, and especially the Wu-Tang Clan would work Five Percent mythology into their lyrics.94 Years before Raekwon would rap about “today’s mathematics,”95 however, the Nation found a sanctuary of sorts in the New York prison system.96

While incarcerated Five Percenters continued teaching the lessons of a group whose outside existence they were uncertain of, they also aroused the suspicions of prison administrators and the government.97 Five Percenters’ suspected involvement with the 1971 Attica prison uprising did not improve their reputation.98 By the 1990s, as prisons became increasingly concerned about gangs, the Five Percenters came under the spotlight.99

2. Legal Analysis

Today, Michigan, New Jersey, North Carolina, South Carolina, and Virginia classify Five Percenters as an STG.100 In their birthplace, New York, Five Percenters were an STG until

their grievance against what they felt was ill treatment by whites.” KNIGHT, supra note 2, at 97. All these misstatements are similar in making the group sound much more militant than their actual beliefs dictate.
90 KNIGHT, supra note 2, at 118.
91 Id. at 121.
92 Id. at 129.
93 Id. at 179.
94 Swedenburg, supra note 7.
95 “The Sun don’t chill, Allah/What’s today’s mathematics, Sun?/Knowledge, God.” RAEKWON, Knowledge God, on ONLY BUILT 4 CUBAN LINX (RCA Records 1995).
96 KNIGHT, supra note 2, at 161.
97 Id.
98 EDWARD E. CURTIS, IV, ENCYCLOPEDIA OF MUSLIM-AMERICAN HISTORY, VOL. 1, at 203 (2010).
99 KNIGHT, supra note 2, at 165.
Marria v. Broaddus was decided in 2003. In that case, the Southern District wrote, with little precedent, “For these reasons, we find that plaintiff’s beliefs as a member of the Nation of God’s [sic] and Earths are both sincere and ‘religious in nature’ and therefore entitled to RLUIPA [the Religious Land Use and Institutionalized Persons Act] and First Amendment protection under the free exercise clause.”101

Although this line of reasoning, which focuses on sincere adherence to beliefs that are religious in nature, has not been applied by other courts,102 the Western District of Virginia recently expressed some willingness to look more closely at the Five Percenters’ classification, albeit in a different way.103 In a section 1983 challenge stemming from the confiscation of a compact disc with Five Percenter content from the petitioner, the district court denied summary judgment and remanded to the magistrate for further fact finding along both First Amendment and RLUIPA grounds.105 In the absence of specific allegations of gang violence committed by Five Percenters, the court found it impossible to apply the Turner factors.106 Having

---

101 Marria v. Broaddus, No. 97 Civ. 8297 NRB, 2003 WL 21782633, at *12 (S.D.N.Y. July 31, 2003). The Broaddus court also noted that Five Percenter literature specifically instructed inmate adherents to use their time in prison productively, and to follow “the rules of [their] respective prison[s] and reap what [they] sow in this righteousness.” Id. at *3 n.6. This case was a direct challenge on New York’s classification of the Five Percenters as an STG and its concurrent ban on Five Percenter literature and memorabilia. Id. at *4. Unlike other states, New York approached STGs via a “nonrecognition” policy, essentially refusing to acknowledge the gang’s existence to reduce its influence. New York simultaneously restricted any outward displays of membership. Id. Finding that the NGE was a religion as far as the Free Exercise Clause was concerned, and that the inmate-plaintiff was a sincere adherent, the court found that the restrictions lacked any evidentiary support sufficient to meet the standards of the RLUIPA. Id. at *18. The court suggested that the prison really had very little understanding of the Five Percenters:

Here, DOCS proposes to treat exclusively as a gang a group that has had a law-abiding existence outside prison for the better part of 40 years, that is an offshoot of another group that DOCS considers a religion, and that has practices that largely resemble those of recognized religious groups . . . .

Id. at *15. For an in-depth account of the background, proceedings, and aftermath of the Marria case, see Knight, supra note 2, at 167-73.

102 See, e.g., Harbin-Bey v. Rutter, 420 F.3d 571 (6th Cir. 2005) (affirming designation of NGE as a prison gang and STG).


104 Civil Action for Deprivation of Rights, 42 U.S.C. § 1983 (2006) (“Every person who, under color of any statute, ordinance, regulation, custom, or usage, of any State or Territory or the District of Columbia, subjects, or causes to be subjected, any citizen of the United States or other person within the jurisdiction thereof to the deprivation of any rights, privileges, or immunities secured by the Constitution and laws, shall be liable to the party injured in an action at law . . . .”).


106 Id. at *4.
found that the prison could not meet its Turner requirements, the court similarly found that under the “compelling interest” standard of RLUIPA, the prison had not presented enough evidence to meet its summary judgment burden. The court expressed concern that the prison’s justification for the Five Percenters’ classification as an STG was “a general statement that other NGE members have committed actions which have undermined the safety of a prison system.”

Perhaps anticipating the potential extension of STG application to other Muslim groups, the court wrote, “Such generalized factual allegations could arguably be applied to any number of religious groups with a few extremist or violent members.”

The Virginia court, like many courts that have considered Five Percenter challenges, acknowledged that the group resembles a religion enough to warrant analysis as one. Nonetheless, courts have often found that the penological interests at stake are sufficient to overcome the Free Exercise or RLUIPA claims. The leading case, Fraise v. Terhune, illustrates a fundamental problem with applying Turner to the Five Percenters. Judge (now Justice) Alito, holding that the New Jersey prisons demonstrated a valid penological interest in classifying Five Percenters as an STG, wrote, “Here, contrary to the suggestion of the dissent that the New Jersey scheme ‘targets members of one religion,’ the STG Policy is entirely neutral and does not in any way take religion into account.” He is not incorrect; the New Jersey STG system took into account only the group’s:

1. History and purpose;
2. Its organizational structure;
3. The propensity for violence of the group and its members;
4. Actual or planned acts of violence reasonably attributable to the group;
5. Other illegal or prohibited acts reasonably attributable to the group;
6. The demographics of the group, including its size, location, and

---

107 RLUIPA articulates the standard:

No government shall impose a substantial burden on the religious exercise of a person residing in or confined to an institution, as defined in section 1997 of this title, even if the burden results from a rule of general applicability, unless the government demonstrates that imposition of the burden on that person (1) is in furtherance of a compelling governmental interest; and (2) is the least restrictive means of furthering that compelling governmental interest.


109 Id.
110 Id.
111 Id. at *4.
pattern of expansion or decline; and (7) the degree of threat that the
group poses.\textsuperscript{113}

The policy does not consider religion and therefore
neatly passes the \textit{Turner} test’s first prong. The \textit{Turner} test,
however, has no mechanism with which to consider what
happens when an STG is also a religious organization. \textit{Turner}
only accommodates shifting conditions in one direction: a
stronger penological interest makes it easier for the prison to
prevail, but a stronger First Amendment claim, as is present
when an entire religious group is designated a security threat,
does not enter into the calculus.

\textit{Marria v. Broaddus} also considers a claim under
RLUIPA that was not raised in \textit{Fraise v. Terhune}.\textsuperscript{114} Though
the First Amendment analysis of prisoners’ rights claims
requires only that the prison demonstrate a “valid penological
interest,”\textsuperscript{115} RLUIPA requires that any “substantial burden on
the religious exercise of a person residing in or confined to an
institution” be supported by a “compelling governmental
interest,” and that the restriction must be “the least restrictive
means of furthering that compelling governmental interest.”\textsuperscript{116}
Applying RLUIPA, the District Court in \textit{Marria} easily found
the state’s classification of the NGE as an STG and its
consequential deprivations to be impermissible.\textsuperscript{117}

Though it predates RLUIPA,\textsuperscript{118} the Fourth Circuit’s
reasoning from the 1999 case \textit{Mickle v. Moore}\textsuperscript{119} has been
persuasive in many subsequent cases. Unlike the New York
case of \textit{Marria v. Broaddus}, where the judge was highly critical
of the prison’s evidence regarding the dangerousness of the
Five Percenters,\textsuperscript{120} South Carolina presented evidence,
sufficient to satisfy the court, of past events involving violent

\textsuperscript{113} \textit{Id.} at 509 (citations omitted).
\textsuperscript{114} \textit{Id.} at 515 n.5 (noting that no RLUIPA claim was raised).
\textsuperscript{115} \textit{See, e.g.}, Lewis v. Casey, 518 U.S. 343, 402 (1996).
\textsuperscript{117} \textit{Marria v. Broaddus}, No. 97 Civ.8297 NRB, 2003 WL 21782633, at *18
\textsuperscript{118} RLUIPA’s application was arguably limited in 2005 by the Supreme Court’s
decision in \textit{Cutter v. Wilkinson}: “We do not read RLUIPA to elevate accommodation
of religious observances over an institution’s need to maintain order and safety. Our
decisions indicate that an accommodation must be measured so that it does not override
\textsuperscript{119} Mickle v. Moore (\textit{In re} Long Term Admin. Segregation of Inmates
Designated as Five Percenters), 174 F.3d 464 (4th Cir. 1999). Here, Five Percent
inmates challenged South Carolina’s policy of designating their group an STG, and
subsequent administrative segregation. \textit{Id.} at 466.
\textsuperscript{120} \textit{Marria}, 2003 WL 21782633, at *18.
behavior by members of the group.\textsuperscript{121} In an opinion rife with deferential language,\textsuperscript{122} the court found that the prison interests easily met Turner’s standards.\textsuperscript{123} Notably, the Fourth Circuit found that there were other avenues available for the Five Percenters to practice their religion,\textsuperscript{124} a suggestion that was expressly rejected by the Marria v. Broaddus court.\textsuperscript{125} The Fourth Circuit’s decision has been cited approvingly by the First,\textsuperscript{126} Third,\textsuperscript{127} Sixth,\textsuperscript{128} Eighth,\textsuperscript{129} and Tenth\textsuperscript{130} Circuits, while Marria v. Broaddus has only been mentioned in passing by the Seventh Circuit on a tangential point\textsuperscript{131} and was strongly repudiated by the Eastern District of Virginia.\textsuperscript{132} New York’s Five Percenters stand essentially alone, though it is appropriate that the only court to find convincingly in their favor is one housed in their birthplace and spiritual homeland of Manhattan.\textsuperscript{133}

Courts, perhaps unsurprisingly, show little sympathy to claims from Five Percenters; a deliberately non-western ideology\textsuperscript{134} with a body of young, black, male practitioners and strains of virulently anti-white sentiment will inevitably be viewed with more skepticism than a traditional religion. The First Amendment, however, applies equally regardless of the

\begin{flushright}
\textsuperscript{121} In re Long Term Admin. Segregation of Inmates Designated as Five Percenters, 174 F.3d at 466-67.
\textsuperscript{122} Id. at 468-69.
\textsuperscript{123} Id. at 469-70.
\textsuperscript{124} Id. at 470.
\textsuperscript{125} Marria, 2003 WL 21782633, at *13-14. Perhaps it is relevant that the inmates in Marria were denied study materials, while the South Carolina inmates apparently were not, despite the severity of their confinement. In re Long Term Admin. Segregation of Inmates Designated as Five Percenters, 174 F.3d at 470. Or, perhaps the heightened standard of the RLUIPA was dispositive, though the Marria court suggests that perhaps the Five Percenters could also prevail under Turner. Marria, 2003 WL 21782633, at *14 (“Even the less restrictive test set forth in Turner v. Safley that governed prisoner free exercise claims prior to the enactment of [RLUIPA] recognized that deference is not warranted when a prison regulation represents an exaggerated response to security objectives.”).
\textsuperscript{126} Figueroa v. Dinitto, No. 02-1428, 2002 WL 31750158, at *1 (1st Cir. Dec. 9, 2002).
\textsuperscript{127} Fraise v. Terhune, 283 F.3d 506, 517 (3d Cir. 2002).
\textsuperscript{128} Harbin-Bey v. Rutter, 420 F.3d 571, 576 (6th Cir. 2005).
\textsuperscript{129} Goff v. Graves, 362 F.3d 543, 551 n.6 (8th Cir. 2004).
\textsuperscript{130} Ajaj v. United States, No. 07-1073, 2008 WL 4192738, at *13 n.3 (10th Cir. Sept. 15, 2008).
\textsuperscript{131} Lindell v. McCallum, 352 F.3d 1107, 1110 (7th Cir. 2003).
\textsuperscript{132} Harrison v. Watts, 609 F. Supp. 2d 561, 573-74 (E.D. Va. 2009) (“Only a single district court has held that a Five Percenter’s beliefs are religious in nature and therefore deserving of protection under the First Amendment and the RLUIPA, and its conclusion in this respect is neither controlling nor persuasive.”).
\textsuperscript{133} See supra Part II.A.1.
\textsuperscript{134} Swedenburg, supra note 7.
\end{flushright}
popularity of the religion in question. While Five Percenters are not Muslims in any traditional understanding of Islam, they do adopt some of the trappings and practices of Muslims. Muslims are disproportionately represented in American prisons, and conversion to Islam is extraordinarily popular amongst inmates. Combined with the ongoing war on terror, the Muslim prison population can be expected to grow and draw intense scrutiny. Given the current public and political hostility to Islam in many segments of the United States, it is foreseeable that other Muslim groups may find themselves subject to treatment similar to the Five Percenters. The United Kingdom is already experiencing problems with Muslim prison gangs.

B. **Black Guerilla Family**

1. **Background**

The Black Guerilla Family was founded in prison by George Jackson. Born in Chicago in 1941, Jackson first encountered racism in elementary school, when he was sent to a segregated inner-city Catholic School. He began dabbling in petty crime as a teenager, when his family moved to the troubled Troop Street housing project. He moved on to more serious lawbreaking after moving to Los Angeles, where he first spent time in prison for allegedly breaking into a department store and attacking a police officer. In 1960, Jackson was arrested for robbing a gas station and sentenced to one year to life; while serving his indeterminate sentence,

---

135 “For the Constitution protects expression and association without regard to the race, creed, or political or religious affiliation of the members of the group which invokes its shield, or to the truth, popularity, or social utility of the ideas and beliefs which are offered.” NAACP v. Button, 371 U.S. 415, 444-45 (1963).
136 Swedenburg, supra note 7.
139 See generally Zoll, supra note 137.
142 GEORGE JACKSON, SOLEDAD BROTHER 6-7 (1994).
143 Id. at 12.
144 Id. at 16.
he was repeatedly denied parole for disciplinary infractions. He was accused of killing a prison guard in Soledad Prison in 1970, a conviction that would have resulted in a mandatory capital sentence, but was killed in an apparent escape attempt at San Quentin prison in August 1971, though the details of his death were controversial.

While in prison, Jackson became interested in Marxist theory, and, over his decade in prison, combined Mao with Huey Newton to develop his own brand of African-American revolutionary ideology. During this period, he also organized the BGF with fellow revolutionaries W.L. Nolen and David Johnson. The group was formed with two primary purposes: The first, immediate goal was to improve the conditions of incarceration for African-Americans by advocating their right to self-defense. The second was to advocate the revolutionary overthrow of the racist American establishment, through violent means if necessary.

Although Jackson was not shy about violence and argued for an armed revolution, the original BGF, as founded by Jackson and Nolen, was indisputably a political organization. Jackson, who, spending much of his time in solitary confinement, was a voracious reader, absorbed the teachings of revolutionaries ranging from Marx and Lenin to Mao, Che

---

145 IMPRISONED INTELLECTUALS, supra note 8, at 84.
146 JACKSON, supra note 142, at 16.
147 Colin Nickerson, Revising a Mystery of the Radical Years, BOS. GLOBE, Aug. 12, 1984, at 1.
149 IMPRISONED INTELLECTUALS, supra note 9, at 85.
150 Id. Although self-defense through violent means was likely not outside the scope of the BGF's aims, it was not the only means advocated. Nolen was in the process of filing civil rights suits on behalf of Soledad inmates when he was killed by a guard during a prison fight. Id. Jackson himself corresponded with Huey Newton's lawyer, Fay Stender, id., in relation to the charges of killing the prison guard. JACKSON, supra note 142, at 206.
151 The inscription to Jackson's first book, Blood in My Eye, reads:

We must accept the eventualty of bringing the U.S.A. to its knees; accept the closing off of critical sections of the city with barbed wire, armored pig carriers crisscrossing the streets, soldiers everywhere, tommy guns pointed at stomach level, smoke curling black against the daylight sky, the smell of cordite, house-to-house searches, doors being kicked in, the commonness of death.

JACKSON, supra note 148, at 1.
152 “People's war, class struggle, war of liberation means armed struggle.”
JACKSON, supra note 142, at 226.
153 JACKSON, supra note 142, at 14.
154 Id.
Guevara, and Huey Newton, as well as more obscure figures like Frantz Fanon and John Gerassi. He was also a student of history, and history to him foreclosed on the usefulness of nonviolent protest. Jackson’s ideas were extreme, and he acknowledged as much, but the BGF’s ideology was roughly in line with other revolutionary groups of the 1960s.

After Jackson’s death, however, the BGF’s character shifted. BGF members were charged with the murder of Jackson’s philosophical mentor Huey Newton in 1989. In 1996, the BGF was named on a list of notorious American gangs by the Justice Department. More recently, a group of two dozen Baltimore residents calling themselves BGF members were indicted on federal conspiracy charges for drug trafficking.

There is some confusion as to the nature of the current incarnation of the BGF. Today, the courts generally treat it as an undisputed prison gang. The group was identified by the Department of Justice as a prison gang in 1996, and the Florida Department of Corrections identifies the BGF as one of

---

155 Id. at 25.
156 Id. at 11.
157 Id. at 25.
158 Id. at 8.
159 JACKSON, supra note 142, at 223.
160 “I am an extremist. I call for extreme measures to solve extreme problems.”
161 Id. at 265.
162 Jackson’s beliefs were similar to many of the ideas espoused by the more radical elements of the New Left movement of the 1960s; for a wide sampling of this ideology, see 15 Years of Radical America: An Anthology, 16 RADICAL AM., no. 3, May-June 1982, available at http://dl.lib.brown.edu/pdfs/1142526171657220.pdf, especially World Revolution: The Way Out, id. at 125. The full archive of Radical America, the leading journal of the New Left movement, is available at http://dl.lib.brown.edu/radicalamerica.
166 See, e.g., Harrison v. Milligan, No. C 09-4665 SI, 2010 WL 1957389, at *1 (N.D. Cal. May 14, 2010) (“[Correction officer’s] description of the materials seized: ‘Upon further searching I noticed two (2) white manila envelopes containing material on George L. Jackson, founder of the BGF (Black Guerrilla Family) and material on the New Afrikan Nationalist movement (the term adopted by the BGF to disguise their gang activity in prison and on the streets).”). It is worth noting that the New Afrikan movement seems to have plenty of life outside the context of prison gangs, see, e.g., PROUDFLESH: NEW AFRIKAN JOURNAL OF CULTURE, POLITICS AND CONSCIOUSNESS, http://www.proudfleshjournal.com/, and within the prison context appears legitimate and not necessarily linked to the BGF, though the ideology is similar to Jackson’s. See Sundiata Acoli, A Brief History of the New Afrikan Prison Struggle, GLOBALAFRICA.COM, http://www.globalafrica.com/Sundiata.htm (last visited Jan. 13, 2012).
167 Id. at 25.
168 Neikirk, supra note 163, at 3.
the six major national prison gangs.\footnote{Gang and Security Threat Group Awareness: Major Prison Gangs, FLA. DEP’T OF CORR., http://www.dc.state.fl.us/pub/gangs/prison.html (last visited Nov. 3, 2010). Similarly, California lists the BGF as one of seven enumerated prison gangs in its Department of Corrections Operations Manual. CAL. DEP’T OF CORR. & REHAB., OPERATIONS MANUAL § 52070.17.2 (2009), available at http://www.cdcr.ca.gov/Regulations/Adult_Operations/docs/DOM/DOM%20Ch%205-Printed%20Final.pdf.} Discussion on African-American revolutionary Internet message boards, however, reveals competing theories about the nature of the BGF today. One prevailing understanding is that the group is a unified prison organization of black inmates who drop their street gang affiliations for protection in numbers, which is not far from Jackson’s original intention.\footnote{“BGF isnt [sic] a gang—its bigger than that, u [sic] got BGF members who are lawyers and doctors, its main purpose is protection of your black brothers (and sisters).” ASSATA SHAKUR SPEAKS FORUMS: BLACK GUERRILLA FAMILY, http://www.assatashakur.org/forum/share-comrades/13092-black-guerilla-family.html (last visited Feb. 3, 2012).} Other posters speculate that the group is no more than a criminal gang, but uses revolutionary ideology to hide their criminal acts and create an appearance of legitimacy, which is also the government’s interpretation.\footnote{Id.} There is also debate as to whether the street version of the BGF is the same organization as the prison version.\footnote{Id. The Latin Kings and Gangster Disciples, two other major prison gangs, have also taken steps to be perceived as legitimate groups, and, in the case of the Gangster Disciples, have created an out-of-prison political organization. KNIGHT, supra note 2, at 166-67.}

2. Legal Analysis

Since reliable information on the group is difficult to come by, and courts are not in the habit of reading revolutionary message boards, BGF members have unsurprisingly had little success challenging STG designation on the grounds that BGF is not a gang. While some, maybe even most, self-identified BGF inmates are involved in the types of violent criminal activities that characterize a traditional prison gang, it is also entirely likely that some are involved out of solidarity with George Jackson’s ideas. Since possession of Jackson’s books can be evidence of STG membership,\footnote{See, e.g., Harrison v. Milligan, No. C 09-4665 SI, 2010 WL 1957389, at *1 (N.D. Cal. May 14, 2010).} it would seem that even an interest only in BGF ideology would be sufficient to warrant STG classification under the \textit{Turner} standard.
This becomes problematic when one considers that political speech is historically one of the most highly valued types of speech.\textsuperscript{172} Even speech advocating the overthrow of the United States government is protected, and cannot be restricted unless the speech is aimed at “inciting or producing imminent lawless action and is likely to incite or produce such action.”\textsuperscript{173} The nature of incarceration would seem to make it impossible for a prisoner to incite imminent lawless action.

Few courts have addressed the speech rights of prisoners. In \textit{Rios v. Lane}, though it predates the development of the STG concept, the Seventh Circuit found that a prison violated an inmate’s due process rights when he was punished for circulating a list of Spanish-language radio stations, along with socialist slogans.\textsuperscript{174} Though he was charged with violating a rule restricting gang activity, the court found that the regulation was impermissibly vague to justify the infringement on Rios’s rights.\textsuperscript{175} Other cases involving political speech by prisoners tend to arise in the context of retaliation claims.\textsuperscript{176} It is illegal for prisons to punish inmates for exercising their constitutional rights in the absence of a valid penological goal,\textsuperscript{177} though the prison is entitled to “appropriate deference.”\textsuperscript{178} There is little other jurisprudence focusing on the political speech rights of prisoners; whether this is because prisoners are no longer particularly political, or because the prisons are generally respecting their speech rights, is unclear.

Prisons’ ability to restrict the political expression of BGF members is sharply illustrated by cases in which the prison used the writings of George Jackson as proof of gang membership. The California Court of Appeals recently held that the prison’s use of George Jackson’s books as evidence of gang membership did not violate an inmate’s First Amendment

\textsuperscript{172} “[P]olitical speech [is] at the core of what the First Amendment is designed to protect.” Virginia v. Black, 538 U.S. 343, 365 (2003).

\textsuperscript{173} Brandenburg v. Ohio, 395 U.S. 444, 447 (1969). The Court’s recent decision in \textit{Holder v. Humanitarian Law Project} arguably limited (or ignored) the \textit{Brandenburg} standard, 130 S. Ct. 2705, 2733 (2010) (Breyer, J., dissenting), though for now, at least, the holding is limited to cases involving monetary support to foreign terrorist organizations, \textit{id.} at 2711-12 (majority opinion).

\textsuperscript{174} Rios v. Lane, 812 F.2d 1032, 1034 (7th Cir. 1987). Note also that this case predates, albeit by a matter of months, the Supreme Court’s decision in \textit{Turner}. It is unclear if the decision would be the same under \textit{Turner}’s due process standard.

\textsuperscript{175} \textit{id.} at 1038-39. The court emphasized, however, that Rios’s free speech rights were not implicated. \textit{id.} at 1039.


\textsuperscript{177} Rhodes v. Robinson, 408 F.3d 559, 567-68 (9th Cir. 2005).

Based on “gang expert” testimony, the first prong of *Turner* was easily met, as BGF members routinely possess Jackson’s writings, and, as they are a recognized prison gang, there is a prison interest in preventing prisoners from indicating their gang affiliation via possession of Jackson’s work. The second prong, the availability of an alternative means to exercise the restricted right, required more work by the California court. By interpreting the liberty interest as “the right to receive and read outside publications,” the court found the prong met. Framed that way, the restriction on Furnace’s ability to read the works of George Jackson did not foreclose his ability to read other things, and he therefore had alternate means to exercise his right.

This view of the inmate’s First Amendment interests only makes sense once the court concludes that the BGF is necessarily a prison gang. Here, the evidence presented as to the nature of the group came from a corrections official. The court referenced “the violent courtroom escape of three California prisoners from a Marin County courtroom in 1970” as the work of the BGF, as well as Jackson’s failed escape attempt in 1971 that “resulted in the killing of three California correctional officers.” The court did not, at any point, consider the political nature of Jackson’s writings or of the BGF as an organization.

The degree of deference inherent to the *Turner* test would likely still make it extremely difficult for the prisoner to prevail, even if the prisoner claimed his rights to political expression were being unconstitutionally restricted. *In re Furnace* illustrates courts’ willingness to defer entirely to the judgment of the prison regarding the nature of groups.

---

179 *In re Furnace*, 110 Cal. Rptr. 3d 820, 833 (Ct. App. 2010).
180 *Id.* at 831.
181 *Id.* at 832.
182 *Id.*
183 *Id.* at 827-28.
184 *Id.* at 827. The court does not mention that Jackson’s younger brother, Jonathan, was killed by police during the incident. *Jackson*, supra note 142, at xiii.
185 *In re Furnace*, 110 Cal. Rptr. 3d at 827. The court similarly failed to note that Jackson himself was killed in the attempt and did not acknowledge the controversy regarding the escape. See *supra* note 147.
186 A federal court in California recently found that prison officials went too far when they suppressed an inmate’s outgoing mail for containing references to “Black August,” a sort of prison memorial recognizing the deaths of Jackson and Nolen, and various “New African Nationalist” organizations, finding an insufficient link to the BGF. *Harrison v. Inst. Gang of Investigations*, No. C 07-3824 SI, 2010 WL 653137, at *6-7 (N.D. Cal. Feb. 22, 2010). Though prisoner mail was analyzed under the slightly tougher standard of *Procunier v. Martinez*, *id.* at *4, it would still appear (at least in California) that the line is drawn at possession of Jackson’s actual books.
designated STGs, without any independent fact-finding. Nonetheless, if a court starts its analysis with the presumption that the BGF is a political organization instead of a criminal gang per se, it becomes more difficult for the court to find that the prisoner retains alternative means of practicing his rights.

Like the religious rights of the Five Percenters, suppression of the political rights of the BGF should require a much greater showing than Turner demands. Far from suppressing the right of a gang member to associate with his fellow gang members and draw gang signs on his walls, the courts in these cases are all too willing to forcefully subdue religious and political affiliation amongst inmates. The application of Turner in these cases is not in error, but the very formulation of the test does not allow courts to consider the significance of the rights being invoked. A peaceful practitioner of the Divine Mathematics or a prisoner with a scholarly interest in George Jackson may be treated no differently from a member of a violent criminal gang. If the liberty interest at stake is the expression of the prisoner’s political beliefs and the writings of George Jackson are foundational to his political beliefs, banning a prisoner from reading any of Jackson’s books makes it extremely difficult for him to exercise his rights to political expression. Prisons understandably do not want inmates possessing written material that is used solely for the purpose of signifying their gang membership; but it is something else entirely to say that the prison interest in safety and gang control can justify suppression of the written materials essential to a prisoner’s political beliefs. But because the Turner test has no mechanism for the court to weigh the importance of the liberty interest against the weight of the penological interest, prisons are able to suppress political expression quite easily.

III. CONSEQUENCES OF STG DESIGNATION

Much as Turner does not allow for consideration of heightened constitutional rights, it also fails to take into full account the severity of the deprivations inherent in the Security Threat Group context. Only the fourth prong of the test, the “absence of ready alternatives” for the prison, hints at the possibility of weighing the competing interests. This

---

187 In re Furnace, 110 Cal. Rptr. 3d. at 827-28.
prong of the test is a reformulation of an earlier due process test, *Bell v. Wolfish*, which requires a showing that the prison response to the alleged penological interest was not exaggerated.\footnote{Bell v. Wolfish, 441 U.S. 520, 551 (1979).} Though the Court in *Turner* addressed the “exaggerated response” idea, it expressed the idea in terms of “obvious, easy alternatives” that the prison could use to meet its penological goal, which the Court noted would be evidence of an exaggerated response.\footnote{Turner, 482 U.S. at 90.} This effectively shifts the burden; rather than requiring the prison to show that its response was proportional to the penological risks, the inmate must show that the prison had “obvious, easy alternatives” that would accomplish the same goal. The dissenters in *Turner* noted the relative meaninglessness of the Court’s interpretation of the fourth prong, which led to the Court’s upholding a sweeping ban on inmate correspondence while striking down a ban on inmate marriage. Justice Stevens wrote:

> The marriage rule is said to sweep too broadly because it is more restrictive than the routine practices at other Missouri correctional institutions, but the mail rule at Renz is not an “exaggerated response” even though it is more restrictive than practices in the remainder of the State . . . . Unfathomably, while rejecting the Superintendent’s concerns about love triangles as an insufficient and invalid basis for the marriage regulation, the Court apparently accepts the same concerns as a valid basis for the mail regulation.\footnote{Id. at 113-14 (Stevens, J., concurring in part and dissenting in part).}

In the years since *Turner* was decided, the fourth prong has not been examined in much detail by the courts and has been interpreted broadly in favor of the prison, particularly in the context of STGs.\footnote{See, e.g., Holley v. Johnson, No. 7:08CV00629, 2010 WL 2640328, at *4 (W.D. Va. June 30, 2010) (finding the third and fourth prongs of *Turner* met by prison’s zero-tolerance gang policy as applied to a Five Percenter, despite inmate’s argument that the prison’s interest could be met by only banning Five Percenter literature that was likely to incite violence).} It is thus useful to consider what sorts of

---

\[189\] Bell v. Wolfish, 441 U.S. 520, 551 (1979). The *Bell* plaintiffs challenged various policies in a short-term, pretrial confinement center in New York, id. at 524, alleging violations of their constitutional rights “because of overcrowded conditions, undue length of confinement, improper searches, inadequate recreational, educational, and employment opportunities, insufficient staff, and objectionable restrictions on the purchase and receipt of personal items and books,” id. at 527. Rejecting the Second Circuit’s “compelling necessity” test for deprivations of detainee rights, id. at 531-32, the Supreme Court held that pretrial conditions must not amount to punishment, a standard that is satisfied “if a restriction or condition is . . . reasonably related to a legitimate goal.” *Id.* at 539. *Turner* effectively approves *Bell v. Wolfish* and emphasizes that the “least-restrictive means” test suggested in the earlier case *Procunier v. Martinez* was not to apply to due process claims brought by inmates. *See* Beerheide v. Suthers, 286 F.3d 1179, 1184 (10th Cir. 2002).
restrictions have been found “unexaggerated” in the context of STG management systems.

A. Lost Privileges

The Supreme Court has upheld the deprivation of reading material and visitation rights from security risk prisoners.\textsuperscript{193} By broadly framing the rights in question, courts have been able to find that a restriction on some kinds of reading material is legitimate, so long as the prisoner has access to other reading materials.\textsuperscript{194} The loose framing of inmate rights, along with the concurrent broad sweep of banned materials, has not gone unchallenged, however. A group of plaintiffs, representing inmates and magazine publishers, along with the ACLU of Colorado, brought suit in Colorado to challenge the state policy of censoring publications that STGs purportedly distribute or that advocate joining STGs.\textsuperscript{195} Gang expert and plaintiffs’ witness Alex Alonso submitted an affidavit in the case, in which he explained that the censorship regulation was “overly broad, subjective and vague,” and stated, “[T]he Colorado [Department Of Corrections] is confused over what is a gang sign and general hand signs used by non-gang members.”\textsuperscript{196} Whether because of prison officials’ ignorance of African-American culture, or simply because of the traditional deference to prison decisions, the sweep of banned materials was particularly broad; among the publications censored were \textit{VIBE}, \textit{Rolling Stone}, and the \textit{Source}, and the alleged “gang-affiliated” personalities in the publications—apparently based on hand signs—included Shaquille O’Neal, Chris Rock, and Derek Jeter.\textsuperscript{197}

Books by or about George Jackson have been used as evidence of gang membership\textsuperscript{198} and are often prohibited in prisons, as is the \textit{Five Percenter}, a magazine published by the


\textsuperscript{194} In re Furnace, 110 Cal. Rptr. 3d. 820, 832 (Ct. App. 2010).


\textsuperscript{197} Id. at 4.

\textsuperscript{198} See James Ridgeway & Jean Casella, Prisoner Sent to Solitary Based on Reading Materials, SOLITARY WATCH (June 16, 2010), http://solitarywatch.com/2010/06/16/prisoner-locked-up-in-solitary-based-on-reading-materials/.
Nation of Gods and Earths organization. A California Prison Focus, a newsletter put out by the non-profit group of the same name, reports that an issue of their newsletter mentioning George Jackson and “Black August,” an annual event memorializing the deaths of Jackson and other black prison leaders, was used as evidence of STG membership.

This sort of systematic censorship of reading materials, especially in light of evidence that prison officials do not always understand the material they are censoring, speaks to both the severity of STG designation and the invasion of First Amendment rights suffered by prisoners. Peaceful Five Percenters and George Jackson followers (or even simply inmates with an interest in the writings of either group) can suffer, at the very least, the inability to study the groups’ teachings and, in some states, may lose considerably more significant privileges, including the bare amount of personal freedom possible in prison.

B. Segregation

A common consequence of STG designation is limited or no contact with other group members or, in some states, placement in segregated housing or solitary confinement. Prisoners designated as STG members are often not allowed to associate with other members of their group, at least to discuss the group or partake in activities related to the group. Particularly for the Five Percenters, this restriction can be extremely limiting. A major element of NGE study, based on the early days when Allah and his first followers pored over the Lost-Found Lessons together, is the act of studying together. Five Percenters form a “cipher,” or a group, usually in a circle, where followers quiz each other on their beliefs, participate in intense one-on-one conversations known as “building,” and

---

201 See infra Part III.B.
202 See, e.g., Ciempa v. Jones, 746 F. Supp. 2d 1171, 1183 (N.D. Okla. 2010) (discussing prison refusal to allow group meetings of Five Percenters); Fraise v. Terhune, 283 F.3d 506, 519 (3d Cir. 2002) (noting that New Jersey correctional policy prohibits “participating in any activity(ies) related to a security threat group”),
203 KNIGHT, supra note 2, at 49, 59-60.
205 KNIGHT, supra note 2, at 81.
hold conventions known as “parliaments.”\textsuperscript{206} Five Percenters can barely be Five Percenters without the ability to work together.\textsuperscript{207}

STG designation can also, in some states, result in placement in solitary confinement.\textsuperscript{208} The use of solitary confinement in the United States has attracted a large amount of legal and political attention in recent years.\textsuperscript{209} In March 2009, Atul Gawande wrote an article for the \textit{New Yorker} with the subtitle, “The United States holds tens of thousands of inmates in long-term solitary confinement. Is this torture?”\textsuperscript{210} \textit{Wired} magazine answered that question a month later in an interview with University of California, Santa Cruz psychologist Craig Haney, who said,

I don’t think correctional administrators always put people in solitary confinement just to make them feel pain. But to the extent that’s done, to the extent they know that people in these environments will feel that pain, then that creeps very close to the definition of what’s understood internationally as torture. I think our sloppiness, our carelessness about how this policy has been implemented, raises very severe ethical concerns about the humane treatment of prisoners by both U.S. standards and international standards.\textsuperscript{211}

A fifty-two-year-old segregated housing unit inmate in the California prison system who has spent seventeen years in solitary confinement as a result of gang validation, writes, “If this is not torture, I don’t know what is. I have seen many fall victim to isolation and sensory deprivation of the SHU environment.”\textsuperscript{212} In 2011, California prisoners in segregated housing, many of them STG members, engaged in a series of hunger strikes to protest the conditions of their confinement.\textsuperscript{213}

\begin{footnotesize}
\begin{enumerate}
\item Id. at 89.
\item See id. at 5 (“Gods love to share knowledge and speak for their nation.”); id. at 131 (“Like the father said, all they needed to do was come together.”).
\item See supra note 19 and accompanying text.
\item See supra note 32.
\end{enumerate}
\end{footnotesize}
Commentators have questioned the constitutionality of long-term segregation. Professor Jules Lobel writes, “The federal courts have not yet definitively addressed the question of whether confining a prisoner permanently or for very long periods of time in a supermax prison, without meaningful periodic review of his or her behavior, constitutes cruel and unusual punishment.” A prison’s penological goals may indeed be met through an STG system, but the question remains whether indefinite solitary confinement or cutting prisoners off from other practitioners of their beliefs are proportional responses. The courts have not yet adequately addressed this issue, though language in *Turner* suggests they could.

C. Reporting to Local Law Enforcement on Release

Security Threat Group designation does not entirely end with the end of incarceration. Even upon release, STG classification follows the person; in some states, local law enforcement must be notified when a gang member is released from prison. Given the difficulty of challenging STG designation, the long-ranging consequences, sticking with the prisoner even after he has served his sentence, are significant. Reporting sincere adherents of BGF or Five Percenter beliefs to local law enforcement implicates the suppression of constitutionally protected speech outside the prison walls. Being a member of the Five Percent Nation is, of course, not illegal.

D. “Debriefing”

In California, and likely other states, the only way to have your STG classification removed, aside from proving you have been inactive in the gang for six years, is to complete a

---


215 *Turner v. Safley*, 482 U.S. 78, 97-98 (1987). (“No doubt legitimate security concerns may require placing reasonable restrictions upon an inmate’s right to marry, and may justify requiring approval of the superintendent. The Missouri regulation, however, represents an exaggerated response to such security objectives.”).


Debriefing involves renouncing the gang, providing information about all gang activity to the prison, and informing the prison of the identity of other, unvalidated gang members. The debriefing process has been challenged, albeit unsuccessfully, on Eighth Amendment grounds. In Castañeda v. Marshall, an inmate argued that reporting the names of other gang members would put his safety at risk from retaliation by his (former) gang. The court rejected this reasoning, deferring to the prison’s assurance that they take steps to protect debriefed former gang members from retaliation. The effectiveness of prison protection for former gang members is debatable; although California provides a special unit in Pelican Bay Prison (the home of its Secure Housing Unit for STG members) for former gang members, prison advocacy groups are unconvincing that the protection is sufficient. The prison also fails to provide any protection once the inmate leaves prison and returns to civilian society, where his former gang may target him on the presumption that he informed on the gang and other members.

Other prisoners have challenged the debriefing process on Fifth Amendment privilege-from-self-incrimination grounds. Courts have found these claims unpersuasive, as debriefing evidence cannot be used in later criminal proceedings. Regardless, in light of the cultural disapproval of “snitches” in the communities that gang members tend to come from, given the choice between staying in secure housing and debriefing, the large majority of inmates choose the former.

The STG system denies prisoners the ability to read materials essential to their religious or political beliefs, and

---

221 Id.
222 Id.
224 Id.
226 Id. at *19.
228 Montgomery, supra note 223.
229 See supra Part III.A.
subjects them to the harshest form of punishment available in American prisons. It follows them after they have served their time, and inmates can only escape it within prison by endangering their own safety. Such a system deserves much stronger checks from the court than are currently in place. When coupled with the significance of the rights being claimed by Five Percenters and peaceful BGF adherents, the Turner test and the Supreme Court’s emphasis on deference runs into a perfect storm of inefficacy.

IV. PROPOSALS FOR CHANGE

A. The Future of Turner

The ability of a prison to impose such severe restrictions on inmates for exercising their rights to political and religious expression, with only minimal justification, raises serious constitutional concerns. As the American prison population evolves, Turner’s application to the Five Percenters and BGF sets a dangerous precedent for future inmates affiliated with religious or political groups deemed to be dangerous. It is fairly clear that, under current due process requirements, it is nearly impossible for a prisoner to present a strong enough liberty interest to overcome the presumption in favor of the prison staff laid out in Turner. In the few cases where a prisoner has won a case involving his STG designation, the decisions have been based on the prison’s failure to meet either the limited standard required by the Supreme Court or evidentiary flaws in the validation process. Marria v. Broaddus overturned New York’s STG system as applied to Five Percenters via RLUIPA, but the Supreme Court’s later restrictions of the interpretation of the Act seem to have foreclosed that line of reasoning.

The Turner test is not likely going anywhere. Turner was merely the first in an over-two-decade-long line of Supreme Court decisions that have restricted the ability of

230 See supra Part III.B.
231 See supra Part III.C.
232 See supra Part III.D.
233 See generally supra Part II.
234 See supra Part I.A.
236 See supra note 18.
prisoners to remedy violations of their constitutional rights.\textsuperscript{237} The Supreme Court, moreover, has been resistant to a facial challenge to \textit{Turner}, with little dissent. While \textit{Beard v. Banks}, a case involving indefinite denial of access to newspapers and magazines, drew two dissenters (Stevens and Ginsburg),\textsuperscript{238} \textit{Overton}'s strict limitations on visitation rights found all nine justices upholding the restriction.\textsuperscript{239} Interestingly, four justices (Stevens, Brennan, Marshall, and Blackmun) rejected the \textit{Turner} formula at the time of the decision, with Justice Stevens writing that the requirement that prison officials find a "logical connection" between the restriction and the penological goal was "virtually meaningless" and "would seem to permit disregard for inmates' constitutional rights whenever the imagination of the warden produces a plausible security concern . . . ."\textsuperscript{240} They would appear to be mostly right, but the courts since then have not seemed to mind.

Although courts have found for the prisoner in some cases, the standard is deferential enough that they could just as easily have found for the prison—courts on the whole err towards deference.\textsuperscript{241} The end effect is to give prison staff effectively unfettered discretion in the STG area. The desire of the courts to take this approach is understandable; prison gang members are not particularly sympathetic figures, and the body of research on the subject seems to indicate that segregating gang members can indeed lead to a reduction in prison violence.\textsuperscript{242} Gang affiliation has been shown to increase the incidence of nearly all types of prison misconduct.\textsuperscript{243} Gang researcher George Knox writes:

\begin{itemize}
  \item \textsuperscript{238} Beard v. Banks, 548 U.S. 521, 542 (2006) (Stevens, J., dissenting).
  \item \textsuperscript{239} Thomas and Scalia refused to join the majority opinion because they believed that only Eighth Amendment claims are judiciable by the Supreme Court over state law. Overton v. Bazzetta, 539 U.S. 126, 141 (2003) (Thomas, J., concurring in the judgment).
  \item \textsuperscript{240} Turner v. Safley, 482 U.S. 78, 100-01 (1987) (Stevens, J., dissenting).
  \item \textsuperscript{241} See supra notes 50-52 and accompanying text.
  \item \textsuperscript{242} See NJDOC's Gang Management Unit a National Model, N.J. DEP'T OF CORR. (Aug. 3, 2005), http://www.state.nj.us/cgi-bin/ corrections/ njnewsline/view_article.pl?id=2661 ("[T]here has been a department-wide drop of 42 percent in staff assaults and an 84 percent decrease in organized violent behavior among NJDOC inmates."). Ironically, the apparently highly regarded New Jersey gang unit was dismantled in early 2010, a victim of budget constraints. Kibret Markos, \textit{Closing Gang Unit to Save State $5M}, RECORD (Woodland Park, NJ) (May 7, 2010), http://www.northjersey.com/news/93049129_Closing_gang_unit_to_save_state__5M.html.
Some academic authors who read and write about prison gangs without doing empirical research on the issue are prone to use the prison gang problem as a platform to criticize the status quo. A common theme in this “gang apologist” approach is to begin with a 1960’s concept of prison rehabilitation and how wonderful the world would be if there were more services and a higher quality of life for inmates and better jobs upon release from prison, and then use the prison gang issue as just another topic to criticize the prison system. These are approaches that ignore the fact that correctional staff are good citizens working for a living and often are the ones brutally assaulted by prison gangs or STGs—this kind of information is not on the minds of the academic critic. 244

Without agreeing with Mr. Knox’s opinion of academic researchers, this note is not intended to argue that the STG system is inherently wrong or that it does not serve important penological interests. Moreover, the nature of imprisonment, as has been pointed out many times, necessarily includes the restriction of constitutional rights. 245 Under the current Turner standard, finding a penological interest that justifies restricting a prisoner’s rights of association and expression is not a difficult task for the prison. The issue therefore becomes the adequacy of the Turner test in the face of the strong interests presented by religious and political groups, like the Five Percenters and the BGF.

The Turner test has two significant shortcomings: first, it does not, as generally applied, take into account the severity of the deprivation of rights; and second, as becomes important in cases involving STGs that have a religious or political nature, it contains no mechanism for adjusting its analysis in response to the significance of the right being restricted. The Five Percenters and BGF create unique cases, where the speech being restricted is religious or political (two varieties of speech classically subject to very strong protection), and the

244 Knox, supra note 243.
245 See, e.g., Beard v. Banks, 548 U.S. 521, 528-29 (2006) (“This Court recognized in Turner that imprisonment does not automatically deprive a prisoner of certain important constitutional protections, including those of the First Amendment. But at the same time the Constitution sometimes permits greater restriction of such rights in a prison than it would allow elsewhere.” (citations omitted)).
246 Though the Supreme Court has never explicitly endorsed a hierarchy of speech, as Justice Stevens wrote in R.A.V v. City of St. Paul, “Our First Amendment
deprivations as a result of speech can reach the point of indefinite seclusion.\textsuperscript{247}

The other element that \textit{Turner} does not, but should, allow into the courts’ calculus is the severity of deprivations faced by inmates subjected to the STG system, particularly solitary confinement. Decisions which casually describe long-term, or even indeterminate, isolation as “administrative segregation,” and thus subject to minimal review, ignore the terrible consequences of this type of punishment, which multiple commentators have suggested should not be allowed in any circumstances.\textsuperscript{248}

\textbf{B. Proposed Approaches}

Given the intractability of the \textit{Turner} test, and the unlikeliness that the Supreme Court will add a new prong to the formula in the near future, how can the rights of inmates be adequately protected under existing due process standards? One simple solution has been proposed by Scott N. Tachiki—the prison must show individualized proof of violation of prison rules \textit{aside from} gang membership as a predicate to STG assignment.\textsuperscript{249} This change would take care of the issue of the “ideological” gang member who, while interested in the philosophical underpinnings of a group, is not involved with its criminal activity. A policy change of this sort, however, would likely have to come from the prison system itself, and the larger fear in the prison system today is that gangs will somehow slip by unnoticed, not that innocent inmates will be swept into the STG process.\textsuperscript{250}

Judge Marjorie Rendell, dissenting in \textit{Fraise v. Terhune}, proposed another means of dealing with “case[s] placing harsh

\textsuperscript{247} See supra Part III.B.

\textsuperscript{248} For a thorough discussion on the impact of solitary confinement and the courts’ inability to deal with it properly, see Christine Rebman, Comment, \textit{The Eighth Amendment and Solitary Confinement: The Gap in Protection from Psychological Consequences}, 49 DePaul L. Rev. 567 (1999).

\textsuperscript{249} Tachiki, supra note 21, at 1145-46.

\textsuperscript{250} See, e.g., \textit{Gang and Security Threat Group Awareness}, Fla. DEP’T OF CORR., http://www.dc.state.fl.us/pub/gangs/prison.html (last visited Feb. 3, 2012) (“Each group is represented in Florida’s prison system population; however some are not readily recognizable . . . . Although their numbers are small in Florida prisons, if left unmonitored they could easily develop into highly predatory groups as they have in states with comparable inmate populations.”).
restrictions upon inmates with certain religious beliefs." Judge Rendell argued that, in these cases, the court should require "a ‘tight’ or ‘closer’ fit between the correctional system’s admittedly legitimate interest and an inmate’s belief." While Judge Rendell felt that Turner allows a flexible application, and that Supreme Court precedent suggests a willingness to require a closer fit in certain circumstances, her reasoning was rejected by the majority as well as many other courts.

It is possible for courts to work within Turner to accommodate the heightened concerns raised by groups like the Five Percenters and the BGF. The second prong of Turner, for example, very much turns on the framing of the right in question. To return to In re Furnace, had the court considered the right in question to be “the right to read the works of the ideological founder of a political organization the inmate adheres to,” rather than “the right to read outside material,” the case would necessarily have been decided differently. Indeed, courts addressing Five Percenters’ claims have occasionally struggled with the second prong; Fraise v. Terhune required both an extended discussion and a fairly expansive understanding of “alternative means” to conclude that even a complete ban on the Lost-Found Lessons, the foundational documents of Five Percenter ideology, was acceptable. Judge Alito reasoned that because “nothing in the STG Policy restricts Five Percent Nation members from discussing or seeking to achieve self-knowledge, self-respect, responsible conduct, or righteous living,” the second prong of Turner was met. While the court was required to consider the right “sensibly and expansively,” it seems fair to consider this reasoning a stretch. Nonetheless, the Supreme Court has

---

252 Id.
253 Id. Note, however, that the Southern District of New York applied a roughly analogous approach in Maria v. Broaddus, finding that the prison had insufficiently shown that Five Percenters were per se dangerous. See supra note 101.
254 See supra notes 179-84 and accompanying text.
255 Fraise, 283 F.3d at 519-20.
256 Id. at 519.
257 Id. at 518 (citations omitted).
258 Indeed, Judge Rendell in dissent noted:

In the course of this treatment, the FPN member is barred from the teachings, which are at the heart of the Five Percent Nation religious experience. Furthermore, to be released from close custody he must promise to never again affiliate with FPN. Thus, the desired result of the treatment is to eradicate the belief. It is difficult to see how, realistically, there are “alternate means” here.
indicated that it does not value this prong as heavily as the first.\textsuperscript{259}

Similarly, the fourth prong is subject to a degree of judicial interpretation. While not requiring a least-restrictive-means analysis,\textsuperscript{260} the inmate may still attempt to show that the prison had the means to solve its problem while “accommodat[ing] the prisoner’s rights at de minimis cost.”\textsuperscript{261} This prong could be used to weigh the penological interest against the constitutional right at stake; a balancing test between the interests of the prison and prisoner seems to be the only way to reach just outcomes, and the fourth prong is the only part of \textit{Turner} that could facilitate such a test. When the right involves religious or political speech and affiliation, the alternate methods for the prison (for example, restrictions on materials related to violence, as opposed to blanket bans)\textsuperscript{262} should be subjected to closer examination. Similarly, the degree of deference granted to the prison allows the imposition of particularly severe deprivations, which should be considered when determining what alternative means are available to the prisons; surely there are ways to handle prison gangs without indeterminate, nearly indiscriminate application of long-term solitary confinement, blanket bans on literature, and isolation from fellow believers.\textsuperscript{263}

\textbf{CONCLUSION}

Regardless of whether the solution is a new test, a tweaking of the interpretation of the first prong of \textit{Turner}, or a degree of flexibility in considering the second and fourth prongs, the current approach is deeply problematic. Punishing people for political or religious views runs afoul of foundational American values,\textsuperscript{264} and even prisoners should not be subject to such treatment. Moreover, the inability of the STG system to adequately protect religious freedom may become more problematic in light of Islamic terrorism.

\textit{Id.} at 529 (Rendell, J., dissenting).
\textsuperscript{259} \textit{See supra} note 38 and accompanying text.
\textsuperscript{260} \textit{See supra} note 38 and accompanying text.
\textsuperscript{261} Id. at 91.
\textsuperscript{263} \textit{See supra} Part III.
\textsuperscript{264} \textit{See, e.g.,} Joint Anti-Fascist Refugee Comm. v. McGrath, 341 U.S. 123, 143 (1951) (Black, J., concurring) (“[T]he system adopted effectively punishes many organizations and their members merely because of their political beliefs and utterances, and to this extent smacks of a most evil type of censorship.”).
There are tens of thousands of Muslims in American prisons, mostly black men, well out of proportion to the Muslim civilian population. While Islam has long been popular in prisons, its prevalence took on a sinister air in light of the September 11 attacks; the FBI views prisons as “fertile ground for extremists.” As the tone of anti-Muslim discourse has ramped up recently, the treatment of the Five Percenters stands as uncomfortable precedent for religion-based prison regulations. Although the decision in Turner begins its analysis with the declaration that “[p]rison walls do not form a barrier separating prison inmates from the protections of the Constitution,” the standard it established allows courts to erect just such a barrier in the face of important constitutional protections. The courts should begin to reconsider Turner’s application before they are faced with attempts at even more severe restrictions.

Justin L. Sowa†

---

265 Zoll, supra note 137.
266 Id.
† J.D., Brooklyn Law School, 2012; B.M., New York University, 2007. I would like to thank Professors Nelson Tebbe and Ursula Bentele for their advice and guidance, as well as Professor Ted Swedenburg of the University of Arkansas for allowing me to use his unpublished research and directing me to other indispensable resources. I would also like to thank the Brooklyn Law Review staff and editors for their tireless work, and for allowing me to cite the Wu-Tang Clan in a work of legal scholarship.
Homebuyer Beware

MERS AND THE LAW OF SUBSEQUENT PURCHASERS

INTRODUCTION

Imagine you are in the market for a new home. You scour local real estate listings, you vet real estate agents, and you canvas neighborhoods until finally a property grabs your attention. You know this is where you will spend the rest of your life. So you make the homeowner an offer, he accepts, and you proceed to the closing. Lawyers, real estate agents, and title insurers all gather around a table. Everything appears to be going smoothly: you distribute your purchase funds to the seller, he pays his remaining balance on the mortgage, and the owner of his mortgage—a company named Mortgage Electronic Registration Systems, Inc. (MERS)—discharges the mortgage and removes its lien on the property. Because your title search revealed that MERS possessed the only outstanding interest against the property, you are satisfied that the title is free from adverse claims. Accordingly, you close the deal.

Six months pass. You have finally finished moving into your new home. You have met your new neighbors, found a new favorite restaurant, and enrolled your children into a new school. Moreover, your finances have remained stable, and you are current on your mortgage payments. As a result, you are perplexed when you return home one day to an unwelcome surprise: a foreclosure notice. You immediately call your bank in protest, but the bank actually confirms its receipt of your mortgage payments. In fact, the foreclosure notice is not even from your lender.

You hire a lawyer to represent you, but to no avail. The law is not on your side. Although you performed a title search and found only MERS’s name in the land records, another unrecorded claim existed. As it turns out, MERS held only “legal title” to the property, while another party actually owned...
the right to the mortgage payments. Accordingly, after the seller tendered the remaining balance to MERS, the company encountered a problem when it could not determine who was entitled to the funds. MERS later distributed the funds to the wrong individual. As a result, the true owner retains a claim against the underlying property. You lose your dream home. Worse still, you may not be alone.

Subsequent purchasers of MERS-mortgaged properties throughout the nation may find themselves in similar danger. MERS is listed on more than 65 million mortgages, or “approximately 60% of all mortgage loans in the United States.” Moreover, due to documentation errors and “shoddy recordkeeping practices,” the likelihood that MERS cannot identify the true owner of a particular mortgage note is quite substantial. This generates tremendous risks for subsequent purchasers. For example, under New York law, a discharge of the mortgage without a corresponding discharge of the mortgage note remains ineffective as against the owner of the note. Therefore, when MERS discharges a mortgage without knowledge of the note owner's identity and later distributes funds to the wrong MERS member, the owner of the note can

---


6 See Congressional Oversight Panel, supra note 4, at 48 (finding that “mortgage companies who filed claims . . . in bankruptcy cases of homeowners did not attach a copy of the note to 40% of their claims”).

redeem his security interest in the property.\textsuperscript{8} Even where the property is conveyed to a subsequent purchaser, courts will refuse to protect that party because he cannot satisfy the requirements of a bona fide purchaser for value.\textsuperscript{9} As demonstrated, this will cause a subsequent purchaser to lose his property.

These problems could have a devastating effect on the real estate market.\textsuperscript{10} As prospective purchasers and title insurers become aware of these risks, owners of MERS-mortgaged properties will suffer from “clouded title”\textsuperscript{11} as potential buyers begin to avoid their properties.\textsuperscript{12} This will affect homeowners’ ability to alienate their property, and it could cause widespread deadlocks in the real estate market—at least to the extent that MERS encounters documentation problems.\textsuperscript{13} Nevertheless, courts can avoid these dangers by utilizing principles of agency law. In particular, if courts accept MERS’s authority to act as an agent for lenders, those lenders would be bound by MERS’s discharges and would lose any claims against the property.\textsuperscript{14} Although the adoption of an agency theory might harm some homeowners by insulating MERS from attack in other areas,\textsuperscript{15} it would actually protect the vast majority of homeowners of MERS-mortgaged

\textsuperscript{8} See Assets Realization Co., 98 N.E. at 458-59; Walentas, 813 N.Y.S.2d at 370; Polomaine, 519 N.Y.S.2d at 934.

\textsuperscript{9} See Curtis v. Moore, 46 N.E. 168, 169 (N.Y. 1897).

\textsuperscript{10} See 1 Joyce D. Palomar, Title Insurance Law § 1.2 (1994) (“Stability of land titles is critical not only to individual property owners, but also to society as a whole. . . . Development will not occur if lenders cannot be relatively certain that their real property collateral will be marketable.”).

\textsuperscript{11} Robo-Signing, Chain of Title, Loss Mitigation, and Other Issues in Mortgage Servicing Before the Subcomm. on Hous. & Cnty. Opportunity of the H. Fin. Servs. Comm., 111th Cong. 1 (2010) (written statement of Adam J. Levitin, Associate Professor of Law, Georgetown University Law Center), available at http://financialservices.house.gov/Media/file/hearings/111/Levitin111810.pdf; see 65 Am. Jur. 2d Quieting Title § 13 (2011) (“A cloud upon title may . . . be defined as . . . an apparent defect in the title that has the tendency, even in a slight degree, to cast doubt upon the owner’s title, and to stand in the way of the full and free exercise of his or her ownership.”).

\textsuperscript{12} See 1 Joyce D. Palomar, supra note 10, § 1.2 (“Purchasers want assurance that the title is good before they invest money, time, and care, not damages from the grantor when the title proves to be defective.”).

\textsuperscript{13} See Congressional Oversight Panel, supra note 4, at 46-51 (explaining that lenders “documentation irregularities” may be quite “pervasive”).

\textsuperscript{14} See 2A N.Y. Jur. 2d Agency § 291 (2011) (“A principal is liable on contracts entered into on its behalf by an authorized agent.”).

\textsuperscript{15} In particular, a court’s recognition of MERS’s agency relationship with lenders would enable the company to initiate foreclosure actions on lenders’ behalf. See, e.g., Pantoya v. Countrywide Home Loans, Inc., 640 F. Supp. 2d 1177, 1188-89 (N.D. Cal. 2009) (internal quotation marks omitted) (holding that “[u]nder California law, a trustee, mortgagee, or beneficiary or any of their authorized agents may conduct the foreclosure process” and finding that “the Deed of Trust expressly designated MERS as the nominee of the lender and as the beneficiary”).
properties, as well as prospective purchasers of those properties.¹⁶

This note, in Part I, provides a brief overview of the MERS registry and how it operates within the mortgage finance industry. Part II provides a historical background of the property doctrines affecting MERS and the law of subsequent purchasers, including law pertaining to mortgages and the recording statutes. Part III discusses modern developments in mortgage finance by tracing the market’s evolution toward mortgage securitization and explaining how MERS alters the traditional securitization framework. Part IV explores public reactions to MERS, including various courts’ decisions relating to MERS, the company’s recent legal battles with county recorders and Attorneys General, scholarly analysis of MERS, as well as attention MERS has received from the public at large. Part V analyzes the frequently overlooked problems MERS poses for subsequent purchasers. In particular, this Note argues that subsequent purchasers of MERS-mortgaged properties will not constitute bona fide purchasers for value and that MERS’s documentation problems pose severe risks to them and to the real estate market generally. Finally, Part VI suggests that courts can avoid these issues by recognizing MERS’s authority to act as an agent for its members.

I. MERS

MERS operates a large, electronic document registry that “track[s] ownership interests in residential mortgages.”¹⁷ Financial institutions¹⁸ created MERS in response to the perceived inefficiency and costliness of the traditional recording system,¹⁹ under which lenders were required to

---

¹⁶ Recent census data reveal that most Americans are current on their mortgage payments, while less than 19 percent are delinquent or involved in foreclosure. See U.S. CENSUS BUREAU, STATISTICAL ABSTRACT OF THE UNITED STATES: 2012, at 743 tbl.1194 (131st ed. 2011) (estimating that 4.6 percent of residential mortgage loans were in the foreclosure process at year-end in 2010, 5.0 percent were entering the foreclosure process at year-end 2010, and 9.3 percent were “delinquent 30 days or more”).


record changes in mortgage ownership with county recorders and pay a fee.20 Indeed, on its website, MERS states that it was “created . . . to streamline the mortgage process by using electronic commerce to eliminate paper”21 and that its “mission is to register every mortgage loan in the United States on the MERS® System.”22 Not surprisingly, financial institutions that are active in the mortgage finance industry find MERS attractive precisely for these reasons: it facilitates the efficient transfer of mortgages and mortgage notes among numerous parties,23 and it avoids the costly and often slow process of recording these transfers with the county clerk.24

MERS alters the traditional mortgage financing system by permitting “[l]enders [to] identify MERS as nominee and mortgagee for its members’ successors and assignees.”25 Once a mortgage is registered with MERS, “the beneficial ownership interest or servicing rights may be transferred among MERS members”26 in order to bundle mortgages into securities more effectively.27 However, throughout this entire process, “MERS remains the mortgagee of record in local county recording offices regardless of how many times the mortgage is transferred, thus freeing MERS’s members from paying the recording fees that would otherwise be furnished to the relevant localities.”28 Only
upon a transfer to a nonmember or a discharge of the mortgage will the company seek to record any change in mortgage ownership. Moreover, notwithstanding its constructive claim of ownership with the county recording offices, “MERS does not lend money, does not receive payments on promissory notes, and does not service loans by collecting loan payments.”

II. PROPERTY DOCTRINES AFFECTING MERS

To fully understand the MERS system, how it operates, and its effects on the real estate market, one must first understand the property doctrines that provide its foundation.

A. Mortgages

“A mortgage is a conveyance or retention of an interest in real property as security for performance of an obligation,” where the obligation “is almost always a loan of money evidenced by a promissory note.” To avoid confusion, one must distinguish between a mortgage, on one hand, and a mortgage note or promissory note, on the other. A “mortgage” is a security interest that a lender holds in the underlying property, whereas a “promissory note” represents a borrower’s obligation to repay his loan. Conceptually, mortgage transactions are structured as follows. First, a lender provides funds to a borrower in order to initiate a property transfer from a third party. Second, the borrower executes the transfer and obtains title to the property. In exchange for the borrowed funds, the

29 See Romaine, 861 N.E.2d at 83 n.4 (“If a MERS member transfers ownership interest or servicing rights in a mortgage loan to a non-MERS member, an assignment from the MERS member to the non-MERS member is recorded in the County Clerk’s office and the loan is deactivated within the MERS system.”).
31 Silverberg, 926 N.Y.S.2d at 536.
33 JOHN G. SPRANKLING, UNDERSTANDING PROPERTY LAW 358 (2d ed. 2008) (emphasis omitted).
34 See Grant S. Nelson & Dale A. Whitman, REAL ESTATE FINANCE LAW § 1.1 (5th ed. 2007).
36 See Nelson & Whitman, supra note 34, § 1.1.
37 See id. § 1.5. Depending on the jurisdiction, the borrower may hold legal title to the property—subject to his satisfaction of the mortgage—or the lender may hold legal title to the property, which reverts back to the borrower upon satisfaction of the mortgage. See id. (explaining the distinction between the “title theory” and “lien theory”).
borrower (mortgagor) delivers to the lender (mortgagee) a
promissory note and a mortgage.38

A mortgage is commonly terminated either by
foreclosure39 or through a satisfaction and discharge.40 Where a
borrower defaults on his obligation to the lender, the lender has
the power to redeem his security interest in the property
through foreclosure.41 By contrast, where a borrower makes
“payment at or before maturity,”42 the borrower’s payment will
terminate the mortgage.43 Nevertheless, whether the owner of
the mortgage—as opposed to the owner of the mortgage note—
can properly discharge a borrower’s debt may prove important to
borrowers and their successors in interest. For example, in New
York, a note secured by a mortgage “will not be discharged by
payment to the record holder if . . . the note and mortgage ha[ve]
already been transferred . . . , even though no assignment has
been recorded.”44 Indeed, courts have cautioned that “[t]he
satisfaction of the mortgage [is] not a blanket release of [a
borrower’s] obligations under the note.”45 Accordingly, a payment
to the wrong individual would be “at [the borrower’s] peril.”46

B. Recordation and the Recording Acts

The recording acts were developed “to secure a
permanent record of landholding, and to prevent fraudulent

---

38 See RESTATEMENT (THIRD) OF PROP.: MORTGAGES § 1.1 cmt. a, illus. 1 (1997).
39 See, e.g., N.Y. REAL PROP. ACTS. LAW § 1301 (McKinney 2009).
40 See, e.g., id. § 1921(1) (“After payment of authorized principal, interest and
any other amounts due . . . , a mortgagee of real property . . . must execute and
acknowledge . . . a satisfaction of mortgage . . . .”).
42 When an obligation secured by a mortgage becomes due, the mortgagee may
either . . . obtain a judgment against any person who is personally liable on
the obligation and, to the extent that the judgment is not satisfied, foreclose
the mortgage on the real estate for the balance . . . or . . . foreclose the
mortgage and, to the extent that the proceeds of the foreclosure sale do not
satisfy the obligation, obtain a judgment for the deficiency against any person
who is personally liable on the obligation . . . .

Id.
43 12 THOMPSON ON REAL PROPERTY, SECOND THOMAS EDITION § 101.03(c)
(David A. Thomas ed., 2008).
44 See id.
2006); see Signal Fin. of N.Y., Inc. v. Polomaine, 519 N.Y.S.2d 933, 934 (Civ. Ct. 1987)
(holding that “[t]he underlying obligation, as evidenced by the note, survive[d] the
Discharge of the Mortgage”); see also 59 C.J.S. MORTGAGES § 628 (2009); 78 N.Y. JUR. 2D
MORTGAGES AND DEEDS OF TRUST § 375 (2003).
46 Assets Realization Co., 98 N.E. at 459.
claims to lands by concealment of transfers.”\textsuperscript{47} In large part, the recording acts effectuate these purposes by providing a “fruitful source[] of notice to a purchaser”\textsuperscript{48} or by documenting which landowners have won the “race to the record.”\textsuperscript{49} United States jurisdictions emphasize these features of recordation to varying degrees through their recording statutes.\textsuperscript{50} In particular, jurisdictions utilize three variations: race statutes,\textsuperscript{51} notice statutes,\textsuperscript{52} and race-notice statutes.\textsuperscript{53}

Race statutes\textsuperscript{54} “protect[] the first purchaser to record.”\textsuperscript{55} Accordingly, a purchaser’s “notice or knowledge of prior unrecorded claims is irrelevant” to determining his protection under the statute.\textsuperscript{56} Simply put, as between two competing interests in land, “the first to record has priority.”\textsuperscript{57}

By contrast, notice statutes\textsuperscript{58} generally protect only those subsequent purchasers who can satisfy the requirements of a “bona fide purchaser for valuable consideration.”\textsuperscript{59} A bona fide purchaser is a purchaser “without notice of prior unrecorded interests that are subject to the recording act.”\textsuperscript{60} Under this standard, “several kinds of notice . . . may disqualify a person from . . . protection,”\textsuperscript{61} including actual notice, constructive notice, and inquiry notice.\textsuperscript{62} Actual notice occurs where the purchaser has “actual knowledge of the prior interest.”\textsuperscript{63} Constructive notice occurs where “a reasonable title

\textsuperscript{47} 4 AMERICAN LAW OF PROPERTY § 17.5 (A. James Casner ed., 1952).
\textsuperscript{48} Id. § 17.17.
\textsuperscript{49} Id. § 17.5.
\textsuperscript{50} See generally id. (explaining the various recording statutes that states have adopted).
\textsuperscript{51} See 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(a).
\textsuperscript{52} See id. § 92.08(b).
\textsuperscript{53} See id. § 92.08(c).
\textsuperscript{54} Race statutes are very uncommon across United States jurisdictions. Only Arkansas, Louisiana, North Carolina, Ohio, and Pennsylvania have enacted race statutes for mortgages. 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.5, 545 n.63; 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(a), 158 nn.283-84.
\textsuperscript{55} 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(a).
\textsuperscript{56} Id.
\textsuperscript{57} See id.
\textsuperscript{58} Notice statutes are far more common than race statutes. Alabama, Arizona, Connecticut, Delaware, Florida, Iowa, Kansas, Kentucky, Maine, Massachusetts, Missouri, New Hampshire, New Mexico, Oklahoma, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, and West Virginia have all enacted notice statutes for mortgages. 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.5, 545 n.63; see also 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(b), 159 n.286.
\textsuperscript{59} 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.10; see 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(b).
\textsuperscript{60} 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(c).
\textsuperscript{61} Id.
\textsuperscript{62} Id.
\textsuperscript{63} Id.
search of the public real estate records would have revealed a prior interest.\textsuperscript{64} Inquiry notice occurs where a purchaser has “notice of facts which would have caused a reasonable person to make further inquiries.”\textsuperscript{65} Pure notice statutes therefore represent the converse of race statutes: a subsequent purchaser must be bona fide,\textsuperscript{66} but he “need not record to qualify for . . . protection.”\textsuperscript{67}

Race-notice statutes\textsuperscript{68} occupy an intermediary position between race and notice statutes, combining elements of both.\textsuperscript{69} In particular, race-notice statutes require both that the subsequent purchaser record first\textsuperscript{70} and that he represent a bona fide purchaser for value.\textsuperscript{71} So long as a purchaser can satisfy these requirements, the statutes will protect his interest.\textsuperscript{72}

III. MODERN MORTGAGE FINANCE: THE RISE OF SECURITIZATION AND MERS

Developments in mortgage finance that have occurred over the last century illustrate why financial institutions created MERS and how MERS facilitates their businesses. Perhaps the most important development for mortgage finance has been the creation of the Federal National Mortgage Association (Fannie Mae) and the Federal Home Loan Mortgage Corporation (Freddie Mac).\textsuperscript{73} As a matter of financial policy, Fannie Mae and Freddie Mac both serve important functions for the mortgage market.\textsuperscript{74} First, “[t]hey facilitate the flow of capital from areas of the country where funds are plentiful to places in which mortgage money is in short supply,” thereby providing liquidity to local banks.\textsuperscript{75} Second, “they move capital investment from other

\textsuperscript{64} Id.
\textsuperscript{65} Id.
\textsuperscript{66} See 4 AMERICAN LAW OF PROPERTY, supra note 47, §§ 17.10-17.11; 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.09(c).
\textsuperscript{67} 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(b).
\textsuperscript{68} Race-notice statutes are the most common recording statute. Alaska, California, Colorado, District of Columbia, Georgia, Hawaii, Idaho, Illinois, Indiana, Maryland, Michigan, Minnesota, Mississippi, Montana, Nebraska, Nevada, New Jersey, New York, North Dakota, Oregon, South Dakota, Utah, Washington, Wisconsin, and Wyoming have all enacted race-notice statutes for mortgages. 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.5, 545 n.63; see also 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(c), 160 n.288.
\textsuperscript{69} See 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(c).
\textsuperscript{70} See id.
\textsuperscript{71} See id.
\textsuperscript{72} See id.
\textsuperscript{73} See NELSON & WHITMAN, supra note 34, § 11.3.
\textsuperscript{74} See id.
\textsuperscript{75} See id.
sectors of the national economy into the mortgage market.” 76
Finally, they “even out regional differences in interest rates” and
“create a means of spreading the risks inherent in mortgage
portfolios that [are] heavily concentrated in one state or region.” 77
Fannie Mae and Freddie Mac accomplish these goals by purchasing “vast amounts of home loans from savings banks
and other lenders” throughout the country. 78 The companies
then issue “mortgage-backed securities,” 79 which are based on
pools of the underlying mortgage loans. 80 This process requires
the cooperation of several different entities. 81 Additionally, prior
to the creation of MERS, most of these transactions were
recorded with the county recorder’s office. 82 Figure A 83 below
represents a securitization chain where MERS is not involved.

In the wake of MERS’s implementation, financial
institutions could execute the securitization process much more

---

76 See id.
77 LEFCOE, supra note 35, at 453 n.205.
78 Id. at 454.
79 See NELSON & WHITMAN, supra note 34, § 11.3.
80 See LEFCOE, supra note 35, at 452.
81 Id. at 454.
83 This diagram is based on a symposium presentation on Securitization and
Governance given by Nancy Wallace. See Nancy Wallace, Presentation at the U.C.
Berkeley Symposium: Private-Label Residential Mortgage Securitization: Recording
Innovations and Bankruptcy Remoteness 7 (Mar. 11, 2011), available at
quickly. Indeed, some have even suggested that this may have played a role in causing the financial crisis. Nevertheless, speed is not MERS’s only advantage for financial institutions. The MERS system also allows its members to avoid paying recording fees on transfers occurring within the registry. Indeed, county recorders around the country have begun to file lawsuits against MERS in an effort to recover lost fees. Moreover, financial institutions appear to lose nothing by utilizing this system. Once a loan enters the MERS system, members authorize MERS to act on their behalf with respect to the property. Accordingly, financial institutions are able to increase the volume of their business and reduce costs while also retaining the ability to initiate legal proceedings through MERS. Figure B below represents a securitization chain where MERS acts as mortgagee of record.

---

84 See Dennis & Cha, Loan Chaos, supra note 27 (“MERS allow[s] big financial firms to trade mortgages at lightning speed . . . .”).
85 See id. (suggesting that “[w]ithout this system, the business of creating massive securities made of thousands of mortgages would likely have never taken off”).
86 See Silverberg, 926 N.Y.S.2d at 536 (“MERS delivers savings to the participants in the real estate mortgage industry by allowing those entities to avoid the payment of fees which local governments require to record mortgage assignments.” (citation omitted)).
88 See Sample MOM Mortgage, supra note 1, at 3 (“Borrower does hereby mortgage, grant and convey to MERS (solely as nominee for Lender and Lender’s successors and assigns) and to the successors and assigns of MERS, the following described property . . . .”)
90 This diagram is based on a symposium presentation on Securitization and Governance given by Nancy Wallace. See Wallace, supra note 83, at 11.
MERS has achieved significant notoriety in the wake of the financial collapse. MERS currently finds itself at the center of a national foreclosure crisis—where approximately one in twenty homeowners faces foreclosure\(^91\) and MERS is listed on more than 60 percent of mortgages nationwide.\(^92\) As a result, MERS’s name has surfaced in thousands of foreclosure and bankruptcy actions throughout the country.\(^93\) Nevertheless, courts remain divided over the legal status of MERS’s business practices.\(^94\) Some have held that MERS lacks standing to

\(^{91}\) See U.S. CENSUS BUREAU, supra note 16, at 743 tbl.1194 (estimating that 4.6 percent of residential mortgage loans were in the foreclosure process at year-end in 2010 and that 5.0 percent were entering the foreclosure process at year-end 2010).


\(^{94}\) Compare Pantoja, 640 F. Supp. 2d at 1188-89 (holding that “[u]nder California law, a trustee, mortgagee, or beneficiary or any of their authorized agents may conduct the foreclosure process” and finding that “the Deed of Trust expressly designated MERS as the nominee of the lender and as the beneficiary” (internal quotation marks omitted)), with In re Agard, 444 B.R. 231, 253 (Bankr. E.D.N.Y. 2011).
pursue foreclosure actions or bankruptcy stay relief motions. Other courts have ruled to the contrary. This lack of uniformity has created potentially devastating uncertainty for the mortgage industry.

Recent lawsuits from Attorneys General and county recorders throughout the country have only added to this uncertainty. In particular, Attorneys General in Delaware and Massachusetts have both filed recent lawsuits against MERS, alleging “deceptive trade practices” and “deceptive business practices,” respectively. Moreover, Nevada’s Attorney General has brought suit against Lender Processing Services, a major “default and foreclosure processor that works behind the scenes for most large banks,” alleging “deceptive foreclosure practices” in the wake of the robo-signing scandal. Finally, county recorders throughout the United States have mounted legal attacks against MERS by claiming that the MERS system “bypass[es] local recording laws” and deprives counties of “millions in property recording filing fees.”

Given the current economic climate, it is not hard to understand why MERS has captured such widespread attention from the public and from publicly elected officials. Although economists declare that the U.S. recession officially...

(citing Steinbeck v. Steinbeck Heritage Found., 400 F. App’x 572, 575 (2d Cir. 2010)) ([T]he record of this case is insufficient to prove that an agency relationship exists under the laws of the state of New York between MERS and its members. . . . [T]he fact that MERS is named ‘nominee’ in the Mortgage is not dispositive of the existence of an agency relationship and does not, in and of itself, give MERS any ‘authority to act.’

95 See, e.g., Silverberg, 926 N.Y.S.2d at 539.
96 See, e.g., In re Agard, 444 B.R. at 253.
97 See, e.g., Pantoja, 640 F. Supp. 2d at 1188-89.
ended in June 2009, household incomes have continued to fall by approximately 7 percent, unemployment rates have remained well above equilibrium levels, and foreclosure rates have continued to surge. Many Americans appear to be upset over a crisis that banks caused primarily through their risky business practices. Now that many Americans are facing foreclosure at the hands of those very same institutions,


106 See Alex Tanzi, Bloomberg U.S. Mortgage Delinquency, Foreclosure Rates, BLOOMBERG (Oct. 17, 2011) (showing increases in foreclosure rates between September 30, 2009 and September 30, 2011 from 7.22 percent to 9.13 percent for prime loans, from 11.68 percent to 13.45 percent for Alt-A loans, and from 16.51 percent to 18.03 percent for subprime loans); Alex Tanzi, Bloomberg U.S. Mortgage Delinquency, Foreclosure Rates, BLOOMBERG (Oct. 15, 2009) (showing increases in foreclosure rates between September 30, 2007 and September 30, 2009 from 0.79 percent to 7.22 percent for prime loans, from 1.56 percent to 11.68 percent for Alt-A loans, and from 6.16 percent to 16.51 percent for subprime loans).


109 See U.S. Census Bureau, supra note 16, at 743 tbl.1194.
they may view MERS as a symbol of the same overzealous financial culture that contributed to the crisis. Indeed, some have even advocated confronting MERS directly in quiet-title actions and foreclosure proceedings.

V. RISKS FOR SUBSEQUENT PURCHASERS OF MERS-MORTGAGED PROPERTY

While MERS has captured widespread public attention for its role in the foreclosure crisis, the company’s business practices raise equally important questions for purchasers of MERS-mortgaged properties. As recent events have shown, failures by financial institutions and MERS to properly document assignments of mortgage notes have created a risk that neither can determine who owns the note or recreate the chain of title for any particular mortgage. Courts in several states have already held that MERS’s inability to produce both the mortgage and the

110 This is particularly true where many of the most prominent banks “played a critical role in the development of MERS.” Shareholders, MERS, http://www.mersinc.org/about/shareholders.aspx (last visited Sept. 23, 2011) (identifying Bank of America, CitiMortgage, Inc., HSBC Finance Corp., Wells Fargo Bank, GMAC Residential Funding Corp., Mortgage Bankers Association, Fannie Mae, and Freddie Mac as MERS shareholders).

111 See Christopher Ketcham, STOP PAYMENT! A Homeowners’ Revolt Against the Banks, HARPER’S, Jan. 2012, at 29 (chronicling homeowners’ efforts to “attack the banking industry with the fine print of real estate law” by “demanding proof of who really own[s] their loan”).


113 See Verified Complaint at 30, Delaware v. MERSCORP, Inc., No. 6987, (Del. Ch. Oct. 27, 2011) (citing Transcript of Hearing at 6-8, 16-17, In re Kemp, 440 B.R. 624, 626 (Bankr. D.N.J. 2010) (No. 08-18700) (Aug. 11, 2009)) (“In a 2009 hearing . . . , an employee for the Bank of America entity responsible for servicing the securitized Countrywide mortgage loans testified under oath that Countrywide did not have a practice of delivering original documents such as the note to the Trustee . . . . In addition, the same employee further testified that allonges are typically prepared in anticipation of foreclosure litigation, rather than at the time the mortgage loans are purportedly securitized.”); see also Gongloff, supra note 4 (“The crisis has been escalating for several weeks, as banks suspend foreclosures across the country, citing flaws they have uncovered, including faulty or missing documentation. Tales of mismanagement within the foreclosure process—including so-called robo-signers, who were paid to rubber stamp documents without properly reviewing them—are emerging daily.”); Gretchen Morgenson, Flawed Paperwork Aggravates Foreclosure Crisis, N.Y. TIMES (Oct. 3, 2010), http://www.nytimes.com/2010/10/04/business/04mortgage.html.
note is fatal to its ability to foreclose. However, MERS’s inability to produce the note generates problems that transcend the foreclosure arena. In particular, it threatens title to MERS-mortgaged properties that are subsequently purchased by third parties, and it places a cloud on title for existing homeowners whose properties list MERS as the mortgagee of record.

A. Framing the Issue

MERS raises a host of issues that could pose risks to subsequent purchasers. For example, imagine the following scenario, represented below by Figure C. A is a homeowner who financed the purchase of his home with a mortgage. A’s lender, Bank Z, provided funds to A in order make the purchase, but in exchange, Bank Z required A to execute two documents. First, A executed a mortgage on the property to MERS, “solely as nominee for [Bank Z] and [Bank Z]’s successors and assigns.” Second, A executed a promissory note to Bank Z, secured by the mortgage to MERS. After closing the transaction, MERS recorded its mortgage at the county recorder’s office. Subsequently, Bank Z assigned the note to Bank Y, who then assigned the note to Bank X, who later assigned the note to Bank W. All of these assignments were executed in order to securitize A’s mortgage note, but none were recorded. Instead, the banks utilized MERS to track their assignments. However, MERS and the relevant banks either lost track of the note or failed to properly execute their assignments. As a result, the banks suffer from missing or inconsistent paperwork with respect to the note.

Meanwhile, A wanted to sell his home and remained current on his mortgage payments. B demonstrated interest in A’s property and decided to buy it. At the closing, B distributed funds to A, which A tendered to MERS in order to satisfy the remaining balance on his note. In exchange for A’s payment,


115 This visual representation is intended to provide the reader with a conceptual understanding of the various relationships that might exist in a common real estate transaction. For the purpose of clarity, many facets of these relationships have been simplified, and extraneous parties have been omitted. For a more comprehensive visual representation of the mortgage securitization process with MERS, see Wallace, supra note 83, at 11; see also Powell & Morgenson, supra note 112 (follow “How a Mortgage Moves Through MERS” hyperlink).

116 Sample MOM Mortgage, supra note 1, at 3.
MERS discharged its mortgage on the property. A then conveyed title to B. However, neither A nor B received any assurance that A’s debt was satisfied from the actual owner of A’s note. Instead, both relied on MERS’s discharge of the mortgage, in its capacity as mortgagee of record. Therefore, if MERS fails to distribute the proceeds to the owner of the note, it may be possible that the owner of A’s note has a claim against B’s property, which was pledged as security for the note. The viability of any claims against B’s property hinges on two critical questions. First, can MERS properly discharge A’s debt obligation and, in the process, extinguish any future claims by the owner of the note? Second, if MERS cannot properly discharge A’s debt, will courts recognize B’s claim to the property as superior to the noteholder’s prior unrecorded interest, on the ground that B is a bona fide purchaser for value?

117 Depending upon the jurisdiction, MERS’s interest in the mortgage may represent legal title to the property or a lien against the property. See Nelson & Whitman, supra note 34, § 1.5 (explaining the distinction between the “title theory” and “lien theory”). Since New York adopts a lien theory, MERS technically does not convey legal title to the borrower but instead releases its lien on the mortgaged property. See Barson v. Mulligan, 84 N.E. 75, 78 (N.Y. 1908) (“The mortgagee, having come to be regarded as a mere lienor, has no legal estate in the land covered by his mortgage...”).

118 See 4 American Law of Property, supra note 47, §§ 17.10-17.11; 11 Thompson on Real Property, supra note 42, § 92.09.
B. MERS’s Ability to Discharge the Debt

As a threshold matter, it is unclear whether MERS can properly discharge the borrower’s obligation, given that MERS holds only an interest in the mortgage. For example, under New York law, a mortgagee is entitled to discharge a borrower’s obligation upon the borrower’s satisfaction of the mortgage debt. Importantly for MERS, the relevant statute defines “mortgagee” as “the current holder of the mortgage of record or the current holder of the mortgage.” Therefore, one who holds only the mortgage without a corresponding interest in the note can nevertheless discharge a mortgage under the statute. Indeed, the statute expressly provides for a discharge of the mortgage even where the mortgagee cannot produce the note.

The inquiry does not end here, however. Notwithstanding MERS’s ability to discharge a mortgage for recording purposes, a question remains as to what legal effect that discharge will have upon the owner of the note. In New York, a note secured by a mortgage “will not be discharged by payment to the record holder if . . . the note and mortgage ha[ve] already been transferred . . . , even though no assignment has been recorded.” In other words, a borrower cannot satisfy his debt by paying the record holder of the mortgage if the record holder has already assigned its interests to a third party, regardless of the third party’s failure to record. To that effect, the New York Court of Appeals has cautioned that “[w]here a

---

119 See Sample MOM Mortgage, supra note 1, at 3 (“Borrower understands and agrees that MERS holds only legal title to the interests granted by Borrower in this Security Instrument . . . .” (emphasis added)); see also Residential Funding Co., LLC v. Saurman, 807 N.W.2d 412, 417 (Mich. Ct. App.) (“MERS, as mortgagee, only held an interest in the property as security for the note, not an interest in the note itself.”), rev’d on other grounds, 805 N.W.2d 183 (Mich. 2011).
120 N.Y. REAL PROP. ACTS. LAW § 1921(1) (McKinney 2011) (“After payment of authorized principal, interest and any other amounts due . . . , a mortgagee of real property . . . must execute and acknowledge . . . a satisfaction of mortgage . . . .”).
121 Id. § 1921(9)(a). The statutory definition of “mortgagee” also includes “any person to whom payments are required to be made” or “their personal representatives, agents, successors, or assigns.” Id.
122 See id.
123 See id. § 1921(4) (“If the mortgagee has delivered such satisfaction of mortgage in a timely manner and has certified that the note and/or mortgage are not in its possession as of such date, the mortgagee shall not be liable under this section if the mortgagee agrees to defend and hold harmless the mortgagor by reason of the inability or failure of the mortgagee to furnish the note or mortgage within the time period prescribed in this subdivision . . . .”).
125 Id.
party makes . . . a final payment in satisfaction of a bond and mortgage without taking a satisfaction and without requiring production of the instruments, or receiving some sufficient excuse for their nonproduction, the payment is at his peril and not good as against an assignee for value under an unrecorded assignment.”126 At bottom, New York law embodies the principle that a borrower should require his mortgagee to demonstrate contemporaneous ownership of the mortgage and note before tendering a final payment on the mortgage.127 Otherwise, the payment is made “at his peril.”128 This appears to flow from the maxim that “a transfer of the mortgage without the debt is a nullity.”129 Given that “a mortgage is but an incident to the debt which it is intended to secure,”130 it follows that the release of a mortgage should remain ineffective as against the owner of the note, who may not have received the final payment.131

Where MERS acts as mortgagee of record, these principles bear directly on MERS’s ability to properly discharge a borrower’s obligation. Since MERS has no interest in the underlying mortgage note,132 the company can release only its interest in the mortgage. However, given that “a transfer of the mortgage without the debt is a nullity,”133 a release of the

---

126 Id. at 459; cf. Restatement (Third) of Prop.: Mortgages § 6.4 cmt. c (1997) (“When payment or tender by the person primarily responsible for the debt has extinguished the mortgage, the payor derives little comfort unless a document can be recorded to clear the public records of the mortgage lien . . . . In some states it is customary for the mortgagee to provide an endorsement on the public records, to display the promissory note, marked ‘paid,’ to the recorder’s office personnel, or to return the original mortgage document.” (emphasis added)).

127 See Assets Realization Co., 98 N.E. at 458-59. It bears mentioning that this principle is analogous to the requirements imposed upon parties seeking foreclosure in New York. See Bank of N.Y. v. Silverberg, 926 N.Y.S.2d 532, 537 (App. Div. 2011) (“In a mortgage foreclosure action, a plaintiff has standing where it is both the holder or assignee of the subject mortgage and the holder or assignee of the underlying note at the time the action is commenced.”); U.S. Bank v. Collymore, 890 N.Y.S.2d 578, 580 (App. Div. 2009); Kluge v. Fugazy, 536 N.Y.S.2d 92, 93 (App. Div. 1988).

128 Assets Realization Co., 98 N.E. at 459.


130 Id.


132 See Sample MOM Mortgage, supra note 1, at 3 (“Borrower understands and agrees that MERS holds only legal title to the interests granted by Borrower in this Security Instrument . . . .” (emphasis added)); see also Residential Funding Co, LLC v. Saurman, 807 N.W.2d 412, 417 (Mich. Ct. App.) (“MERS, as mortgagee, only held an interest in the property as security for the note, not an interest in the note itself”), rev’d on other grounds, 805 N.W.2d 183 (Mich. 2011).

133 Bartholich, 36 N.Y. at 45.
mortgage without a release of the debt would likewise appear to be a nullity. Under these circumstances, a borrower's failure to require MERS to “produce and deliver up the instruments which are being paid and satisfied”—in particular, the mortgage note—could place his payment in jeopardy.134

Nevertheless, some authority exists135 to suggest that MERS may have the ability to discharge the borrower's debt as an agent “for [the] Lender and Lender’s successors and assigns.”136 In MERSCORP, Inc. v. Romaine,137 the New York Court of Appeals held that MERS had authority to record discharges “[a]s the nominee for the mortgagee of record or for the last assignee.”138 Moreover, in an opinion dissenting in part, Chief Judge Kaye also appeared to accept MERS's agency theory, suggesting that “the use of a nominee as the equivalent of an agent for the lender is apparent, and not unusual.”139 Nevertheless, Chief Judge Kaye agreed with the majority by stating,

[[Issues concerning the underlying validity of the MERS mortgage instrument—in particular, whether its failure to transfer beneficial interest renders it a nullity under real property law, whether it violates the prohibition against separating the note from the mortgage, and whether MERS has standing to foreclose on a mortgage—are best left for another day.140

The extent to which courts will recognize MERS's agency powers beyond the context of the recording statutes remains to be seen. On one hand, some courts appear to have followed the Court of Appeals’ formulation of MERS's agency authority.141 In particular, one court has held that “the language of the mortgage appoints MERS as nominee, or agent, for the lender and its successors and assigns for the purposes set forth therein.”142

134 Assets Realization Co., 98 N.E. at 459.
135 See MERSCORP, Inc. v. Romaine, 861 N.E.2d 81, 84 (N.Y. 2006).
136 Sample MOM Mortgage, supra note 1, at 3; see Brief for Petitioners-Appellants at 27, MERSCORP, Inc., v. Romaine, 743 N.Y.S.2d 562 (App. Div. 2002) (No. 2001-04792), 2001 WL 34687001 (citing RESTATEMENT (SECOND) OF AGENCY § 1 (1958)) (“MERS's relationship with its member lenders is that of agent and principal. This is a fiduciary relationship, which results from the manifestation of consent of one person to allow another to act on his behalf and subject to his control, and consent by the other to so act. The principal is the one for whom action is to be taken, and the agent is the one who acts.”).
137 Romaine, 861 N.E.2d 81.
138 Id. at 84.
139 Id. at 87 (Kaye, C.J., dissenting in part).
140 Id. at 87 n.* (emphasis added).
141 See, e.g., Deutsche Bank Nat'l Trust Co. v. Pietranico, 928 N.Y.S.2d 818, 829-30 (Sup. Ct. 2011).
142 Id.
Accordingly, the same court found that MERS was entitled “to exercise any and all of the interests granted by the borrower under the mortgage,” including “the right to foreclose and sell the property[,] and to take any action required of the lender.”\textsuperscript{143}

However, other courts appear to have contradicted the Court of Appeals’ approach in favor of a more exacting standard.\textsuperscript{144} In particular, one court has refused to find that an agency relationship existed between MERS and its members by explaining that “the fact . . . MERS is named ‘nominee’ in the Mortgage is not dispositive of the existence of an agency relationship and does not, in and of itself, give MERS any ‘authority to act.’”\textsuperscript{145} Additionally, another court has held that, “as nominee, MERS’s authority was limited to only those powers which were specifically conferred to it and authorized by the lender.”\textsuperscript{146} Although the mortgage gave MERS the right “to exercise any or all of those rights, [granted by the Borrowers to Countrywide] including, but not limited to, the right to foreclose and sell the Property,”\textsuperscript{147} the court nevertheless insisted that no party can initiate foreclosure without holding both the mortgage and the note at the time of the action.\textsuperscript{148} The implication of these cases is clear. Both suggest that the power to foreclose must be specifically authorized by the principal and that the language contained in the mortgage provides insufficient authorization. The same logic would also seem to extend to MERS’s ability to discharge the security instrument, which is another power the mortgage confers upon the mortgagee.\textsuperscript{149}

Nevertheless, the fact remains that no court has squarely addressed the issue of whether MERS’s discharges would be binding upon the owner of the note, particularly where the owner of the note is unknown\textsuperscript{150} and perhaps

\textsuperscript{143} Id. (internal quotation marks omitted).


\textsuperscript{145} \textit{In re Agard}, 444 B.R. at 253.

\textsuperscript{146} Silverberg, 926 N.Y.S.2d at 538 (internal quotation marks omitted).

\textsuperscript{147} Id. at 534 (alterations in original).

\textsuperscript{148} Id. at 557.

\textsuperscript{149} See Sample MOM Mortgage, \textit{supra} note 1, at 3 (“MERS (as nominee for Lender and Lender’s successors and assigns) has the right . . . to take any action required of Lender including, but not limited to, releasing and canceling this Security Instrument.”).

\textsuperscript{150} MERS’s principal might be unknown to MERS in two ways. First, MERS presumably cannot determine, upon origination of the mortgage, who will later receive an assignment of the note and on whose behalf MERS will act when discharging the obligation. Therefore, at the time of origination, the mortgage purports to create an agency relationship between MERS and a class of unknown third parties, one of whom
unknowable to MERS. However, as courts continue to examine MERS’s agency authority, they should recognize that their decisions carry immense implications for subsequent purchasers of MERS-mortgaged properties. In particular, where the note has been lost or improperly assigned, two dangers emerge. First, MERS may lack the ability to identify the note’s true owner, which precludes the company from recording a proper discharge on behalf of the owner of the note. Second, MERS may lack the ability to identify all assignees in the chain of title to the note, which means that MERS cannot rule out the existence of potential claims by unknown assignees to a portion of the note’s proceeds. Under these circumstances, unsatisfied parties to the original mortgage note may have claims against the property, as security for the note, if those parties do not receive the funds to which they are entitled. Because MERS cannot determine how to distribute funds where it cannot identify the note’s owner, this likelihood does not appear to be remote.

Accordingly, a subsequent purchaser may be at risk of losing his property when the note’s owner attempts to redeem the security interest, even though he takes the property totally unaware of any prior unrecorded assignees of the note or any potential claims they might possess. Indeed, when MERS releases its mortgage lien and the borrower then conveys the property to a third-party purchaser, the third party relies on the fact that all recorded interests appear to have been satisfied. Therefore, the question arises whether the recording statutes will protect the subsequent purchaser against prior unrecorded interests. In most jurisdictions, the third party must represent a bona fide purchaser for value.

will later receive the note. Second, MERS may also remain unaware of its principal’s identity at the time of discharge, given that MERS has the apparent authority to discharge the obligation itself and may refrain from investigating the note’s ownership. MERS’s principal might be unknowable to the extent that documentation errors preclude MERS from ascertaining the true identity of the note’s owner.

152 See supra notes 132-34 and accompanying text.

153 See supra note 113 and accompanying text.

154 See supra notes 132-34 and accompanying text.

155 See 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.5, 545 n.63 (demonstrating that nearly every United States jurisdiction requires some form of notice to a subsequent purchaser).

156 See id. §§ 17.10-17.11; 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.09.
C. The Recording Acts and Bona Fide Purchasers for Value

1. Race Jurisdiction

In race jurisdictions, the recording acts generally “protect[] the first purchaser to record.” Accordingly, a purchaser’s “notice or knowledge of prior unrecorded claims is irrelevant” to determining his protection under the statute. For example, North Carolina’s statute provides that “instruments registered in the office of the register of deeds shall have priority based on the order of registration as determined by the time of registration.” Under this statute, a subsequent purchaser of a MERS-mortgaged property would be secure in his title to land only if he records his interest first in the local recorder’s office. However, if the true owner of the note beats the subsequent purchaser to record, the true owner’s prior unrecorded interest in the land would prevail.

2. Race-Notice Jurisdiction

In race-notice jurisdictions, a subsequent purchaser must satisfy two requirements. First, he must “record [his interest] before the prior unrecorded claim is recorded.” Second, he must be “a purchaser for value without knowledge or notice of the prior unrecorded claim.” This second requirement demands, in other words, that he represent a bona fide purchaser for value.
value. For example, New York’s statute provides that “[e]very . . . conveyance not . . . recorded is void as against any person who subsequently purchases . . . the same real property . . . in good faith and for a valuable consideration, . . . and whose conveyance . . . is first duly recorded.” Of course, New York courts have supplied these statutory requirements with a judicially-crafted meaning. In particular, courts have refused to recognize a party’s status as a bona fide purchaser where that party possesses one of three features: actual notice, constructive notice, or inquiry notice. A purchaser has actual notice where he purchases land “with knowledge of a prior outstanding title by an unrecorded deed.” A purchaser has constructive notice based on documents found within the record and “is presumed to have investigated the title, . . . to have examined every deed or instrument properly recorded, and to have known every fact disclosed or to which an inquiry suggested by the record would have led.” A purchaser has inquiry notice where he “had knowledge of facts or circumstances that would have [led] a reasonably prudent person to make inquiry into a possible defect of title.” Finally, courts have also warned that “[a] bona fide purchaser or encumbrancer for value is protected in its title unless the deed is void and conveys no title . . . or it had previous notice of the alleged fraud.”

In the case of subsequent purchasers of MERS-mortgaged properties, regardless of whether such purchasers could win the “race to the courthouse,” it is unlikely that they could meet the requirements of a bona fide purchaser. First, the subsequent purchaser does not have actual notice of the existence of the note owner’s unrecorded claim to the

---

165 See 4 AMERICAN LAW OF PROPERTY, supra note 47, §§ 17.10-17.11. While a purchase “for value” is required, this note assumes that all payments made by subsequent purchasers will satisfy this additional requirement. For a discussion of the “for value” requirement, see 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.09(c).
166 N.Y. REAL PROP. LAW § 291 (McKinney 2011).
170 Eighmie, 41 N.Y.S. at 1015-16.
173 Id.
However, the subsequent purchaser may have constructive notice. In particular, a thorough investigation of the record would have uncovered the fact that MERS recorded a discharge of mortgage in favor of the borrower, which “so state[d]” that the mortgage had not been assigned “of record.” A thorough investigation of the record also would have revealed that the note was never produced to accompany MERS’s discharge, and that the note’s owner never executed a separate discharge with respect to the note. These observations would have been sufficient to make the subsequent purchaser suspicious as to the state of his title. Moreover, the subsequent purchaser was aware of facts that might put him on inquiry as to the state of his title—in particular, that MERS was named on the mortgage. In light of the recent notoriety MERS has received for its documentation problems, this fact might cause a “reasonably prudent person to make inquiry into a possible defect of title.” Finally, it is also possible that, in some circumstances, courts will refuse to protect the subsequent purchaser because his deed is void by reason of fraud or forgery. This may be particularly relevant in MERS’s case if there were an allegation of forgery by a robo-signer. Accordingly, a subsequent purchaser of MERS-mortgaged property is unlikely to warrant the protection of a race-notice recording statute under these circumstances, regardless of whether he records his interest first.

175 See Eighmie, 41 N.Y.S. at 1015-16.
176 N.Y. REAL PROP. LAW § 321(3) (McKinney 2011); see MERSCORP, Inc. v. Romaine, 861 N.E.2d 81, 84-85 (N.Y. 2006).
177 See Curtis v. Moore, 46 N.E. 168, 169 (N.Y. 1897) (“[Defendant] was not a bona fide purchaser . . . because the record of the mortgage was notice to him that the mortgage was outstanding and unsatisfied, and it was no concern of his who happened to be the owner at the time. In dealing with the property on the assumption that [the record owner] still owned the mortgage, he acted at his peril, and assumed the risk that [the record owner] might have transferred the mortgage to someone else. He was put upon his inquiry, and it was not enough for him to examine the record, and see that no assignment of mortgage appeared thereon, but he should have required a satisfaction piece in due form, or the delivery of the mortgage and the note.”); see also HSBC Mortg. Servs., Inc. v. Alphonso, 874 N.Y.S.2d 131, 133 (App. Div. 2009).
178 See supra notes 4-6, 113 and accompanying text.
181 Robo-signing has captured significant public interest around the country and has also garnered recent attention from prosecutors. See Michael Kraus, Nevada Attorney General Pursuing Criminal Charges Against Robo-signers, TOTAL MORTG. SERVS. (Nov. 18, 2011), http://www.totalmortgage.com/blog/mortgage-rates/nevada-attorney-general-pursuing-criminal-charges-against-robo-signers/14741.
3. Notice Jurisdiction

In notice jurisdictions, a subsequent purchaser need only meet the requirements of a bona fide purchaser for value in order to receive protection. Indeed, an important distinction from a race-notice statute is that, under a pure notice statute, “[o]ne need not record to qualify for . . . protection.” Nevertheless, a subsequent purchaser is unlikely to receive protection from a notice statute under these circumstances. Following a substantially similar analysis to that outlined above, subsequent purchasers of MERS-mortgaged properties are likely to have constructive notice and inquiry notice. Moreover, subsequent purchasers may face similar challenges due to robo-signing forgeries. Accordingly, a notice statute is also unlikely to provide any refuge.

VI. IMPLICATIONS FOR SUBSEQUENT PURCHASERS

Because recording statutes cannot provide subsequent purchasers of MERS-mortgaged properties with adequate assurance to title, subsequent purchasers may suffer significant harms. Moreover, given MERS’s prevalence in today’s real estate market, courts will inevitably address this issue. Agency theory offers the best solution for all parties in resolving this conflict.

A. Where MERS Cannot Properly Discharge the Mortgage

MERS’s ability to properly discharge a borrower’s debt plays a critical role in the real estate market. If courts follow the common law rule and find that a borrower must receive a discharge from the owner of the note—as opposed to a discharge from MERS as owner of the mortgage—then the debt will remain effective and the owner of the note will have

---

182 As discussed in Part II, the following jurisdictions have enacted notice statutes for mortgages: Alabama, Arizona, Connecticut, Delaware, Florida, Iowa, Kansas, Kentucky, Maine, Massachusetts, Missouri, New Hampshire, New Mexico, Oklahoma, Ohio, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, and West Virginia. 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.5, 545 n.63; see also 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(b) n.286.
183 See 4 AMERICAN LAW OF PROPERTY, supra note 47, § 17.11; 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(b).
184 11 THOMPSON ON REAL PROPERTY, supra note 42, § 92.08(b).
185 See supra Part V.C.2.
186 See supra Part V.C.2.
recourse against the borrower’s property.\textsuperscript{188} Indeed, this remains true under New York’s recording statutes.\textsuperscript{189} As we have seen, this may have disastrous effects for subsequent purchasers under each of the different recording statutes.\textsuperscript{190} In race jurisdictions, subsequent purchasers will face risks unless they can beat the note owner to the recorder’s office.\textsuperscript{191} And in the forty-six states\textsuperscript{192} that have adopted some form of notice or race-notice statute, subsequent purchasers face risks that a note owner might seek to enforce his rights against the property on the ground that purchasers had constructive or inquiry notice of the prior unrecorded claim.

This scenario could have devastating consequences for subsequent purchasers and for the real estate market generally. First, subsequent purchasers throughout the country could face both economic loss and physical loss of property as a result of their defective titles.\textsuperscript{193} If a prior unrecorded interest remains unsatisfied, that party might pursue foreclosure against the property to redeem its security interest,\textsuperscript{194} leaving subsequent purchasers with nothing. In essence, they have “purchase[d] [not property but] a lawsuit.”\textsuperscript{195} However, it is likely that subsequent purchasers would have obtained title insurance to protect against the risk of defective title.\textsuperscript{196} If subsequent purchasers have obtained title insurance, the title policy would likely insure “against loss or damage . . . sustained or incurred by the insured by reason of . . . [a]ny defect in or lien or encumbrance on the title.”\textsuperscript{197} This provision includes any liens created by a prior


\textsuperscript{189} Although section 1921 of the New York Real Property Actions and Proceedings Law protects the borrower from claims by a mortgagee who cannot produce the note, it does not appear to protect the borrower from claims by the owner of the note, whose security interest was discharged without his knowledge or assent. N.Y. REAL PROP. ACTS. LAW § 1921(4) (McKinney 2011).

\textsuperscript{190} See supra Part V.C.

\textsuperscript{191} See supra Part V.C.1.

\textsuperscript{192} See supra notes 162, 182.

\textsuperscript{193} See, e.g., Curtis v. Moore, 46 N.E. 168 (N.Y. 1897) (affirming judgment of foreclosure against subsequent purchaser based on purchaser’s inferior interest in the property).

\textsuperscript{194} See, e.g., id.

\textsuperscript{195} See 1 Joyce D. Palomar, supra note 10, § 1.2 (“The basis for making a real property transaction contingent on evidence of clear, marketable title is the consensus in the law that a buyer should not have to purchase a lawsuit.”).

\textsuperscript{196} See id. § 5.5 (“The title insurer . . . assumes the risk of loss to the insured by reason of any defect in or lien or encumbrance on the insured title.”); 11 Thompson on Real Property, supra note 42, § 93.03(a)(2).

\textsuperscript{197} 2 Joyce D. Palomar, supra note 10, app. C2.
mortgage on the property.\textsuperscript{198} Since the subsequent purchaser would suffer a loss when the owner of the prior unrecorded interest initiates foreclosure proceedings,\textsuperscript{199} the title insurer would indemnify this loss and the purchaser would receive remuneration.\textsuperscript{200}

Nevertheless, subsequent purchasers are still removed from their homes and businesses,\textsuperscript{201} and any indemnification they receive may fail to account for the subjective value they derived from the property.\textsuperscript{202} Homeowners “commonly value their own property in significant ways that the market does not recognize.”\textsuperscript{203} These values may include “investments in networks of friends and the development of social capital”\textsuperscript{204} as well as “the search costs of finding shops and services in the new location.”\textsuperscript{205} Perhaps most significant, however, is the homeowner’s personal attachment to the property\textsuperscript{206} and the loss of autonomy he would experience by losing it.\textsuperscript{207} Although the homeowner might receive reimbursement from a title insurer in order to purchase a new home, “the price of a replacement will not restore the status quo,”\textsuperscript{208} and “perhaps no amount of money can do so.”\textsuperscript{209} Therefore, a title insurer’s

\textsuperscript{198} 11 THOMPSON ON REAL PROPERTY, supra note 42, § 93.03(a)(2) (“A mortgage creates a voluntary lien.”).
\textsuperscript{199}  Id. § 93.04(b) (“An actual loss appears most clearly after foreclosure occurs, when a deficiency can subsequently be attributed to a covered superior lien or defect in title to the mortgage.”).
\textsuperscript{200} See 1 JOYCE D. PALOMAR, supra note 10, § 10.8 (“[T]he insurer owes . . . a duty to reimburse for actual losses the insured has incurred because of the title defect.”).
\textsuperscript{201} See id. § 1.2 (noting that “[p]urchasers of real property expect to be able to build businesses, establish homes, or otherwise improve their properties” and do not wish to “invest money, time, and care” into their properties if they will be subject to loss).
\textsuperscript{202} In large part, I owe many of my thoughts on this subject to Professor Brian Lee, whose notion of an “idiosyncratic premium” in eminent domain proceedings explores the problem of compensating property owners for subjective value. Brian Lee, Faculty Workshop at Brooklyn Law School: The Idiosyncratic Premium in Eminent Domain (Sept. 8, 2011).
\textsuperscript{203} John Fee, Eminent Domain and the Sanctity of Home, 81 NOTRE DAME L. REV. 783, 790 (2006); see Christopher Serkin, The Meaning of Value: Assessing Just Compensation for Regulatory Takings, 99 Nw. U. L. REV. 677, 700 (2005) (“Sentimental attachments, or a unique business enterprise, may prevent the market from accurately reflecting the value of property to its owner.”); see also Fee, supra, at 793 (explaining that “some business owners, like homeowners, [also] become personally attached to their business property in ways that the market . . . do[es] not value”).
\textsuperscript{204} Lee Anne Fennell, Taking Eminent Domain Apart, 2004 MICH. ST. L. REV. 957, 964.
\textsuperscript{205} Id. at 963.
\textsuperscript{206} See Margaret Jane Radin, Property and Personhood, 34 STAN. L. REV. 957, 992 (1982) (recognizing “society’s traditional connection between one’s home and one’s sense of autonomy and personhood”).
\textsuperscript{207} Id. at 959.
\textsuperscript{208} Id.
\textsuperscript{209} Id.
reimbursement of the homeowner may prove inadequate in compensating for his subjective loss.\textsuperscript{210}

The title insurer has also suffered an economic loss as a result of the defective title.\textsuperscript{211} Given the title insurer's newfound appreciation for the title risks posed by MERS-mortgaged properties, most title insurers will simply refuse to insure them.\textsuperscript{212} When a title search reveals that the owner of the note has not recorded a discharge, the title insurer will inform the prospective purchaser of the defect and refrain from issuing a policy.\textsuperscript{213} To the extent that MERS and its members cannot produce documentation with respect to the note, the underlying properties will begin to suffer from the problem of clouded title.\textsuperscript{214} Aware of potential defects in the chain of title, title insurers will refuse to issue policies on the properties,\textsuperscript{215} and prospective purchasers will be unwilling to bear the risks of prior claimants.\textsuperscript{216} Accordingly, they will seek to purchase property from another homeowner, and the initial borrower will be unable to alienate his property.\textsuperscript{217} This will have significant adverse consequences for property owners and for the real estate

\textsuperscript{210} See id. ("[I]f a wedding ring is stolen from a jeweler, insurance proceeds can reimburse the jeweler, but if a wedding ring is stolen from a loving wearer, the price of a replacement will not restore the status quo—perhaps no amount of money can do so.").

\textsuperscript{211} See supra notes 197-200 and accompanying text.

\textsuperscript{212} Some title insurers have taken this position in the past as a result of lender document irregularities. See David Streitfeld, Company Stops Insuring Titles in Chase Foreclosures, N.Y. TIMES (Oct. 2, 2010), http://www.nytimes.com/2010/10/03/business/economy/03foreclose.html ("Old Republic National Title Insurance[] told its agents Friday that it would not write policies on foreclosed Chase properties until 'the objectionable issues have been resolved,' according to a memorandum sent out by the firm's underwriting department.").

\textsuperscript{213} See generally 11 THOMPSON ON REAL PROPERTY, supra note 42, § 93.05(b) (surveying the title insurer's "duty to search title and report adverse matters").

\textsuperscript{214} See 65 AM. JUR. 2D Quieting Title § 13 (2011).

\textsuperscript{215} See supra note 212 and accompanying text.

\textsuperscript{216} See 1 JOYCE PALOMAR, supra note 10, § 1.2 ("Purchasers of real property expect to be able to build businesses, establish homes, or otherwise improve their properties without the risk of having to begin again in another location because someone appears with a superior claim to the title. Purchasers want assurance that the title is good before they invest money, time, and care, not damages from the grantor when the title proves to be defective.").

\textsuperscript{217} In this regard, it is worth noting the moral hazard concerns that this situation creates. If the borrower has a firm enough belief in the defunct status of his mortgage note—particularly where the borrower is unable to sell his property—he may be willing to cease payments on his mortgage altogether, betting on the fact that his lender cannot produce the note in order to pursue foreclosure proceedings. See Ketcham, supra note 111, at 36 (suggesting that current uncertainty surrounding MERS is "an opportunity to be embraced"). This particular danger arises in jurisdictions that require the lender to prove ownership of the note prior to foreclosure. See, e.g., Bank of N.Y. v. Silverberg, 926 N.Y.S.2d 532, 537 (App. Div. 2011).
market generally.\textsuperscript{218} If purchasers will not bear the risks of
defective title and title insurers will not write policies on the
affected properties, buyers and sellers will become stuck as their
property loses its alienability. Given that MERS holds an
estimated 65 million mortgages,\textsuperscript{219} or approximately 60 percent
of mortgages nationwide,\textsuperscript{220} this could have a significant impact
on the country’s real estate market.

B. Where MERS Can Properly Discharge the Mortgage as
Agent

However, if courts find that MERS has authority to act
as agent for the original lender and the lender’s assigns,\textsuperscript{221}
many of the problems for subsequent purchasers may disappear.
A central principle of agency law is that a principal is “bound by
the acts of its agent.”\textsuperscript{222} Accordingly, if courts accept MERS’s
authority to act as agent for lenders, those lenders would be
bound by MERS’s discharges to borrowers and would be
prevented from bringing any claims against the property.\textsuperscript{223}
Where MERS has discharged the mortgage but the note owner
fails to receive the proper funds, the remedy is no longer against
the property but instead is against MERS as the note owner’s
fiduciary.\textsuperscript{224} This solution alleviates any issues for subsequent
purchasers since they would now be entitled to rely on MERS’s
discharge\textsuperscript{225} and the note owner’s privity to that transaction.\textsuperscript{226}
As a result, agency theory removes the cloud on title for owners

\begin{flushright}
\begin{footnotesize}
\textsuperscript{218} See 1 Joyce Palomar, supra note 10, § 1.2 (“Stability of land titles is

| \multicolumn{2}{c}{critical not only to individual property owners, but also to society as a whole.”.).}.
\textsuperscript{219} Dennis & Cha, Foreclosure Chaos, supra note 2.
\textsuperscript{220} See Silverberg, 926 N.Y.S.2d At 539.
\textsuperscript{221} See, e.g., Deutsche Bank Nat’l Trust Co. v. Pietranico, 928 N.Y.S.2d 818,

\textsuperscript{222} See supra Parts V.B, V.C.2.
\textsuperscript{223} Cange v. Stotler & Co., 826 F.2d 581, 590 (7th Cir. 1987); see 2A N.Y. Jur.

\textsuperscript{224} See, e.g., Restatement (Third) of Agency § 8.08 (2006) (“Subject to any

\textsuperscript{225} See supra Parts V.B, V.C.2.
\textsuperscript{226} See Restatement (Third) of Agency § 6.02 (2006).

Agency § 291 (2009) (“A principal is liable on contracts entered into on its behalf

by an authorized agent.”); Restatement (Third) of Agency § 6.02 (2006) (“When an

agent acting with actual or apparent authority makes a contract on behalf of an

unidentified principal, . . . the principal and the third party are parties to the

contact.”).

\textsuperscript{227} See 3 Am. Jur. 2d Agency § 262 (2002) (citations omitted) (“A principal is

bound by the act of its agent if the agent acts within the scope of the agent’s

authority, whether the authority of the agent is actual or real, or apparent. . . . [A]n

agent acting with actual or apparent authority who enters a contract on behalf of a

principal binds the principal but not himself.”).

\textsuperscript{228} See supra note 10, § 1.2 (“Stability of land titles is
critical not only to individual property owners, but also to society as a whole.”.).
\end{footnotesize}
\end{flushright}
of MERS-mortgaged properties and breathes liquidity back into the real estate market.227

CONCLUSION

The MERS system could have dire effects on subsequent purchasers of MERS-mortgaged properties. Given the potential for widespread documentation errors and the fact that courts may refuse MERS’s authority to discharge a mortgage without proof of the note, subsequent purchasers could be at risk of surprise foreclosures against their property—despite their timely mortgage payments. This would have severe effects on the real estate market and on homeowners’ ability to alienate their property. No title bearing MERS’s name would be certain.

Nevertheless, courts possess the ability to sidestep these problems. If courts recognize MERS’s authority to act as agent for its members, they can achieve more desirable outcomes. In particular, this solution would alleviate the risks that subsequent purchasers otherwise face when acquiring MERS-mortgaged properties. It would also remove the cloud from homeowners’ title. Perhaps most importantly, it would hold MERS accountable for its documentation errors and force the company to internalize the cost of its own mistakes.

Joshua J. Card†

227 See supra Part VI.A.
† J.D. Candidate, Brooklyn Law School, 2013; B.A., Amherst College, 2008. I would like to thank Professor Christopher Serkin, David De Gregorio, and the members of the Brooklyn Law Review, for their thoughtful insights and guidance. I would also like to thank my family and Jaclyn DeMais for their love and support.
“The Power to Tax Involves the Power to Destroy”

HOW AVANT-GARDE ART OUTSTRIPS THE IMAGINATION OF REGULATORS, AND WHY A JUDICIAL RUBRIC CAN SAVE IT

INTRODUCTION

Had they been there, Dan Flavin and Bill Viola might have felt as naked as the fabled emperor when the customs officials in London declared their artwork, carefully packed in crates for international shipping, was just not art. Not art? No—only light fixtures and DVD players. Dan Flavin, hailed as the “conjurer of light and lyrical poet of Minimal Art,” and considered part of the inner core of Minimalism, passed away in 1996 after decades of making neutral, geometric works with fluorescent light tubes. Seminal video artist Bill Viola today continues to make haunting, existential video works exploring a central theme—birth and death. In 2006, United Kingdom customs (Her Majesty’s Revenues & Customs or H.M. Revenues & Customs) handed the London art gallery, Haunch of Venison, a £36,000 bill when the gallery imported these two works from the United States; Customs reasoned that the works, when disassembled, fail to qualify for the discounted import tax normally given to “works of art.” So disassembled, the works amount to little more than electronic parts—light bulbs, various lighting fixtures, projectors, and their accoutrement.

1 McCulloch v. Maryland, 17 U.S. 316, 431 (1819).
2 Manfred Schneckenburger, Sculpture, in ART OF THE 20TH CENTURY, VOL. II, at 528 (Germany, 1998).
4 Christiane Fricke, New Media, in ART OF THE 20TH CENTURY, supra note 2, at 615.

1665
Under U.K. law, if classified as sculpture (a subcategory of the general category of works of art), the works would only be subject to an import value-added tax (VAT) of 5 percent and would be completely exempt from customs duty.\(^7\) When U.K. Customs declared that the parts of these artworks did not qualify as artworks, or more specifically sculptures, the importers were subjected to the standard import VAT rate\(^8\) (17.5 percent at the time,\(^9\) and 20 percent as of January 1, 2011\(^10\)).

In 2008, the gallery brought a lawsuit in the London VAT and Customs Tribunal (Tribunal) challenging the customs bill.\(^11\) At this proceeding, favorable U.K. and European Union precedent, statements from U.S. tax officials, and testimony from gallerists and critics convinced the court that the works should be classified as sculptures, bringing them within the VAT reduction and duty exemption.\(^12\) The Tribunal held that it would be “absurd” to classify the components of the works as something different from what they would be—works of art—when they were fully assembled.\(^13\) But, in a 2010 reversal that shocked the art world, the European Commission (Commission) issued a regulation that reversed the 2008 decision of the Tribunal and reaffirmed the original customs determination.\(^14\) According to the Commission, the light and video works of Flavin and Viola were no more than the sum of their non-art

---


\(^8\) Valentin & McClean, supra note 6.


\(^12\) Valentin & McClean, supra note 6.

\(^13\) Haunch of Venison Partners Ltd., 2008 WL 5326820, at [49]; see also Valentin & McClean, supra note 6.

parts.\textsuperscript{15} Not only was this regulation—passed “behind closed doors . . . . [and] contrary to the jurisprudence of the European Court of Justice”—highly suspect, but it suggests that artists’ right to even call their work art is subject to the prevailing opinions and mores of lawmakers.\textsuperscript{16} Additionally, the regulation did not result from any appeal from the defendant, H.M. Revenues & Customs. The Commission acted on its own,\textsuperscript{17} and contrary to court rulings in two separate European Community member states that held that these and similar works should be classified as art.\textsuperscript{18} Lawmakers with no vested interest in the litigation passed this objectionable law. The implications following from this regulation have riled the art world. Far more radical contemporary artworks could face even greater difficulty gaining widespread recognition as artworks. Adding to the mystery, the Commission did not attempt to make clear its overall policy or reasoning behind the regulation. This note argues that legislating what does and does not qualify as an “artwork” without any objective or specific rubric goes against legal precedent and is dangerous to the livelihood of artists.\textsuperscript{19} Implementing a universal standard for the assessment of “art” as a formal classification would not be as difficult as it might seem: U.S. copyright standards already possess the necessary structure,\textsuperscript{20} and the European recognition of artists’ moral rights is mirrored in these standards.\textsuperscript{21}

This note will begin in Part I with brief overviews of Minimalist Art and Conceptual Art, paying particular attention to the public reception of—and reactions to—shifting trends in artworks over time and geography. Part II will give a brief explanation of the legislative systems at work in the European

\textsuperscript{15} Kennedy, \textit{supra} note 10.


\textsuperscript{17} \textit{Id.}; Valentin, \textit{Not Art}, \textit{supra} note 9.

\textsuperscript{18} Valentin, \textit{Not Art}, \textit{supra} note 9.

\textsuperscript{19} For an in-depth treatment of art and judicial subjectivity through the lens of various aesthetic theories, see Christine Haight Farley, \textit{Judging Art}, 79 TUL. L. REV. 805 (2005). Farley suggests that judges and courts could rely more directly on these various aesthetic theories in developing their art-judging jurisprudence. \textit{Id.} at 808. Aesthetic philosophies filter into opinions subconsciously, even sometimes reluctantly, according to Farley. \textit{Id.} at 845. Being more upfront about the use of these theories could clarify courts’ rulings, and the “rich discourse” could give courts more support. \textit{Id.} at 809.

\textsuperscript{20} For a discussion of “recognized stature” under U.S. copyright law, see infra notes 262-80 and accompanying text.

\textsuperscript{21} For a discussion of “droit morale” under the Berne Convention, see infra notes 245-46 and accompanying text.
Community (EC), as well as an introduction to the European value-added tax system. Part III of this note will discuss various instances of courts, both in the United States and abroad, attempting to navigate the intersection of artwork and customs duties and taxation. Part IV will explore various approaches to protection for conceptual and visual artworks, giving special attention to problems encountered by the more difficult cases. That part will conclude with a suggested method for evaluation and classification of artworks that can be applied by courts and legislators, domestically and abroad, that leaves intact both the artists’ intentions and their artworks’ integrity. This note will conclude with a brief discussion of how similar VAT or flat-tax systems implemented in the United States could lead to comparable difficulties in U.S. courts and legislatures.

I. THE ART BACKDROP

New art, and particularly conceptual art, has a history of challenging the public with works that have not yet been received as art. Legislators and customs officials are not typically on the cutting edge of the art world, and they can be as hostile as the general public to new works by avant-garde artists. Flavin’s and Viola’s works, though hardly new artforms, represent conceptual methods in art that fit into a rich history of progressive upstarts challenging the established norms of the art world. The criticisms of many late twentieth and early twenty-first century works of art—the Damien Hirst vitrines, the Allan Kaprow “happenings,” the Chris Burden performance works—are numerous, but discord between newer artists, who experiment in form, material, and content, and the canon of art history, is almost a tradition itself.

For purposes of this note, I will be using EC and EU interchangeably—EC when referring to cases where the EC is specifically mentioned, and EU when referring to the member states. The EC was absorbed into the EU when it was created, but current case law and legislation still refers to the EC without caveat. 

See [source]

22 See William Fleming, Arts & Ideas 583 (6th ed. 1995) (describing twentieth century modern artists as “command[ing] attention” by “intend[ing] to delight or irritate, to arouse or denounce, to exhort or castigate, to surprise or excite, to soothe or shock”).
A. Flux Is Constant in the Art World: A Brief History

Take the seventeenth century turn from the late Renaissance’s “bold humanistic thinking” into the “bitter self-reproach” of the Counter-Reformation. Caravaggio and others imbued this baroque-period work with a dark humanism that was not always well received. The religious populace favored a “more conventional elegance and illusionism . . . .” Even the nineteenth century romantic Delacroix’s color-intensive, emotional, and macabre works were once the subject of “storms of protest and abuse[,]” though they now rest comfortably among other works of the movement. These “rebels” draw attention for the blatant and unapologetic ways in which they push the edge of the previous movement into the infancy of the next. Manet’s Realist paintings “drew all manner of critical vituperation from the press and public.” Academics of the day “held [Raphael] to be the perfect painter” even though all agreed the “ugliness” of realism should be put aside for a more uplifting style of art. Rodin had trouble as well, suffering accusations that his sculpture was “[t]oo lifelike”—or worse— “[t]oo good.” Even when Impressionism was first introduced it was “a term of critical derision.”

Twentieth century modernists often intended for their works to be overt challenges to the form of art de mode, and “[j]udging from the reactions to Picasso’s early exhibits [and Stieglitz’s famous Armory Show] . . . some artists succeeded beyond their wildest expectations.” Artists of the early twentieth century had to fight hard for acceptance, holding “street demonstrations against museums and art galleries that refused the modernists recognition.” The protests subsided halfway through the century, and many of “[t]hese artists [are now considered] old masters of modern art.”

27 FLEMING, supra note 23, at 377.
28 Id.
29 Id. at 383-85.
30 Id. at 385.
31 Id. at 516-17.
32 Id. at 551.
33 Id. at 553.
34 Id. at 554.
35 Id. at 560.
36 Id. at 583, 630.
37 Id. at 585.
38 Id. at 583, 630.
Even the most open modernist mind was challenged by the cubists’ and futurists’ abstractions, but the Dadaists’ and Surrealists’ reactions to World War I pushed it even further. The Dadaists and Surrealists both advocated abandoning the prevailing styles of art. The constant state of flux of the early twentieth century continued well into the middle of the century, and the technological innovations of the greater society continued to have a profound effect on the arts. Facing alienation from a society deeply affected by war, modernists in New York developed the abstract expressionist “school,” expanding on the breakthroughs of the cubists and futurists. And even these Abstract Expressionists found acceptance difficult at first, with “some of the avant-garde commercial galleries hesitat[ing] to accept their paintings for exhibition.” The action painting of Jackson Pollock and the color-field painting of Mark Rothko and Barnett Newman eventually put some of this resistance to rest.

When the Minimalists emerged later in the century, the hard-fought emotion and humanism was gone from their works, traded instead for “primary structures’ [and] basic geometric volumes . . . .” What most distinguished minimalism was the increased importance of the works’ physical or environmental context—not just hanging on walls or atop pedestals, but “rest[ing] on the floor . . . or occupy[ing] a whole room.” Another shift in focus characteristic of minimalism was the presentation of “art items that are indistinguishable from the raw material they were made from.” There is minimal differentiation between the artwork and the non-art materials used in its creation. Duchamp and his “readymades”—vacuum cleaners, bicycle wheels, urinals—were the inspiration for this shift. The problem for

30 Id. at 605-07.
31 Id. at 605-09.
32 Id. at 629-30.
33 Id. at 632-33.
34 Id. at 635.
35 Id. at 635-40.
36 Id. at 645 (citation omitted).
37 Id. at 647.
38 Id.
39 Id.
minimalists was that viewers would often be puzzled by a lack “of the things [they] tend to expect to find within artworks.”  

B. The Artists, Themselves

Dan Flavin was an American minimalist who devoted the latter part of his career to, and garnered a great deal of international attention from, a series of art installations crafted primarily from fluorescent lighting fixtures one could buy in any hardware store. The Museum of Modern Art and the Guggenheim and Whitney museums hold a number of Flavin’s works in their permanent collections. His work occupies an important place in the conceptual-art movement, and his installations are generally abstract and untitled except for dedications to various people. This note concerns a work by Flavin entitled *Six alternating cool white/warm white fluorescent lights vertical and centred (1973)*. Flavin’s minimalist style is very context-dependent, because his works are generally large enough to fill an entire room.

Bill Viola is an American conceptual and experimental artist whose work has been dedicated primarily to installations involving video. Viola’s work is also held by museums internationally, and he is considered a major figure in contemporary art, particularly for his experimental and moving work with video. This note concerns a work by Viola entitled *Hall of Whispers*. While Flavin and Viola challenge “traditional sculpture,” they are far from rebels in the worlds of minimalist art and video art.

One of the ironies of this litigation was that the materials Flavin and Viola used in their sculptures, while not very “traditional,” were hardly the most unusual. For over a

---

55 Valentin, *Not Art*, supra note 9; see also Haunch of Venison Partners Ltd., 2008 WL 5326820, at [2].
56 For purposes of this note, “traditional sculpture” is defined as three-dimensional forms cast or carved of wood, metal, or clay, often in human form.
century artistic rebels have been “expand[ing] the vocabulary of visual media . . . . [T]he twentieth century produced a succession of artistic innovators, each in turn pushing expressive norms to their limits and beyond.” 57 Examples of other non-art materials used by minimalists and other artists are raw wood, Legos, bricks, aluminum, “rope, cigarette butts, newsprint, taxidermic animals, latex,” 58 crockery,60 bodily fluids,60 pieces of candy,61 chewing gum,62 human hair,63 chocolate,64 soup cans,65 tiger sharks,66 money,67 and post-it

These materials appear with increasing regularity on the tags next to artworks in major museums. These “atypical” materials are undeniably a part of “our cultural heritage.”

II. INTRODUCTION TO THE VALUE-ADDED TAX SYSTEM

The fees imposed by U.K. Customs on the Haunch of Venison gallery included “value added tax.” Value-added tax (VAT) is a form of consumption tax which is levied against goods and services both “within the territory of [an EC member state] . . . [and through] the importation of goods” into a member state. The Sixth Council Directive implemented the present VAT system in 1977. This directive “aim[ed] at a further harmonization of the various national laws” that developed from the first five directives. The implementation of these first five directives was intended to establish “a common market [among the EC member states] . . . whose characteristics are similar to those of a domestic market.” A “harmonized” system of tariffs (TARIC) developed out of these directives, which established a “uniform basis for assessment” in a “common system of value-added tax.” Each member state was required to adopt a version of the legislation into its individual legal system. In the U.K., the “common market” system was adopted in 1973 after implementation of the Second Directive.

69 Davis, supra note 57, at 218.
70 Id.
71 Valentin & McClean, supra note 6.
74 TERRA & KAJUS, supra note 72, at 8.
75 Id. at 5.
77 The two primary ways in which the European Commission passes legislation are through directives and regulations. EUROPEAN UNION LAW AFTER MAASTRICHT: A PRACTICAL GUIDE FOR LAWYERS OUTSIDE THE COMMON MARKET 5 (Ralph H. Folsom et al., eds., 1996) [hereinafter MAASTRICHT]; see also art. 249 (189) of the EC Treaty. A directive sets a policy that each member state must then adopt in a manner appropriate for its individual system. MAASTRICHT, supra, at 5. The member states are not always required to take any action in response to the issuance of a directive. Id. “[D]irective[s] are ‘binding as to the result to be achieved’ but ‘leave[] to the national authorities the choice of form and methods.’” Id. (quoting art. 189 of the Treaty of Rome).

Regulations differ from directives in that they are immediately binding on all member states upon their publication in the Official Journal of the European Community (Official Journal or O.J.) and “must be complied with fully by those to whom they are addressed.” EUROPEAN PARLIAMENT FACT SHEETS, EUR. PARLIAMENT (Oct. 16, 2000),
The Seventh VAT Directive was implemented in 1994 upon the recognition that, in regards to works of art, the different systems of the various member states were causing “distortion of competition and deflection of trade . . . .” This directive established “[s]pecial arrangements applicable to . . . works of art.” The Seventh Directive added language to the previous version that lowered the taxable amount on works of art to “a fraction of the amount” of tax applied to all other imported goods. This directive had the effect of raising the import tax on works of art in the U.K. from 2.5 percent to 5 percent. Of course, U.K. art dealers were upset by this increase in import costs. They worried that this tax increase would “undermine [their] competitiveness with the New York market.”

The Commission, however, felt strongly that harmonizing EC “artists’ resale rights [would] . . . put an end to various kinds of discrimination and inequality . . . which currently exist across the Community for visual artists and . . . allow them to achieve parity with other categories of creative artists . . . who could expect ongoing copyright royalties.”

“Works of art” were defined in Article 26a, added in the Seventh Directive, and this new article fit within the system of “combined nomenclature” (CN) instituted by a 1987 amendment to the Sixth Directive. This new article defined works of art, in

---

79 Id. The Seventh Directive was published on February 14, 1994 as an amendment to the Sixth Directive. Id. For a brief discussion of the Sixth VAT Directive, see DAVID W. WILLIAMS, EC TAX LAW 82-83, 852-86 (1998).
81 Id.
82 Id. (art. 1(b)). Generally the taxable amount was either the price paid by the importer or the open-market value of the goods. See Sixth Council Directive, supra note 73, at 7 (art. 11(B)(1)(a-b)).
83 Maren Günther, Written Question to the Commission E-0551/98, 1998 O.J. (C 223) 93.
84 Id.
85 Mr. Monti, Answer Given on Behalf of the Commission, 1998 O.J. (C 223) 94.
87 Council Regulation 2658/87/EEC, 1987 O.J. (L 256) 1-5. Regulation 2658 created a system of nomenclature and classification code numbers in accordance with previously adopted regulations in order to update and assume the management of the TARIC system (a system of integrated tariffs of the European Communities). Council
particular sculptures, as “original sculptures and statuary, in any material, provided that they are executed entirely by the artist.” The structure of this system of combined nomenclature allowed for expansion of the tariff classifications at the member state level and regularized the system. This 1987 regulation also “laid down the general rules . . . [of] interpretation” for this harmonized tariff system. The regulation affecting Flavin and Viola (2010 regulation) was adopted as further clarification of the 1987 regulation.

Sellers of goods and services in the United States will only encounter the VAT when “importing goods [or] services into the

Regulation 2658, supra, at 1-2. “The combined nomenclature shall comprise: (a) the harmonized system nomenclature; (b) Community subdivisions to that nomenclature, referred to as ‘CN subheadings’ in those cases where a corresponding rate of duty is specified; and (c) preliminary provisions, additional section or chapter notes and footnotes relating to CN subheadings.” Id. at 2 (art. 1(2)).

Each CN was assigned an eight-digit code, the first six of which indicate the heading and subheading of the harmonized system nomenclature, with the last two indicating the CN subheading if one is present. Id. art. 3; see also Integrated Tariff of the European Communities (TARIC), 1993 O.J. (C 143) 8; Commission Regulation 2793/86/EEC, 1986 O.J. (L 202) (setting out the numeric codes to be used in the unified system). Absent a subheading the last two digits would be “00.” Council Regulation 2658, supra, art. 3(1)(b). Additional digits would be added onto this code for purposes of member state statistical subdivisions and any additional community subdivisions. Id. art. 3(2)-(4).


Council Regulation 2658, supra note 87, at 1.

Commission Regulation 731, supra note 14, at 2.

The Flavin/Viola amendments were not the only ones made by the Commission at that time. See Commission Regulation 732, 2010 O.J. (L 214) 4 (EU). Many other amendments have been made over time.
EU [member states]” from the United States. This import tax is generally paid by the buyer of the goods or services, and does not apply to goods with only a de minimis value. The tax is “chargeable at the time when [the] goods enter the [member state].” The standard VAT rate among EU member states ranges from 15 to 25 percent.

III. THE INTERSECTION OF ART AND THE LAW

When courts approach the problem of classification of artworks for tax and import duty purposes, deference is generally correctly given to the vicissitudes of the art world. Courts, both in the United States and abroad, have customarily declined to act as arbiters of taste. What may not have been considered “art” ten, twenty, or fifty years ago might now be recognized as valuable by experts and the art world. There is a continued risk, as evidenced by the 2010 regulation, that vanguard contemporary artworks may still be denied their status as art. Neither the U.S. judiciary nor courts and legislative bodies abroad should permit this unfortunate and shortsighted result. Part III explores the U.K. approach to this problem, both in the context of the Flavin/Viola matter and through earlier EU jurisprudence. As to the domestic approach, the United States has had fewer occasions to address this question, though one case in particular concerned the importation of an abstract sculpture that was initially charged

---

92 Tax Advisors Planning System, Title 43, 43:13:01(E) Imports Into the EU (RIA 2011) (referencing Sixth EC VAT Directive, art. 70).
93 Id.
95 Tax Advisors Planning System, Title 43, 43:13:01(C) Value Added Tax (VAT) Rates (RIA 2011). As of Directive No. 77/388, the minimum tax must be at least 15 percent in each member state. Sixth VAT Directive, supra note 94. The minimum rate was increased as of January 1, 2011. This caused much consternation among those affected by the undiscoun ted VAT rate on imports, as the minimum then increased to 20 percent. See Accounting for VAT When the Standard Rate of VAT Returned to 17.5 Per Cent, HM REVENUE & CUSTOMS, http://www.hmrc.gov.uk/vat/forms-rates/rates/ rate-changes.htm (last visited Jan. 19, 2012); Increase in the Standard Rate of VAT to 20 Per Cent, HM REVENUE & CUSTOMS, http://www.hmrc.gov.uk/vat/forms-rates/rates/rate-increase.htm (last visited Jan. 19, 2012).
97 See, e.g., Bleistein, 188 U.S. at 251; Haunch of Venison Partners Ltd., 2008 WL 5326820 at [50]-[51].
import tax on the basis of its component material. The U.S. court, just as the U.K. Tribunal did with Flavin and Viola, found the work product to indeed be art. There are still other methods by which courts and legislatures can approach this issue, and those are suggested and discussed in Part IV.

A. Haunch of Venison Partners Ltd. v. H.M. Revenue & Customs

The heart of this inquiry is the dispute over customs bills charged to London’s Haunch of Venison gallery when it imported the contemporary sculptures of Dan Flavin and Bill Viola. Both works were imported in a disassembled state and subject to a tariff rate normally charged to goods and component parts intended for further sale or assembly. In other words, the customs officials assessed the light tubes and video equipment as light tubes and video equipment, not as disassembled artworks packed in boxes. Ironically, they assessed the tariff on the estimated value of the assembled pieces—in their final form as artworks. Thus the customs officials acknowledged the imports were valuable as artwork, but simultaneously denied them the reduced tariff rate.

1. 2006–2008: Trouble with U.K. Customs

The Haunch of Venison gallery appealed the customs bill because Flavin’s and Viola’s sculptures were not assessed under the discounted rate normally reserved for works of art. When the artworks were first imported, customs officials declared that the works were subject to the full VAT (then 17.5 percent). The gallery brought a lawsuit against HM Revenue & Customs (HMRC) protesting this classification before the London VAT and Duties Tribunal in 2008. The gallery sought relief from the tariff assessment by arguing that the artworks—even in their disassembled state—qualified for the discounted rate.

98 See Brancusi, 54 Treas. Dec. at 428-29; see also infra Part III.B.
99 See Brancusi, 54 Treas. Dec. at 431; see also infra Part III.B.
100 Haunch of Venison Partners Ltd., 2008 WL 5326820 at [18].
102 See sources cited supra note 101.
103 Haunch of Venison Partners Ltd., 2008 WL 5326820 at [18].
104 Tarsis, supra note 101.
105 See generally Haunch of Venison Partners Ltd., 2008 WL 5326820.
reserved for artworks. The trial lasted four days, during which the Tribunal visited the Tate galleries to view similar Flavin and Viola works. The Tribunal also heard testimony from “[s]everal high profile witnesses” as to the correctness of the customs officials’ classifications. HMRC’s central concern was that treating imported goods that do not appear to be traditional artworks as art would open the door to “any importer . . . declar[ing] any goods to be works of art and thereby circumvent[ing] the positive rates of duty.” HMRC argued that the particular components in this case should be considered not as unassembled artworks but instead according to their “objective characteristics,” namely as light bulbs and DVD players. HMRC also argued that sculptures could not be only two dimensional. This related particularly to the Viola works, projection[s] emanating from the screen which is itself a flat object and [about which HMRC contended] . . . it is simply incorrect as a matter of fact to consider the mechanism by which that image is realized as being part of a sculpture even if that mechanism does have a three dimensional form.

The Tribunal dismissed both of these concerns. These pieces qualified as sculpture by virtue of “all the components [having been chosen] deliberately and as part of the artistic process with a view to achieving [the] artistic intention with the greatest effect.” The Tribunal did not expressly require engagement with the third dimension. It found instead that these components included the structure by which the projectors and screens would be hung, the materials and equipment selected for the projectors and screens, and the specific spatial arrangement of these elements for the actual installation. The combination of these elements satisfied the Tribunal’s definition of sculpture.

---

106 Id. at [4].
107 Valentin, Not Art, supra note 9.
108 Id.
109 Id., 2008 WL 5326820, at [39].
110 Id. at [36], [40] (citing Case C-35/93, Develop Dr Eisbein GmbH & Co. v. Hauptzollamt Stuttgart-West, 1994 E.C.R. I-2655).
111 Id.
112 Id. at [35].
113 Id.
114 Id. at [48].
115 Id. at [49]-[51].
116 Id.
The Tribunal also dismissed HMRC’s more central concern—“that importers might declare just anything as works of art” to evade duty rates. The Tribunal declared it absurd to classify any of these works, in their unassembled form, as components ignoring the fact that the components together make a work of art. . . . It stretches the objective characteristics principle too far to say that a work of art is no longer a work of art if it is dismantled for transportation . . . .

The Tribunal also noted that “where there is doubt about a classification . . . , preference should be given to one of the Chapter 97 headings over those of any other Chapter.” Having negated both of HMRC’s main arguments, the Tribunal concluded that these works should be classified as sculptures and are thereby subject to the lesser VAT rate. Appeals were allowed by the Tribunal, but as of this writing HMRC has not appealed.

The Tribunal relied upon Develop Dr Eisbein GmbH & Co. v. Hauptzollamt Stuttgart-West, which addressed tariff classifications for articles imported in an unassembled or disassembled state. The parts at issue in Eisbein GmbH were photocopier parts and accessories. The classifications, although made by German customs officials, ultimately utilized the same EU harmonized tariff system that the U.K. customs officials applied to Flavin and Viola. Imports of fully-assembled apparatuses into Germany owe an additional antidumping duty, and this duty was imposed in Eisbein GmbH. The importer appealed, arguing that the parts were not fully-assembled photocopiers and thus not subject to the duty. The importer relied on the Explanatory Notes to the antidumping rule, which included simply-assembled articles (i.e., apparatuses) in the “fully assembled” category. These

---

117 Id. at [50].
118 Id. at [49]-[51].
119 Id. at [44] (discussing Case 155/84, Onnasch v. Hauptzollamt Berlin-Packhof, 1987 E.C.R. 1449). For a discussion of the various classification “Chapters,” see supra notes 87-88 and accompanying text; see also infra note 149.
120 Id. at [51]. Such a classification would also exempt the works from a separate customs duty of 3.7 percent, which was charged to the gallery. Valentin, European Definition, supra note 142.
122 Id.
123 Id. at [1].
124 Id. at [4] (discussing the antidumping duty under Regulation 2640/86/EEC, 1986 O.J. (L 239) 5).
125 Id. at [51].
126 Id.
were described as “articles the components of which are to be assembled” using only simple instructions.\textsuperscript{127} The importer argued that the components of a photocopier require highly specialized staff for assembly, and that therefore assembly cannot be accomplished with simple instructions.\textsuperscript{128} Thus, the photocopiers were neither “unassembled” requiring simple assembly nor “fully assembled,” the antidumping rule did not apply, and the parts should be treated solely as components.\textsuperscript{129} Customs officials argued that the requirement of highly specialized staff did not take these articles out of the “simple assembly” category.\textsuperscript{130} They argued that this only occurs when “changes to . . . the part in question [would be required] in the course of the production process.”\textsuperscript{131}

The court disagreed with the importer, asserting instead that “highly qualified specialized staff” does not mean the assembly was not “simple,”\textsuperscript{132} and further, that neither “the assembly technique [nor] the complexity of the assembly method” should be taken into account when classifying components for customs purposes.\textsuperscript{133} The court defined simple assembly in the negative: “parts [that] have to undergo major processing before assembly . . . [do] not have the relevant essential character” of the final product.\textsuperscript{134} The inverse of that argument is that parts that do not have to undergo major processing do “have the relevant essential character of the final product.”\textsuperscript{135}

Flavin’s and Viola’s installation components may have technical and specific assembly instructions, but they do not require major processing before assembly, as raw materials might. Thus, these component light bulbs, projectors, and screens can be understood to possess the “relevant essential character” necessary for classification under \textit{Eisbein GmbH} as their final forms—sculptures and works of art. This distinction implies that, unless changes to the forms of Flavin’s and Viola’s installation components needed to take place, they should be classified as the full apparatus when imported. The rule in \textit{Eisbein GmbH} makes this even plainer: “[F]or tariff purposes an article presented unassembled or disassembled must be

\textsuperscript{127} \textit{Id.} (internal quotation marks omitted).

\textsuperscript{128} \textit{Id.}

\textsuperscript{129} \textit{Id.} at [6]-[7].

\textsuperscript{130} \textit{Id.} at [8].

\textsuperscript{131} \textit{Id.}

\textsuperscript{132} \textit{Id.} at [17]-[23].

\textsuperscript{133} \textit{Id.} at [19].

\textsuperscript{134} \textit{Id.} at [12].

\textsuperscript{135} \textit{Id.}
regarded as a complete article. No reference is made to the assembly technique which must be applied in order to produce the finished product.” 136 Simple assembly or major processing requirements aside, if disassembled products “must be regarded as a complete article,” 137 it is clear that the Tribunal relied on the holding in Eisbein GmbH in making its decision regarding Flavin’s and Viola’s works.

2. 2008–Today: The Reaction of the European Commission

Within weeks of the Tribunal decision, “the issue [of the Flavin and Viola imports] was on the agenda of the EC Customs Code Committee . . . .” 138 The committee knew that two member states (the U.K. and Holland) “had held that video installations should be classified as sculptures[,]” while other member states “expressed the view that components of video installations should be taxed individually (e.g. as video projectors).” 139 Within one year, and “without apparent further consideration or consultation, the committee decided that a draft regulation [would] be prepared for a future meeting—[one that would] overturn the UK and Dutch National Court decisions.” 140

This proposal was before the Customs Code Committee by June 2010; 141 acting under the 1987 regulation it “was adopted at the [August] meeting and supported by the UK.” 142 The new regulation’s sole concern was the classification of these specific artworks by Flavin and Viola. Although the 1987 regulation allows for changes to be made to the harmonized system that relate to “changes in requirements relating to statistics or commercial policy,” 143 there are no claims or references made in the text of the 2010 regulation to any such changes in “statistics or commercial policy” 144 driving this amendment. As a regulation, this change to the tariff classification system CN was effective immediately upon its

136 Id.
137 Id.
138 Valentin, Not Art, supra note 9. This was discovered through a Freedom of Information request made by Pierre Valentin, the attorney representing the gallery in the 2008 Tribunal matter. Id.
139 Id.
140 Id. (internal quotation marks omitted).
141 Id.
142 Id.; Commission Regulation 731, supra note 14, at 2.
143 Council Regulation 2658, supra note 87, at 4 (art. 9(b)).
144 Id.
publication in the Official Journal, which made it binding on all member states.\textsuperscript{145}

The artworks at the center of the 2010 regulation were described objectively, without any reference to the artists’ names or the names of the works.\textsuperscript{146} Under the heading “Description of the goods,” Bill Viola’s video work was listed as:

A video-sound installation consisting essentially of the following components:

- 10 video reproducing apparatus of the digital versatile disc (DVD) type,
- 10 projectors using matrix liquid crystal display (LCD) technology, of a kind also capable of displaying digital information generated by an automatic data-processing machine,
- 10 single self-powered loudspeakers, mounted in their enclosures, and
- 20 digital versatile discs (DVDs) containing recorded works of “modern art” in the form of images accompanied by sound.

The appearance of the video reproducing apparatus, the projectors and the loudspeakers has been modified by an artist with a view to appear as a work of “modern art” without altering their function.\textsuperscript{147}

Dan Flavin’s sculpture was described as:

A so-called “light installation” consisting of six circular fluorescent lighting tubes and six lighting fittings of plastics.

It has been designed by an artist and operates in accordance with instructions provided by the artist. It is intended to be displayed in a gallery, fixed to the wall.

The fittings are separate from each other and are intended to be mounted vertically.

The tubes are to be placed into the fittings, providing two alternating shades of white.\textsuperscript{148}

Each of these two descriptions was assigned a classification code: Viola’s work was assigned to TARIC Chapter 85 (Electrical Machinery and Equipment) and Flavin’s to

\textsuperscript{145} Commission Regulation 731, \textit{supra} note 14, at 2. The regulation actually provided that the regulation was “to enter into force on the 20th day following its publication in the Official Journal.” \textit{Id.} (art. 3) (emphasis omitted). With the date of publication at August 14, 2010, the regulation would be effective as of September 3, 2010.

\textsuperscript{146} Nevertheless, those familiar with the matter would have no trouble making out to what the EC was referring. \textit{Id.} at 3.

\textsuperscript{147} \textit{Id.}

\textsuperscript{148} \textit{Id.}
Chapter 94 (Miscellaneous Manufactured Articles).\textsuperscript{149} The reasons given for the classifications range from the presumptive to the “absurd.”\textsuperscript{150} For Flavin’s work, the Commission claimed that “[c]lassification under [Chapter 97] as a sculpture is excluded, as it is not the installation that constitutes a ‘work of art’ but the result of the operations (the light effect) carried out by it.”\textsuperscript{151} The Commission asserted that Viola’s “video-sound installation is neither composite goods, as it rather consists of individual components, nor goods put up in sets for retail sale . . . . Consequently, the components of [both] installation[s] are to be classified separately.”\textsuperscript{152} The thrust of the Commission’s reasoning can be found in the following section of the regulation:

Classification under [Chapter 97] as a sculpture is excluded, as none of the individual components or the whole installation, \emph{when assembled}, can be considered as a sculpture. The components have been slightly modified by the artists, but these modifications do not alter their preliminary function of goods of Section XVI. It is the content recorded on the DVD which, together with the components of the installation, provides for the “modern art.”\textsuperscript{153}

The concern here is whether these works are classifiable as “sculpture.” Chapter 97 of TARIC only specifies that a sculpture be of “any material.”\textsuperscript{154} Nothing in that chapter attempts to define or give features to any of the categories of art mentioned (sculpture is not the only category).\textsuperscript{155} The Tribunal judges in 2008 had no trouble recognizing these works

\textsuperscript{149} Chapter 85 under Section XVI of TARIC is entitled “Electrical Machinery and Equipment and Parts Thereof; Sound Recorders and Reproducers, Television Image and Sound Recorders and Reproducers, and Parts and Accessories of Such Articles.” Council Regulation 2658, supra note 87, at Annex Sec. XVI; see also TARIC Consultation, EUR. COMM’N: TAX’N & CUSTOMS UNION, http://ec.europa.eu/taxation_customs/dds2/taric/taric_consultation.jsp?Lang=en&Taric=85219000&Area=US&Expand=true&SimDate=20110923#8521900000 (last visited Mar. 23, 2012), Viola’s work was classified under codes 8521 90 00, 8528 69 10, 8518 21 00, and 8523 40 51.

\textsuperscript{150} Council Regulation 2658, supra note 87, at Annex Sec. XX; see also TARIC Consultation, EUR. COMM’N: TAX’N & CUSTOMS UNION, http://ec.europa.eu/taxation_customs/dds2/taric/taric_consultation.jsp?Lang=en&Taric=9405&Area=US&Expant=true&SimDate=20110923#9405000000 (last visited Mar. 3, 2012), Flavin’s work was classified under code 9405 10 28.

\textsuperscript{151} Haunch of Venison Partners Ltd. v. Revenue & Customs Comm’rs, [2008] UKVAT (Customs) C-00266, 2008 WL 5326820, at [49] (Dec. 11, 2008 VAT & Duties Tribunal (London)).

\textsuperscript{152} Commission Regulation 731, supra note 14, at 3.

\textsuperscript{153} Id.

\textsuperscript{154} Id. (emphasis added).

\textsuperscript{155} Council Regulation 2658, supra note 87, at Annex sec. XXI.
in their assembled states as sculptures and artworks.\footnote{156 Haunch of Venison Partners Ltd. v. Revenue & Customs Comm'rs, [2008] UKVAT (Customs) C-00266, 2008 WL 5326820, at [47]-[49] (Dec. 11, 2008 VAT & Duties Tribunal (London)).} The Tribunal also found it a “stretch” to refuse to recognize the components as the equivalent of their finished, assembled form.\footnote{157 Id. at [51].} The dictionary defines sculpture as “the product of the sculptor’s art,”\footnote{158 Sculpture Definition, Merriam-Webster.com, http://www.merriam-webster.com/dictionary/sculpture (last visited Mar. 16, 2011).} as well as “a three dimensional work of art.”\footnote{159 Sculpture Definition, Oxford Eng. Dictionary Online, http://oxforddictionaries.com/definition/sculpture?q=sculpture (last visited Mar. 23, 2012).} While neither definition clarifies the matter, the judicial system has traditionally relied on the art maker and the art community to affirm a work as art.\footnote{160 See infra notes 194-200 and accompanying text.} The members of the Commission, however, applied some other analysis to this same question, though the regulation does not make clear the principles they based their decision upon.\footnote{161 Id. at [51].}

The Commission’s reasoning disregards the U.K. VAT Tribunal’s 2008 decision. Since the component parts of Flavin’s and Viola’s installations are both presented disassembled, the Commission holds, they must be classified by their individual components.\footnote{162 Id. at 3.} But the Commission also declares that, even assembled, these works are not art.\footnote{163 Id.} The Commission believes displaying the images contained on the Viola DVDs is the final step necessary for his work to become sculpture.\footnote{164 Id.} It reasons similarly with Flavin’s work, claiming that the addition of a “light effect” to the arrangement of light tubes creates the art.\footnote{165 Id.} The Commission distinguishes between “art effects” and assembled, non-art components, which it finds necessary but not sufficient to constitute a finished work. If this distinction holds, the “art effect” of, for example, Flavin’s work could never be conveyed: it is a “light effect,” something that could not be shipped or taxed because it is not a material. Even more troubling, the Commission does not follow its own rule when it assesses the nondiscounted VAT rate for the supposed non-art components; it assesses VAT on the value of the finished artwork, which is much higher than the value of industrial
light bulbs and DVD players combined.\textsuperscript{166} The Commission is willing to use the value of the art works as a basis for the tax, but it refuses to grant the components of unassembled works the discounted status granted to art.\textsuperscript{167}

Non-art components are necessary components of certain artworks. “Art effects” require their non-art components, and the Commission acknowledges this critical interplay, though only nominally.\textsuperscript{168} These artists expressly desired the “light effect” and the DVD images to be produced with these screens and light tubes arranged in a certain way, and the Commission also acknowledges this intention.\textsuperscript{169} And while true that these works rely on certain non-material, non-art components, under the Commission’s standard, it is impossible to import any kind of artwork that employs electricity in its final form under the discounted VAT rate for artworks. This standard is too high. If a court accepted the Commission’s standard, any work that uses infrastructure upon its completion—pressurized water for a fountain, electrical plugs for a neon sculpture, wind for chimes—may not be classified as art, but rather a composition of non-art components. A work would only become art in its intended context or ultimate form—when the play button is pressed or the electrical current is live. This is akin to legislating that artworks that are intended for a particular context are “not art” until they are actually placed in that context. Under this standard, a fully assembled Flavin work would cease to be an artwork once the gallery closed for the night and the janitor turned off the lights.\textsuperscript{170} This standard thus requires that an artwork be continually connected to every element of its context in order to remain art.\textsuperscript{171}

\textsuperscript{166} Valentin, \textit{Not Art}, supra note 9.
\textsuperscript{167} \textit{Id}.
\textsuperscript{168} Commission Regulation 731, \textit{supra} note 14, at 3.
\textsuperscript{169} \textit{Id}.
\textsuperscript{170} Pierre Valentin asked a similar question in a series of opinion pieces published by \textit{The Art Newspaper}. See, e.g., Valentin, \textit{European Definition}, \textit{supra} note 16.
\textsuperscript{171} Interestingly, Duchamp’s \textit{Fountain} was never intended to be connected to a plumbing source. His intention was to keep it freestanding and unconnected to any piece of infrastructure, but for its platform. William Camfield, \textit{Marcel Duchamp’s Fountain: Its History and Aesthetics in the Context of 1917}, in \textit{MARCEL DUCHAMP: ARTIST OF THE CENTURY} 78 (Rudolf E. Kuenzli and Francis M. Naumann, eds., 4th prtg. 1996). Under the Commission’s ruling, there would be no “art effects” of Duchamp’s piece because it is free of infrastructure and context; thus it would not be ruled as a protected sculptural artwork.
More significantly, this standard separates the artist’s work product from the art itself. The Commission asserts that the art is found when the artist’s work product is added to a particular infrastructure or context, but of course that final element is generally outside of the artist’s control. Creating the art is not the accomplishment of the gallery owner who turns on the lights! The problem is that this standard creates no account in the law for what “art” is at all. Recall the definition of sculpture under the TARIC classification—“original sculptures and statuary, in any material, provided that they are executed entirely by the artist . . . .” The Commission’s standard, which places the work done by the context and infrastructure ahead of the work done by the artist, is clearly operating outside of the harmonized system.

B. U.S. Analogues in Customs Duties

The United States does not have a consumption tax or VAT system, but there are customs duties imposed on imports. By statute, art works are exempt from these customs duties. Courts will occasionally need to determine if an import is a work of art in order to decide upon the proper customs duty. There are a handful of cases in the United States that address relevant import duties, but they were decided under an earlier iteration of the customs code. This historical precedent is still instructive, however, when evaluating the recent Commission regulation.

One of the most famous of these cases involved a simple sculpture cast from bronze that the customs officials had trouble classifying as a “sculpture.” Interestingly, this rather “sensational lawsuit . . . captured the attention of American public opinion for two years.” This 1928 U.S. Customs Court decision concerning Constantin Brancusi’s Bird in Space sculpture is the U.S. case that most closely parallels the Flavin and Viola matter in both the facts and the court’s reasoning.

---

175 Id.
178 Brancusi, 54 Treas. Dec. at 428. In the case, the sculpture is referred to as Bird in Flight. Id.
Brancusi’s sculpture was subject to an import duty upon its entry into the United States because the sculpture did not exactly resemble a “bird.” Since the sculpture did not resemble a “bird,” the Customs refused to classify it as a “zero-rated work of art.” Customs, refusing to classify the sculpture as such, subjected the “bird” to a 40 percent import duty, the rate generally applied to manufactures of metal. Photographer Edward Steichen had purchased the sculpture; he paid the duty and then went to court to protest the determination. The court eventually agreed with Steichen, and his $240 was returned.

The Customs Court acknowledged that under the Tariff Act of 1922 artworks were entitled to duty-free entry, as opposed to “manufacture[s] of metal” which would be subject to a tariff of “40 per cent ad valorem.” The court wrote into its opinion the entire list of materials and processes included under the Tariff Act’s “art and sculpture” heading. While the Tariff Act does not make this list explicitly exclusive, all of the “traditional” materials of sculpture are included, and the statute did not indicate that any “non-traditional” artwork or materials should be read into the list. Even while the court conceded that “the exercise of rather a vivid imagination” is required to see the Brancusi sculpture as even resembling a bird, it held that the sculpture, crafted from traditional materials, warranted the discounted duty under the Tariff Act.

179 Id. at 429.
180 Kennedy, supra note 14.
181 Id.
182 Id.
183 Id.
184 Id. at 428. “Ad valorem” taxes are taxes “based on the assessed value of [the] real estate or personal property” at issue. Ad Valorem Tax, INVESTOPEDIA, http://www.investopedia.com/terms/a/advaloremtax.asp#axzz1rCwgDuPW (last visited Apr. 5, 2012). A familiar example of an ad valorem tax is municipal property tax.
185 Id. at 428-29. Paragraph 1704 of the Tariff Act of 1922 specified that “sculpture” was to be understood as:

professional productions of sculptors only, whether in round or in relief, in bronze, marble, stone, terra cotta, ivory, wood, or metal, or whether cut, carved, or otherwise wrought by hand from the solid block or mass of marble, stone, or alabaster, or from metal, or cast in bronze or other metal or substance, or from wax or plaster, made as the professional productions of sculptors only . . . .

186 Id. at 428.
187 Id.
188 Id. at 431.
The court applied a three-part test: (1) was this the work of a professional, (2) was this an original work, and (3) was this an article of utility?\textsuperscript{189} Despite contradictory testimony, the court answered the first two questions affirmatively: “There is no question in the mind of the court but that the man who produced the [sculpture] is a professional sculptor, . . . [and w]e also find it is an original production.”\textsuperscript{190} The court would need to answer the third question in the negative for the work to qualify as an artwork under the Act.\textsuperscript{191} Interestingly, the court addressed this question somewhat inversely. Instead of saying whether this piece was an article of utility, the court set out to determine whether it was a “work of art.”\textsuperscript{192} The court acknowledged a strong precedent that would suggest this work be denied this categorization.\textsuperscript{193}

Nevertheless, the court favored a more contemporary approach.\textsuperscript{194} In recognizing the “so-called new school of art” the court accepted an art movement “whose exponents attempt to portray abstract ideas rather than to imitate natural objects.”\textsuperscript{195} While the majority stated that it did not have to agree or be “in sympathy with these newer ideas[,]” the court held that it must recognize the change.\textsuperscript{196} The court stated that “the fact of [the new art movements’] existence and their influence upon the art world as recognized by the courts must be considered.”\textsuperscript{197} The Brancusi object was “beautiful and symmetrical in outline, and while some difficulty might be encountered in associating it with a bird, it is nevertheless pleasing to look at and highly ornamental . . . .”\textsuperscript{198} Thus, the court recognized the metal production as a sculpture and a piece of art, and granted it free entry under the Tariff Act.\textsuperscript{199} The court relied on the principle of “objective acceptance,” which subordinates conflicting subjective responses of the court to expert testimonials. This principle recognizes the shifting trends of the art world, and should have been employed by the Commission when assessing Flavin’s and Viola’s works, though this rubric likely needs additional structure.

\textsuperscript{189} Id. at 430.
\textsuperscript{190} Id. at 429.
\textsuperscript{191} Id. at 428 (citing Paragraph 1704 of the Tariff Act of 1922).
\textsuperscript{192} Id. at 430.
\textsuperscript{193} Id. at 428 (citing Paragraph 1704 of the Tariff Act of 1922).
\textsuperscript{194} Id. at 430.
\textsuperscript{195} Id. at 430.
\textsuperscript{196} Id. at 430-
\textsuperscript{197} Id. at 430-31.
\textsuperscript{198} Id. at 431.
\textsuperscript{199} Id.
At the very least, the Commission should have explained why it did not analyze the works under this “objective acceptance” principle. In Part IV of this note, other rubrics will be discussed that offer more structure but still maintain this objective character.\textsuperscript{200}

Two other U.S. Customs Court cases (one that predated \textit{Brancusi} and another that followed four decades after) addressed questions that relate to the \textit{parts versus wholes}, \textit{components versus finished products} issues. In one case, the intention to use the pieces of the work as a unit determined the imports being considered a whole.\textsuperscript{201} In the other, a marble “sculpture,” which was only decorative and failed to rise to the level of fine art, was found \textit{not} to be an artwork by the court.\textsuperscript{202}

In \textit{Miniature Fashions, Inc. v. United States}, importers appealed a decision classifying patterned cotton shirts and shorts imported from Japan as separates.\textsuperscript{203} This classification subjected the clothing to a rate of 25 percent ad valorem under the Tariff Act of 1930.\textsuperscript{204} Plaintiffs in the suit contended that the sets—designed, manufactured, and sold together as “cabana sets”—were instead “entireties for tariff purposes.”\textsuperscript{205} They argued that these “entireties” should be assessed under the “Other” category under the same section of the Tariff Act, which covered articles of clothing manufactured “wholly or in part . . . of cotton, and not specially provided for [elsewhere in the Act].”\textsuperscript{206} Articles falling under this “Other” category were only subject to a rate of 20 percent ad valorem.\textsuperscript{207} Witnesses for the plaintiffs testified that the sets were “inexpensive . . . [and] have very little . . . value when separated.”\textsuperscript{208} The Customs Court found the articles to be separates because the shirts and shorts would remain functionally the same even upon separation: “[A]lthough these cabana sets were designed . . . for sale together, . . . the functions of the several parts of the set were no different from what they would otherwise have been had the sets not been so coordinated. . . . [T]he shirt continued to be a shirt and the shorts remained shorts.”\textsuperscript{209}

\textsuperscript{200} \textit{See infra} Part IV.
\textsuperscript{201} \textit{Miniature Fashions, Inc. v. United States}, 55 Cust. Ct. 154 (1965).
\textsuperscript{203} \textit{Miniature Fashions}, 55 Cust. Ct. at 155.
\textsuperscript{204} Id.
\textsuperscript{205} Id. (emphasis added).
\textsuperscript{206} Id.
\textsuperscript{207} Id.
\textsuperscript{208} Id.
\textsuperscript{209} Id. at 156.
The importers’ appeal focused on proving that trends in fashion, with an emphasis on dual-purpose apparel manufacturing, were determinative of whether these pieces were “entireties” and not separates. The court admitted “that [while] a changing popular attitude played a role in the conclusions reached in [cases like Brancusi v. United States], we do not believe that [those] decisions actually rested upon this factor.” The court declared that “a designer’s conception of ‘fashion’ or ‘eye appeal’” is not sufficient to overcome previous policy of the court:

If what is imported as a unit is actually . . . two or more individual entities which, even though imported joined or assembled together, nevertheless, retain their individual identities and are not subordinated to the identity of the combination, duties will be imposed upon the individual entities in the combination as though they had been imported separately. Conversely, if there are imported in one importation separate entities, which by their nature are obviously intended to be used as a unit, or to be joined together by mere assembly, and in such use or joining the individual identities of the separate entities are subordinated to the identity of the combined entity, duty will be imposed upon the entity they represent.

Therefore, the question of subordinated identity was the determining factor for the Customs Court, and applying that concept to Flavin and Viola is instructive. The Commission argues that the imposed VAT rate should apply to the functional components of Flavin’s and Viola’s works. But, the Commission also concedes that the works take on a separate, unified identity when assembled and “used as a unit[,]” in the words of the Miniature Fashions court. This qualified concession as to the “unified identity” of the Flavin and Viola works is inconsistent with the Commission’s claim that the value on which tax should be calculated is the value of the unit as an artwork rather than the value of the individual components.

Closer to the realm of sculpture, the question of whether carved marble objects were dutiable as “manufactures of marble” valued on their component material or as works of art arose in United States v. Olivotti. Valued on their component material, the marble boxes and plinths at issue would have

---

210 Id. at 157.
211 Id. at 158.
212 Id. at 160 (emphasis added) (quoting Donalds Ltd. v. United States, 32 Cust. Ct. 310 (1954)).
213 Id.
214 Commission Regulation 731, supra note 14, at 3.
215 See Tarsis, supra note 101.
been subject to a 45 percent ad valorem duty the Tariff Act of 1913; but, if classified as works of art under a different paragraph the applicable duty rate would have been 15 percent ad valorem. The government appealed the decision of the Board of General Appraisers, which found the marble works to fall within the 15 percent rate as works of art. The board had reasoned that since Greek temples were art by virtue of simply being sculpture, these pieces must be art as well. The Court of Customs Appeals was not as easily satisfied, reasoning that one of the pieces being “the work of a sculptor[,] . . . . fashioned from solid marble[,]” and “artistic and beautiful” was insufficient “to constitute a sculpture.”

The court in Olivotti was ultimately unwilling to expand the reach of Paragraph 376 to include the decorative and undeniably sculptural, even beautiful, qualities of the marble pieces in question, holding that neither the marble font nor the marble seats were sculptures or works of fine art dutiable at the lower rate. “That everything artistic and beautiful can not [sic] be classed as fine art was well established in [a Supreme Court decision],” which held that concededly beautiful paintings on glass windows “were representative of the decorative and industrial rather than of the fine arts.”

In Olivotti, the court addressed decorative elements that, when added to functional objects (or precious stone), did not rise to the level of artworks. This is distinguishable from Flavin’s and Viola’s use of nondecorative components that themselves comprise an artwork. Olivotti holds that the decorative elements of a finished commercial product may not elevate it to the classification of “sculpture” or “artwork.” But, the European Commission correctly understands that Flavin’s light tubes and Viola’s DVD players were not decorative elements of the installations. Haunch of Venison did not argue that decorative elements made these sculptures art; instead, the gallery argued it was the artists’ intentions, along with their

---

217 Id; see also Paragraphs 98 (“Manufactures of marble, etc.”) and 376 (“Paintings, drawings, etc.”) of the Tariff Act of 1913, 38 Stat. 114, 123, 151 (1913), ch. 16, paras. 98, 376.
218 Olivotti & Co., 7 Ct. Cust. App. at 47.
219 Id.
220 Id. at 48. At this time the court was still hewing closely to the conception of sculpture being mainly representative of “natural objects, chiefly the human form.” Id. Twelve years later in Brancusi, the court recognized this traditional understanding but gently put it to the side in light of the changing tides of contemporary art. See supra notes 176-98 and accompanying text.
222 Id. (citing United States v. Perry, 146 U.S. 71, 74 (1892)).
reception by the art world, that made the works art. Thus *Olivotti* is not on point in this matter. In any event, *Brancusi* removes any lingering doubts that even basic materials can be deserving of tax discounts offered to works of art by virtue of their artistic manipulation by artists into artworks.

IV. CRITIQUES AND SUGGESTIONS

The “tension between the law and the evolution of ideas in modern or avant garde art” can lead to the protectionist tendencies seen in the U.K. customs officials’ worries about retail imports classified under false pretenses. The law is insistent upon “taxonomiz[ing and classifying] artistic creations,” while the avant-garde is making valiant efforts to be “whatever [one] can get away with.” On a fundamental level, this tension might always exist because “law is about precedent whereas art is about the evolution of ideas.” The law cannot be expected to accommodate such a broad (and shifting) definition of art, but, significantly, “[a]rt is not apart from the law.” Often with conceptual art, “extrinsic circumstances” and context must be taken into account to properly classify the works. If a work is accepted within the context of the “art world,” that is sometimes the full extent to which it is validated as art. The law’s inherent structure is at odds particularly with a school of art that requires a certain amount of context in its presentation. The troubles foreseen by U.K. customs officials—an open door for importers to call any shipped product “art”—is rooted here. For some of the more conceptual artworks, testimony as to their legitimacy will grow in necessity and importance. The alleged or potential burden of such a requirement, however, is not a justification for blanket legislation that makes even legitimate imports impossible.

---

223 Haunch of Venison Partners Ltd. v. Revenue & Customs Comm’rs, [2008] UKVAT (Customs) C-00266, 2008 WL 5326820 (Dec. 11, 2008 VAT & Duties Tribunal (London)).


226 *Id.*

227 Farley, *supra* note 19, at 807.

228 Cronin, *supra* note 225, at 213.

229 Farley, *supra* note 19, at 808.

230 Cronin, *supra* note 225, at 236.

231 See generally Arthur C. Danto, *The Artworld*, 61 J. PHILO. 571 (1964); see also Farley, *supra* note 19, at 844 (discussing Danto and “institutionalism”).
The Visual Artists Rights Act (VARA), adopted by Congress in 1990 to broaden general copyright protections to include artists producing physical “work[s] of visual art,” attempts to resolve some of the art versus law dispute by establishing criteria which obviate subjective classification. While VARA is not without criticism, it makes inroads to recognizing the “moral rights” of artists in a way that the United States has not previously done. VARA gives credence to both the claims of the artists themselves (in declaring a visual work “art”) and the claims of the art world and its specialists (in recognizing or lauding a visual work as “art”). The foreign nations that are signatories to the Berne Convention accept a similar convention in the “droit morale” protections offered to artists in those jurisdictions. The copyright protections under VARA—echoes of droit morale—bridge art and law, and may offer an effective legal rubric for the assessment of art for tax and other purposes. The Tribunal that heard the 2008 Haunch of Venison lawsuit has already employed this standard: it took testimony from experts in the art world and made sure to ascertain that Flavin and Viola were indeed bona fide artists—that they had “recognized stature.” This part will further discuss these and other possible “bridges” between art and law.

A. Problems for Conceptual Art

Contemporary, conceptual art has a difficult status in society and in the law. Part of the problem is that “the ‘plain and ordinary’ meanings of words describing modern art” cannot keep pace with the developments within these art styles and types. The law has equal difficulty determining how to treat these types of developing works. As discussed below, VARA offers protection of certain artists’ rights for works of a “recognized stature.” Some scholars read this to mean that VARA only protects “the most revered work of the Old Masters.” Yet others see the low bar on creativity in the

---

232 17 U.S.C. § 106A(a) (2006); see also Cronin, supra note 225, at 209.
Copyright standards as allowing for artwork with only a minimal level of original, authored expression to garner copyright protections.238

Charles Cronin, in a 2009 article concerning VARA and conceptual artworks made with living materials, asserts that in order to classify these conceptual works as art, “extrinsic circumstances [are relied upon] to a much greater extent than . . . works in traditional genres.”239 He argues that the nature of conceptual works requires a contextual approach: the works cannot be interpreted without these extrinsic circumstances. A Monet would easily be recognized, even if “stripped of its sumptuous frame[,]” but Jeff Koons’s balloon figures or Duchamp’s urinal would be seen as having little or no aesthetic appeal if they were encountered outside of their “frame” or “art” context.240 For Cronin, the reliance on extrinsic circumstances and context means this avant-garde art garners less protection from VARA. Cronin’s argument is limited, however, because these more conceptual works are intentionally moving away from the traditional confines of gallery walls and museum spaces.241 In fact, “[c]ertain current art practice is about breaking down the doors of art’s exalted cloister and exploding the definition of art, especially definitions that envision a narrow ‘high’ art.”242 This artistic practice should be protected even if it presents a challenge to the current copyright scheme.

Cronin also argues that conceptual works garner less copyright protection because they are primarily concepts or ideas. It is universally understood that ideas are not copyrightable, yet this overlooks the material components of these works. Artists over the last century have been stretching the form of artworks but have not ceased creative expression through their chosen medium. Conceptual and contemporary artists have utilized customarily functional materials to access artistic expression that the more removed and rarified “traditional” art materials sometimes cannot. Conceptual artists present finished works that

238 See infra notes 254-53 and the accompanying text.
239 Cronin, supra note 225, at 236.
240 Id. at 235-36. What Cronin considers a demerit toward the classification of the artwork Arthur Danto argues is a baseline required for the interpretation of art objects. Danto and other “institutionalists” would not distinguish between Monet and Duchamp; for them, all art objects require the context of the “art world” in order to be seen as art. See generally Danto, supra note 231; see also Farley, supra note 19, at 844.
241 For a discussion of intentional changes in and departures from the art-world ethos, see supra Part I.A.
242 Farley, supra note 19, at 814.
are products of creative expression, but Cronin argues that artists should not be allowed to “elevate[] the status of these [non-art] materials to that of art by addressing them as such . . . .” Cronin asserts that works employing materials that have not been in use “[s]ince time immemorial” do not rise to the level of art just because an artist has done something creative with them. However, this assertion is simply untenable. Nowhere in the copyright-protection statutes exists such a requirement that art materials satisfy some preapproved list of acceptable and traditional materials. Cronin offers two pointed criticisms of avant-garde art, both of which focus on contextual and material components, but he leaves out artistic intention and how the art world receives the work. These conceptual works should not fail to garner protection simply because they are focused on the idea, and not the materials employed in their execution.

B. Moral Rights, Foreign and Domestic

International copyright protections were established in the Berne Convention for the Protection of Literary and Artistic Works (Berne Convention). This doctrine is particularly focused on the “[r]ights of [a]tribution and [i]ntegrity.” Berne Convention rights “are commonly called ‘moral rights’ because they ‘are a constellation of rights that ensure an ongoing relationship between the author and the creative work outside [of] economic issues.’” For instance, the right of integrity “allow[s] an artist to prevent changes to her work that would affect her honor or reputation negatively.” This right, granted to artists, serves to protect their finished work product from

243 Cronin, supra note 225, at 252.
244 Id. at 243.
245 17 U.S.C. § 106A(a) (2006); 7 W ILLIAM F. PATRY, P ATRY ON COPYRIGHT § 23:16 (2011); Berne Convention for the Protection of Literary and Artistic Works, art. 6bis, Sept. 9, 1886, S. Treaty Doc. No. 99-27 (1986), 1160 U.N.T.S. 30. The United States adheres to the Berne Convention with certain important exceptions, namely that the U.S. Berne Implementation Act of 1988 insists that “the Copyright Act, the Lanham Act, and state law” are the “sole source of rights” for copyright protection in the United States, rejecting the additional “moral rights” under Article 6bis of the Berne Convention. 7 PATRY, supra, § 23:23. “The obligations of the United States under the Berne Convention may be performed only pursuant to appropriate domestic law.” Id. § 23:45 (quoting Pub. L. No. 100-568, 100th Cong., 26 Sess. § 2(2), 102 Stat. 2853). This is not to say that certain moral rights have not been extended in the Copyright Act to works of visual art over time (e.g., VARA), but rather that the United States’s recognition of artists’ moral rights came well after the international community’s. Id. § 23:23.
247 Id. at 220 (quoting H.R. Rep. No. 101-514, at 6).
being changed into something else (e.g., a collection of non-art components). Though a layperson might try to destroy an artist’s work by denigrating or taxing it, this kind of clumsy effort might indeed enhance the prestige of the artist. Still, the honor of artists as artists lives through their works; the act of denying that their work is art may play well in the press, but it indeed negatively affects their honor.

European copyright laws appear “better adapted to address the interests of fine artists because they protect not only the finished work, but also the artist’s control over the creative process and ultimately her persona and reputation.”

It is ironic, then, that EU copyright better respects the persona and reputation of the artist while the EU VAT system, which recognizes both the existence and cultural importance of fine artwork, does not. These two systems should ideally be working in concert; interpretations under one (i.e., TARIC) should respect and heed the protections of the other (i.e., the Berne Convention). In passing the 2010 regulation, the Commission seems to be doing anything but that. Considering the increasing use of “atypical” materials in contemporary artworks, the implications of the 2010 regulation become all the more alarming. The next time a customs official in the U.K. is confronted with a box of bricks or cords of wood imported by a gallery or museum and listed as an artwork, it is very likely that they will be subject to the import VAT appropriate for their raw industrial components. The problem of appropriate taxation will continue to arise as the frequency of inclusion of non-art materials continues to increase in paintings, sculptures, and other types of artworks.

In the United States, VARA extended certain copyright protections to “author[s] of work[s] of visual art.” Some of the rights granted through VARA were “[r]ights of attribution and integrity.” These rights match the Berne Convention system, but standard U.S. copyright requirements must first be met before the VARA rights are even reached. The basic tenets of copyright protection under the Copyright Act require that a work be an original, authored expression in a fixed and

---

248 Sapolich, supra note 47, at 455.
249 For a discussion of the increasingly common use of nontraditional materials in artworks, see supra notes 58-69.
251 Id.
252 Kelley v. Chi. Park Dist., 635 F.3d 290, 292 (7th Cir. 2011), cert. denied 132 S. Ct. 380 (2011). The court in Kelley called these basic copyright requirements “foundational.” Id.
tangible form. In U.S. copyright law, “the requisite level of creativity is extremely low; even a slight amount will suffice.” Courts find “creativity is not a high bar to copyright, and... a ‘work of art’ exists when ‘by the most generous standard [it] may arguably be said to evince creativity.’ This implies that moral rights of conceptual artists could be given all the more protection: if only a modicum of creativity is required, many alleged artworks will be accepted as such, and many alleged artists will have their moral rights protected.

Though Flavin’s and Viola’s sculptures are only “minimally differentiated from [their] non-art materials,” these artists have contributed more than a little creativity in detailing how to construct the installations. So long as this minimal level is met, it would seem U.S. copyright protection could be asserted. How is it, then, that the U.S. copyright provisions can be seen as more assertive of an artist’s moral rights than the European system and its droit morale protections? This demonstrates that something is amiss in the 2010 regulation.

A copyright can be obtained with a “low standard of originality.” This “is intended to minimize the possibility that judges would interject their own ideas of what is and is not art.” Setting the creativity bar low keeps judges from having to make subjective determinations, thus “ensur[ing] that judges remain objective and neutral.” Unpopular or controversial art movements could be at risk if judges needed to make subjective determinations as to whether a work warrants copyright protection. Certain trends in “artistic development might be stultified by ignorant or outdated legal evaluations.” “Judges [could not] make artistic decisions while remaining objective,” since the heart of artistic decisions is individual taste, an inherently subjective concept. The U.S. copyright system properly considers a work worthy of protection even if it only

---

255 Id. (second alteration in original) (quoting MELVILLE B. NIMMER & DAVID NIMMER, 1 NIMMER ON COPYRIGHT § 2.08(B)(1) (Matthew Bender & Co. 2006)).
256 Id. at 460.
257 Id. at 472.
258 Id.
259 Id.
260 Farley, supra note 19, at 814.
261 Sapolich, supra note 47, at 473 (quoting Farley, supra note 19, at 812-13).
evinces creativity. That the Commission, operating within a system granting greater protections for artists’ moral rights, does not acknowledge this creativity as a contributing factor for being considered “art” further demonstrates the mysteriousness of the Commission’s standard.

C. VARA’s “Recognized Stature” Protection

Section 106A(a)(3) of VARA has particular relevance for Flavin and Viola. This section provides authors of works of visual art the rights:

(A) to prevent any intentional distortion . . . or . . . modification of that work which would be prejudicial to his or her honor or reputation, and any intentional distortion . . . or modification of that work is a violation of that right, and (B) to prevent any destruction of a work of recognized stature, and any intentional or grossly negligent destruction of that work is a violation of that right.

Destruction is defined as “the state or fact of being destroyed” or “the action or process of destroying something.” Destroy means “to put out of existence” or “neutralize.” Admittedly these definitions are narrow in that they bring to mind tangible, visceral destruction. In the context of more conceptual and experiential work, what is destruction? Perhaps it is simply seeing only the components and not the whole. When a party—not the artist—acts to remove or collapse or disassemble an artwork, provided it fulfills the other requirements under VARA, that party can become liable for destruction under the statute. When the discussion centers on minimal or conceptual works of art, the question of destruction is often tricky. Damien Hirst’s “trash installation” following an opening-night party is a near-perfect example. The morning after the opening, the gallery cleaning crew threw away bags containing spent wine cups and cigarette butts, thinking them

265 VARA briefly discusses what destruction or modification is not, holding conservation and relocation efforts acceptable "modifications" of an artwork that do not actually destroy it. 17 U.S.C. § 106A(c)(1)-(2). For conceptual artworks, this does not go far enough. In dismissing something as small as an idea, a conceptual work may in fact be destroyed.
nothing more than garbage. But, Hirst had arranged the “trash” after the party, considering it a part of the installation. Although Hirst found it humorous, in terms of VARA protection it is possible that the cleaning crew, or even the gallery, would have been liable for the installation’s “destruction.”

In the context of artwork preservation, these definitions also chafe against the realities of some contemporary works of art. Certain works “must be disassembled in order to preserve their value and ensure their continued existence.” Certainly Flavin’s and Viola’s installations must be “destroyed” (i.e., disassembled) in order to move them from museum to gallery or vice versa. In fact, “disassembly is [often] required to conserve the work consistently with the artist’s vision.”

Nathan Davis’s article challenged a Southern District Judge’s decision for hewing too closely to the dictionary definitions of “remove” or “destroy” in the context of artwork preservation. Davis asserts, “There is a difference between dismantling the sculpture never to recompose it, and dismantling a sculpture intending to put it back together once a part of it has been fixed.” This distinction applies directly to the Flavin and Viola installations. Both the Flavin and Viola works were dismantled with the intention that the London gallery which was importing them would put them back together. When the works arrived at U.K. customs, this should have been obvious. Had the European Commission understood this distinction, it is possible that the classification of these installations as “sculpture,” by the U.K. Tribunal would have gone undisturbed. Sadly, the Commission did recognize that the Flavin and Viola components were intended to be reassembled, but they nevertheless denied these installations “sculpture” status. Instead, the Commission required the act of switching the lights and projectors on in order to achieve a finished artwork. As discussed in Part III, this standard is impossibly

---

267 Id.
268 Id.
269 Id.
270 Davis, supra note 57, at 241.
271 Id. (discussing a Robert Morris minimalist sculpture entitled Rope Piece consisting of a piece of “rope draped between two painted wooden elements”). Id. at 240.
272 Id. at 241 (discussing Bd. of Managers of Soho Int’l Arts Condo. v. City of New York, No. 01 Civ. 1226, 2005 WL 1153752 (S.D.N.Y. May 13, 2005)). That decision seems predicated on the idea that the work is destroyed once disassembled, thereby being unable to “return to existence.” Id.
273 Id.
274 See supra Part III.A.
275 See supra Part III.A.
high, and the distinction the Commission forces on these works is not really one of \textit{assembled versus disassembled}, but one of \textit{on versus off}. Conservationists of any other type of artwork would hardly make such a distinction as “whether the electricity is flowing.”

The “moral rights” secured through VARA “protect[] the right of an artist to preserve a work of art even after that work is sold.”\textsuperscript{276} It is VARA’s requirement of “\textit{recognized stature}” that gives an artist’s moral-rights claim any legitimacy\textsuperscript{277}: “where a particular work of art has achieved recognized stature, VARA gives the artist the right to prevent its destruction.”\textsuperscript{278} The two-part test for prevailing on these VARA-violation claims requires the plaintiff to prove (i) that the piece is a “work of recognized stature,” and (ii) that the “[d]efendants destroyed the piece in an intentional or grossly negligent manner.”\textsuperscript{279} Recognized stature “is generally established through expert testimony” that proves that the “artistic merit” of the piece has “been recognized by . . . the artistic community and/or the general public.”\textsuperscript{280} For VARA protection, the artist’s particular piece at issue must “have achieved such stature[,]” though there are circumstances imaginable in which an artist “is of such recognized stature that any work by that artist would be subject to VARA’s protection.”\textsuperscript{281} Nathan Davis reads the statute and concludes it is “[a]n imperfect solution to the problem”\textsuperscript{282} because it leaves out too many conceptual artists, honoring only the “Old Masters.”\textsuperscript{283} Though not perfect, accepting recognized stature of the artist as proof that the artist’s finished products are “art” is certainly more defensible than the Commission’s treatment of Flavin and Viola.


\textsuperscript{277} 17 U.S.C. § 106A(a)(3)(B) (2006); see also Scott, 309 F. Supp. 2d at 400.

\textsuperscript{278} Id. (citing 17 U.S.C. § 106A(a)(3)(B); Carter, 861 F. Supp. at 325); see also Martin v. City of Indianapolis, 192 F.3d 608, 612 (7th Cir. 1999) (discussing Carter's test interpreting VARA).

\textsuperscript{279} Scott, 309 F. Supp. 2d at 400 (internal quotation marks omitted).

\textsuperscript{280} Id. In the instance of the Hirst party trash being thrown away, Hirst’s recognized stature would likely impute onto any work he claims to have authored.

\textsuperscript{281} Davis, supra note 57, at 221.

\textsuperscript{282} Id. at 228.
“Self-Expression” and Arbiters of Taste

Besides recognized stature, U.S. courts have required self-expression when granting protections to an artwork under the First Amendment. The Ninth Circuit Court of Appeals held that “[a]ny artist’s original [artwork] holds potential to ‘affect public attitudes,’ by spurring thoughtful reflection in and discussion among its viewers. So long as it is an artist’s self-expression, [an artwork] will be protected under the First Amendment, because it expresses the artist’s perspective.”

This “self-expression” need not be singular or narrow. The Supreme Court has said that “a narrow, succinctly articulable message is not a condition of constitutional protection . . . .” Requiring such a message, the Court reasoned, would invalidate entire wings of major museums that show artists like Jackson Pollock. Rather, self-expression can be broadly construed. The Court has also “distinguished between restrictions on expression based on subject matter and restrictions based on viewpoint, indicating that the latter are particularly pernicious.”

A “bedrock principle” of First Amendment jurisprudence “is that the Government may not prohibit the expression of an idea simply because society finds the idea itself offensive or disagreeable.” When legislation’s “suppression of speech . . . attempt[s] to give one side of a debatable public question an advantage[,] . . . the First Amendment is plainly offended.”

The Second Circuit takes a similar view of legislation that “looks upon visual art as mere ‘merchandise’ lacking in communicative concepts or ideas.” In Bery v. City of New York, the court addressed an appeal concerning regulations the City of New York had adopted prohibiting sales of art in public places without a general vendor’s license. The court found the approach of the city “myopic[,] . . . [and] fundamentally

---

284 White v. City of Sparks, 500 F.3d 953, 956 (9th Cir. 2007) (citation omitted) (quoting Joseph Burstyn, Inc. v. Wilson, 343 U.S. 495, 501 (1952)) (discussing the right of an itinerant painter to be granted the necessary permits to sell his work in a restricted park).
286 Id.
288 Id. (quoting Texas v. Johnson, 491 U.S. 397, 414 (1989)).
289 Id. at 430-31 (quoting First Nat’l Bank of Bos. v. Bellotti, 435 U.S. 765, 785-86 (1978) (internal quotation marks omitted)).
290 Bery v. City of New York, 97 F.3d 689, 695 (2d Cir. 1996).
291 Id. at 691.
misperceiv[ing] the essence of visual communication and artistic expression.”

Not so surprisingly, these courts have not required the aesthetic opinions of the judges themselves. Judges have been hesitant to take steps that would “destroy” an artwork (though they are also practicing some self-protection). The majority in Martin v. City of Indianapolis, a Seventh Circuit case, began, “We are not art critics, do not pretend to be and do not need to be to decide this case.” The concurrence expressed a similar sentiment: “Like my colleagues, I am not an art critic. So I begin with the well-worn adage that one man’s junk is another man’s treasure. No doubt [the artist] treasured what the city’s bulldozers treated as junk.”

These judges, and many others, rely on Justice Oliver Wendell Holmes’s famous statement on the intersection of aesthetics and judicial restraint. In Bleistein v. Donaldson, one of Holmes’s first Supreme Court opinions, he said that “[it] would be a dangerous undertaking for persons trained only to the law to constitute themselves final judges of the worth of [artworks] . . . . [S]ome works of genius would be sure to miss appreciation.” Holmes’s opinion understood that many art movements are “repulsive until the public [learns] the new language in which [the artist] spoke.” Judges fear exposing themselves as “culturally elite” by revealing their aesthetic views.

---

292 Id. at 695.
293 Id. Ten years later, in Mastrowinci v. City of New York, the Second Circuit declined to extend the “Bery injunction” to plaintiffs who were selling articles of clothing painted with graffiti. 435 F.3d 78, 82 (2d Cir. 2006). The court reasoned that articles of clothing painted with graffiti were not necessarily expressive, and that the Bery injunction should be narrowly read not to include “clothing painted with graffiti” under the category of “paintings.” Id. Additionally, objects that are utilitarian in nature or are promotional/advertising materials do not fall within the protections of VARA. 17 U.S.C. § 101 (2006); Kleinman v. City of San Marcos, 597 F.3d 323, 327, 329 (5th Cir. 2010).
294 Martin v. City of Indianapolis, 192 F.3d 608, 610 (7th Cir. 1999). In Martin, the court discussed the city’s argument that the artist had waived this VARA protection through their contract. Id. at 614. Ultimately, the court was not persuaded by the city’s argument. Id.
295 Id. at 615 (Manion, J., concurring in part and dissenting in part).
296 Farley, supra note 19, at 815.
297 Id. (quoting Bleistein v. Donaldson Lithographing Co., 188 U.S. 239, 251 (1903)).
298 Bleistein, 188 U.S. at 251.
opinions in court—and no judge wants their opinion to be vulnerable to later attack.299

In Martin there was a contract between the artist and the city, which had included the possibility of future removal of his sculpture from its original site.300 In fact, the sculpture was created with this specific possibility in mind, “engineered . . . so that it could be disassembled for removal and later reassembled.”301 Nowhere in the Martin opinion did the court indicate that the artwork, or the VARA protections afforded to it, suddenly ceased to exist upon any potential disassembly for relocation. Once afforded the recognized stature, the VARA protections for Martin’s sculpture could not be easily undone; taking the piece apart for relocation would not have compromised those protections; neither would have, for example, moving the piece on a flatbed travelling down the highway. The legal protections would not disappear simply because the artwork was disassembled for removal. Clearly they would not cease to exist during shipment, either. Such an implication would be—to recall the language of the U.K. VAT Tribunal—absurd. It would mean that an artwork’s VARA protections could be circumvented if the artwork was first disassembled. Somehow the components could be destroyed without violating VARA but the assembled work could not. Such machinations would be the equivalent of removing a work’s copyright protections simply by disassembling it. The courts’ VARA interpretations provide that an artist’s legal protections extend to the disassembled components of an artwork because the artwork retains its status as an artwork even when disassembled. The European Commission’s declaration that Flavin’s and Viola’s works in disassembled form (and even once re-assembled) are void of artistic content is in direct contradiction to this learned jurisprudence.

While both the U.S. and EU legal systems accommodate the artist’s moral rights—either through droit morale, VARA and copyright protections, or the First Amendment concept of self-expression—what underpins them all is an acceptance that once a work is recognized by the art community as art, copyright and other protections should be afforded to it. Conceptual works of art may find trouble in language and definition, but having satisfied the baseline original and

---

299 Farley, supra note 19, at 814-15.
300 Martin, 192 F.3d at 611.
301 Id.
creative expression requirements under the U.S. copyright system, these works should at a minimum not be denied protections under the law. The question of what qualifies as destruction under VARA, while important to investigations of liability under the Act, is secondary to the first requirement under VARA—the recognized stature provision. It is this standard that courts, both domestic and foreign, are primed to apply, and many have already proven their ability to do so. Legislators can easily get behind this standard because it removes any subjectivity they might need (or want) to employ in their lawmaking. Lawmakers should decline to act as arbiters of taste, just as many learned judges have done, and leave instead the expert determinations of an artwork’s status to the art experts.

CONCLUSION

While some of the pitfalls of VARA have been discussed above, the standards set out in the legislation—the requirements of “recognized stature” in particular—would make courts’ evaluations of the merits of an object, installation, or experience as an artwork more reliable. This same structure could also be applied to the imposition of import taxes and duties, particularly in jurisdictions that provide for lower tax rates on artwork. The means for applying taxes on the basis of an object’s status as “artwork” would be well served by the requirements imposed under this recognized-stature condition.

While the United States does not currently impose a tax system similar to the harmonized system of the European Community, leading economists have proposed a “consumption tax” as an answer to the nation’s economic woes. Though the idea “offends many conservatives” for its enabling of an expanded spending power of the government and an increase in government overall, liberals find it equally unfavorable for imposing taxes on citizens’ consumption. It is possible that the VAT “appeal to liberals can be enhanced . . . by exempting items

302 See supra notes 282-83 and accompanying text.
304 Id.
such as food and housing.” Economists claim that implementing this type of tax on consumer spending could raise “revenue [of] . . . roughly five percent of G.D.P.” In fact, some economists see a VAT as “more efficient than an income tax.” But, regardless of one’s opinion, implementing such a system would leave the nation vulnerable to a host of classification questions. Of course, the classification of artworks for taxation purposes is not the main concern of a government seeking to dig itself out of a sluggish economy, but allowing further sacrifice of the value of our cultural works would have its own deleterious effects on the nation.

For Flavin’s and Viola’s works, the matter is not yet closed, even though the European Commission went “to such elaborate lengths to overturn the decision of [the Tribunal] in relation to a relatively small amount of import tax in relation to artworks.” There is hope: David Zwirner gallery, which brought a seldom-seen Flavin series to the International Contemporary Art Fair (FIAC) in October 2011, and which represents the Flavin estate, has recently retained the law firm of Mayer Brown to “explore the gallery’s legal options regarding the . . . ruling.”

Rachel J. Tischler†


Immense gratitude to the members of the Brooklyn Law Review for the care with which they edited this note. Also thanks to my many readers and editors for their tireless help: Professor Steven Dean, Steven Glazer, Mark Merriman, Hans R.F. Vermy, Amy Tischler, and the 5-College PATF. Additional thanks to Professor Heidi Gilpin for opening my mind in college. To my family for making sure I kept a level head throughout law school—my love to you. Finally, for his boundless and unconditional encouragement, support, and most of all love, I dedicate this note to Owen Freeman-Daniels. Any errors remain my own.

---

306 Barro, AAA Rating, supra note 305.
307 Barro, Save the Economy, supra note 303.
308 Id.