TABLE OF CONTENTS

ARTICLES

Policing School Discipline  
*Catherine Y. Kim*  861

Not Sick Yet: Food-Safety-Impact Litigation and Barriers to Justiciability  
*Diana R. H. Winters*  905

From Bad to Worse: Assessing the Long-Term Consequences of Four Controversial FCC Decisions  
*Rob Frieden*  959

NRSRO Nullification: Why Ratings Reform May Be in Peril  
*Jason W. Parsont*  1015

A Modest Proposal: The Aged of Death Row Should Be Deemed Too Old to Execute  
*Elizabeth Rapaport*  1089

NOTES

Finding a Better Analogy for the Right of Publicity  
*Andrew T. Coyle*  1133

Disability or Identity?: Stuttering, Employment Discrimination, and the Right to Speak Differently at Work  
*Jared Gilman*  1179

A Standardless Standard: How a Misapplication of *Kelo* Enabled Columbia University to Benefit from Eminent Domain Abuse  
*Justin B. Kamen*  1217
Smoking Out Racism in the FDNY: The Dwindling Use of Race-Conscious Hiring Remedies
Hal Budnick 1249

Cleaning Up the Muck: A Takings Analysis of the Moratorium on Deepwater Drilling Following the BP Oil Spill
Edward W. Thrasher 1285
Policing School Discipline

Catherine Y. Kim†

INTRODUCTION

Notwithstanding the frequent admonishment that “students [do not] shed their constitutional rights . . . at the schoolhouse gate,” courts routinely defer to school officials in cases involving the investigation and punishment of youth in schools. Consequently, youth accused of school misconduct are not entitled to the same procedural protections to which they would be entitled outside the school context: school officials may search their belongings or persons without a warrant or probable cause, and officials may question them without first providing Miranda warnings. Courts and scholars alike defend such

† Assistant Professor of Law, University of North Carolina School of Law at Chapel Hill. I am grateful to Tamar Birckhead, Caroline Brown, Aaron Caplan, Brietta Clark, John Coyle, Charles Daye, Dana Thompson Dorsey, Maxine Eichner, Michael Gerhardt, Myriam Gilles, Don Hornstein, Joe Kennedy, Bill Marshall, Eric Muller, Richard Myers, Teri Ravenell, Mark Weidemaier, Kimberly West-Faulcon, and the participants in the George Washington University Law School faculty workshop and the SMU Dedman School of Law faculty workshop for their time and invaluable feedback. All errors are my own. I would also like to thank Josh Kinard, Tiffany Brown, Robert Lamb, and Kiril Kolev for their excellent research assistance.

2 See, e.g., Vernonia Sch. Dist. 47J v. Acton, 515 U.S. 646, 656 (1995) (holding that the scope of constitutional rights for public school students is limited by “what is appropriate for children in school”).
4 New Jersey v. T.L.O., 469 U.S. 325, 340 (1985) (sustaining student search where school officials have a reasonable suspicion that the search will uncover evidence of criminal activity or a violation of school rules).
5 Although the Supreme Court has not directly ruled on this issue, lower courts consistently find no custodial interrogation where a youth is questioned by a
restrictions on students’ constitutional rights on the ground that school discipline, unlike law enforcement, serves the educational interests of youth. Under this view, the educational value of discipline and consequent alignment of interests between official and student render the constitutional protections guaranteed outside of the school context inapposite in schools.

Recent observations of a “school-to-prison pipeline” resulting in the increased criminalization of student misbehavior, however, cast doubt on this characterization of school discipline. Today, police officers routinely patrol public school hallways on a full-time basis as “school resource officers”; and school officials refer a growing number of youth to the juvenile and criminal justice systems for school-based misconduct. These developments call for a critical reassessment


See infra Part I.A; see also Buss, supra note 3, at 570 (describing judicial deference to the “mystique of the educational institution”); Anne Proffitt Dupre, Should Students Have Constitutional Rights? Keeping Order in the Public Schools, 65 Geo. Wash. L. Rev. 49, 64 (1996) (conceptualizing debate over restrictions on students’ constitutional rights as a debate over competing educational goals); James A. Ryan, The Supreme Court and Public Schools, 86 Va. L. Rev. 1335, 1340-41 (2000) (analyzing limits to students’ constitutional rights as measured against academic function of schools).

See Schulhofer, supra note 3, at 118 (analyzing restriction of probable cause requirement for student searches as based on the view that Fourth Amendment protections reflect[] a balance appropriate mainly to cases in which private activity and public controls are posed in conflict,” which is not the case when students are searched for wrongdoing); see also Debra Livingston, Police, Community Caretaking, and the Fourth Amendment, 1998 U. Chi. Legal F. 261, 294-96 (arguing that alignment of individual and collective government interests renders criminal procedural protections, such as warrant requirement, unnecessary). But see Laurence Tribe, Structural Due Process, 10 Harv. C.R.-C.L. L. Rev. 269, 312 n.28 (1975) (critiquing assumption that interests of school official and punished student are aligned rather than in conflict for purposes of school discipline).

See CATHERINE Y. KIM ET AL., THE SCHOOL-TO-PRISON PIPELINE: STRUCTURING LEGAL REFORM (2010); MARTHA MINOW, IN BROWN’S WAKE: LEGACIES OF AMERICA’S EDUCATIONAL LANDMARK 28 (2010) (utilizing term “school-to-prison pipeline” to refer to the way in which “school systems, police, and juvenile justice programs combine in a process that removes students from mainstream schools and puts them in separate programs that often involve lockup, searches, and little educational value”); Lisa H. Thurau & Johanna Wald, Controlling Partners: When Law Enforcement Meets Discipline in Public Schools, 54 N.Y.L. Sch. L. Rev. 977, 981 (2010) (noting that “the term school-to-prison pipeline” has become “part of our national lexicon,” used to describe “the growing trend of school officials to refer students to law enforcement for acts committed while in school, and the increasing deployment of police in schools”).

of the extent to which contemporary school discipline practices advance the educational goals that historically justified their insulation from judicial scrutiny. 

This article evaluates empirical evidence on contemporary discipline practices and finds that, in a growing number of jurisdictions that rely on law enforcement to maintain order in schools, it can no longer be said that the investigation and punishment of school misconduct serves the accused student’s educational interests, or even the interests of the larger student body.

These changes in the operation of school discipline parallel the changes to the juvenile justice system addressed in the landmark case of In re Gault. Traditionally, youth in juvenile court were not entitled to the procedural protections guaranteed to adults in criminal court, on the ground that juvenile courts, unlike criminal courts, were assumed to be nonadversarial institutions designed to further the best interests of the youth; young people would receive the benevolent protection of court officials in exchange for giving up their procedural rights. Accumulating evidence of juvenile courts’ failure to achieve those beneficent goals, however, led the U.S. Supreme Court in Gault to reconsider prior doctrine and extend to youth at least some of the procedural rights


See Holland, supra note 5 (arguing for consideration of increased policing in schools in determining Miranda rights for youth questioned at school); Josh Kagan, Reappraising T.L.O.’s “Special Needs” Doctrine in an Era of School-Law Enforcement Entanglement, 33 J.L. & EDUC. 291, 304, 321 (2004) (contending that probable cause should be required for student searches in schools where police officers are permanently staffed at school, security cameras are prevalent, and school officials are required to report criminal incidents to police); Michael Pinard, From the Classroom to the Courtroom: Reassessing Fourth Amendment Standards in Public School Searches Involving Law Enforcement Authorities, 45 ARIZ. L. REV. 1067, 1070-71 (2003) (arguing that student searches should require probable cause if search could result in criminal liability); Jacqueline A. Stefovich & Judith A. Miller, Law Enforcement Officers in Public Schools: Student Citizens in Safe Havens?, 1999 BYU EDUC. & L.J. 25 (exploring doctrinal issues arising from police involvement in public schools).

387 U.S. 1 (1967). This parallel between juvenile justice and school discipline was presciently drawn as early as 1971 by William G. Buss. Buss, supra note 3.

In re Gault, 387 U.S. at 14-17.
formerly limited to adults in criminal court. Courts today likewise should evaluate evidence of school discipline’s achievement of its beneficent goals, and modify accordingly the procedural protections available to youth in public schools.

Part I explores the development of the educational theory of school discipline in legal doctrine, focusing on the role that social science has played to restrict the procedural rights of students. The Supreme Court has reasoned that school discipline—in stark contrast to law enforcement—serves the educational interests of the student who is investigated and punished; for this reason, constitutional rights that would be available to youth outside of school are not available to them in the context of school discipline. In the early foundational cases establishing these restrictions, members of the Court relied on personal intuitions about school discipline, even when those intuitions conflicted with empirical evidence properly presented before them. More recent cases, however, suggest an increased willingness to scrutinize the impact of school discipline practices in determining whether the deference traditionally afforded to school officials remains warranted.

Part II analyzes empirical findings on contemporary school discipline practices and their pedagogical impact, focusing on school-based arrests and other forms of referral to law enforcement. Analyzing a number of recent empirical studies, this part finds that schools increasingly rely on law enforcement to maintain order, although the extent to which they do varies. It then explores scholarship from related disciplines in education, sociology, and criminology to conclude that the use of law enforcement in schools has a negative impact on educational outcomes, not only for the investigated youth, but also for the larger student body. These findings suggest that the investigation and punishment of students in at least some jurisdictions no longer serves the pedagogical interests that traditionally justified exempting students from procedural protections.

Part III sets forth a framework for courts and nonjudicial actors to take such social science evidence into account. It proposes that courts engage in a factual assessment of school discipline practices to determine whether the traditional rationale for denying youth in schools constitutional

---

\[13\] Id. at 18-24.
procedural rights remains warranted.\textsuperscript{14} Given the significant variance across jurisdictions in school discipline practices, the analysis employed should be location-specific. The presumption that school discipline serves pedagogical goals would be preserved, but could be rebutted with evidence showing that disciplinary practices in the particular school or district at issue do not further the educational interests of accused youths. Where a court finds that school discipline operates primarily to further law enforcement goals rather than pedagogical goals, investigations of student misconduct should be presumed to be adversarial and thus subject to the full scope of constitutional protections that would be available to youth outside the school context. By contrast, where school discipline practices are found to adhere to the traditional model of discipline in furtherance of pedagogical goals, doctrinal restrictions on students’ constitutional rights would remain in place. Part III then considers the role of nonjudicial actors, arguing that those who make the substantive determination as to whether certain forms of conduct should be criminalized in the first instance will play a critical part in any reform effort.

I. DOCTRINAL RESTRICTIONS ON STUDENTS’ PROCEDURAL RIGHTS

Courts routinely defer to school officials in cases involving the investigation and punishment of students.\textsuperscript{15} A schoolchild accused of bringing a water pistol to school or tearing a page out of a book does not enjoy the same constitutional rights as an adult or child suspected of a criminal act on the street.\textsuperscript{16} School officials may search the


\textsuperscript{15} See, e.g., Vernonia Sch. Dist. 47J v. Acton, 515 U.S. 646, 656 (1995) (holding that scope of constitutional rights for public school students is limited by “what is appropriate for children in school”).

\textsuperscript{16} See Buss, supra note 3, at 640-41 (discussing constitutional rights of a student accused of tearing a page out of a book); Schulhofer, supra note 3, at 115 (comparing the rights of a “student with a water pistol” with those of an adult suspected of a crime).
youth's backpack without a warrant or probable cause and question the youth without first providing Miranda warnings. Moreover, courts impose these restrictions on rights regardless of the relative seriousness of the offense or even the prospect of criminal prosecution. Indeed, some courts have denied these criminal procedural guarantees to youth in schools even where a uniformed police officer participated in the investigation.

While there has always been substantial disagreement within the scholarly literature over the extent to which constitutional rights should be restricted in the public school context, both sides of the debate share a common starting point: such restrictions must be justified, if at all, by pedagogical goals. For example, James E. Ryan has argued that courts grant special deference to school officials when—and only

---

17 See supra note 4 and accompanying text.
18 See supra note 5 and accompanying text.
19 See Wofford v. Evans, 390 F.3d 318, 326-27 (4th Cir. 2004) (rejecting requirement for probable cause where student was suspected of bringing firearm to school); Commonwealth v. Snyder, 597 N.E.2d 1363, 1369 (Mass. 1992) (rejecting Miranda warnings for questioning by school official where principal intended to turn over incriminating evidence of the student's drug dealing to police); State v. Tinkham, 719 A.2d 580, 583 (N.H. 1998) (holding that school officials need not provide Miranda warnings prior to questioning student suspected of dealing drugs); In re Harold S., 731 A.2d 265, 268 (R.I. 1999) (concluding school principal who conferred with police was not required to provide Miranda warnings prior to questioning the student).
20 The Supreme Court has expressly reserved the question of the standard for student searches conducted "in conjunction with or at the behest of law enforcement agencies." New Jersey v. T.L.O., 469 U.S. 325, 341 n.7 (1985). Absent such guidance, lower courts have split. Compare People v. Dilworth, 661 N.E.2d 310, 317 (Ill. 1996) (holding search conducted by school resource officer subject to reasonable suspicion standard rather than ordinary probable cause requirements), and In re Josue T., 989 P.2d 431 (N.M. Ct. App. 1999) (same), and R.D.S. v. State, 245 S.W.3d 356, 365 (Tenn. 2008) (same), and In re Angela D.B., 564 N.W.2d 682 (Wis. 1997) (rejecting probable cause requirement for search by police officer at the request of and in conjunction with school officials), with A.J.M. v. State, 617 So. 2d 1137, 1138 (Fla. Dist. Ct. App. 1993) (applying probable cause standard to search by school resource officer), and Patman v. State, 537 S.E.2d 118 (Ga. Ct. App. 2000) (applying probable cause to search by police officer on special detail to the school).
21 With respect to students' right to Miranda warnings prior to school interrogation, compare State v. Schloegel, 769 N.W.2d 130, 133-34 (Wis. Ct. App. 2009) (finding no custodial interrogation during questioning of student at school by police because if the student was "in custody at all, [he] was in the custody of the school and was not being detained by the police at the time"), with In re T.A.G., 663 S.E.2d 392, 396 (Ga. Ct. App. 2008) (concluding that involvement of school resource officer transforms interrogation by school principal into custodial interrogation requiring Miranda warnings). But see J.D.B. v. North Carolina, 131 S. Ct. 2394 (2011) (rejecting argument that interrogation of youth at school by police never qualifies as custodial interrogation). For scholarly discussion of the procedural rights applicable to school-based investigations involving police officers, see Holland, supra note 5, at 45-58; Kagan, supra note 10, at 316-20; Pinard, supra note 10, at 1080-90; Peter Price, When Is a Police Officer an Officer of the Law?: The Status of Police Officers in Schools, 99 J. CRIM. L. & CRIMINOLOGY 541, 560-67 (2009).
22 See supra note 6.
when—they are acting in their privileged role as educators and transmitters of knowledge.\textsuperscript{22} Similarly, Ann Proffitt Dupre has characterized the Supreme Court’s jurisprudence in school discipline cases as reflective of a larger debate about competing educational goals.\textsuperscript{23}

Central to this defense of restrictions on students’ rights is the assumed alignment of interests between students and school officials.\textsuperscript{24} Procedural protections guaranteed in the criminal context have been deemed unnecessary in the school context to the extent that the investigation and punishment of student misconduct—in stark contrast to the investigation and punishment of ordinary crime—is for the youth’s own educational benefit, teaching the importance of respect for others and acceptance of responsibility. Therefore, the standard calculus applicable outside the school discipline context—balancing the tradeoff between the individual interest and the competing collective or state interest—has been deemed inapplicable in schools. Analyzing restrictions on students’ privacy rights during school searches, Stephen Schulhofer has reasoned that “[b]oth the investigating authority and the person searched are participants in a shared mission,” rendering inapposite ordinary constitutional protections that “reflect[] a balance appropriate mainly to cases in which the private activity and public controls are poised in conflict.”\textsuperscript{25} Under this view, the convergence of interests between school official and student, unlike the adversarial interests of the adult criminal suspect and law enforcement, obviates the need for the robust protections guaranteed in the law enforcement context.

This part traces the development of this pedagogical theory in the three foundational cases in school discipline—\textit{Goss v. Lopez}, involving school suspension; \textit{Ingraham v. Wright}, involving corporal punishment; and \textit{New Jersey v. T.L.O.}, involving student searches. Each of these cases relied on the view that school discipline educationally benefits the punished youth to justify restrictions on students’ rights. Interestingly, although the Court frequently relies on social science evidence in determining educational rights in other contexts, most famously in footnote eleven of \textit{Brown v. Board of }

\textsuperscript{22} Ryan, supra note 6, at 1341. He continues, “the further a policy moves away from the core academic function of schools, the more likely the Court will apply traditional constitutional rules to judge the policy and strike it down.” \textit{Id.}

\textsuperscript{23} Dupre, supra note 6, at 64.

\textsuperscript{24} See supra note 7 and accompanying text.

\textsuperscript{25} Schulhofer, supra note 3, at 117-18.
these foundational school discipline cases are notable for the conspicuous absence of social science support for their conclusions. Rather, in these cases, members of the Court relied almost exclusively on personal intuitions regarding the operation of school discipline, even when those intuitions conflicted with empirical evidence properly before the Court. However, more recent cases demonstrate an increased willingness to factually assess the operation of school discipline to determine whether the judicial deference traditionally afforded to school officials remains warranted.

A. The Foundational Cases

Judicial reliance on the perceived educational value of school discipline to impose limits on students' procedural rights dates at least to *Goss v. Lopez,* decided in 1975, the first time the Supreme Court addressed the scope of these rights. During race-related tensions at a high school, lead plaintiff Dwight Lopez was in the school lunchroom when a group of black students entered and began overturning tables. Lopez claimed he immediately left the lunchroom and did not participate in any of the disruptive activities. After he was suspended for this incident, he filed a class action lawsuit arguing that the refusal to provide students with any opportunity to assert innocence and challenge a school suspension violated procedural due process rights.

On certiorari, a majority of the Court held that school discipline proceedings are subject only to minimal due process protections. Suspensions of up to ten days require only an “informal give-and-take” between the principal and the student, which need not occur prior to the suspension. Acknowledging that these limitations provide individuals facing a denial of education with fewer procedural protections than those afforded to individuals facing a denial of, for example, welfare benefits or a driver’s license, the Court emphasized that unlike other forms of state sanctions, school

---

29 Id. at 1285.
30 Id. at 1281.
31 Goss, 419 U.S. at 584.
discipline serves a pedagogical purpose designed for the benefit of the child: “Suspension is considered not only to be a necessary tool to maintain order but a valuable educational device.”\(^{32}\) The Court expressed concern that imposing additional due process requirements on school discipline would “destroy its effectiveness as part of the teaching process.”\(^{33}\)

Justice Powell’s dissent would have gone further, reasoning that the educational value of school discipline justifies the denial of any procedural due process protections.\(^{34}\) His position expressly emphasized what the majority had only implied: both the punished student and the disciplining school official have a shared interest in swift and informal punishment, rendering traditional due process protections inappropriate. Justice Powell stated, “When an immature student merits censure for his conduct, he is rendered a disservice if appropriate sanctions are not applied or if procedures for their application are so formalized as to invite a challenge to the teacher’s authority.”\(^{35}\)

Accusing the majority of “misapprehending the reality of the normal teacher-pupil relationship,” the dissent insisted that “[u]nlike the divergent and even sharp conflict of interests usually present where due process rights are asserted, the interests here implicated—of the State through its schools and of the pupils—are essentially congruent.”\(^{36}\)

Importantly, neither opinion cited any social science support for its assumptions about the benefits of school discipline—or could it. The record before the Court—far from validating the educational value of school suspensions—was replete with facts indicating that school suspensions harm students. Students testified at trial to the negative impact that the suspension and subsequent loss of instruction time had on their academic progress, and two prominent psychologists gave expert testimony regarding the adverse consequences of suspensions.\(^{37}\) Based on this evidence, the district court entered factual findings that school suspensions are harmful to students and may compromise academic achievement.\(^{38}\)

\(^{32}\) Id. at 580.

\(^{33}\) Id. at 583.

\(^{34}\) Id. at 585-86 (Powell, J., dissenting).

\(^{35}\) Id. at 593.

\(^{36}\) Id. at 591, 594.

\(^{37}\) Appellees’ Brief on the Merits at 12-13, 33-34, Goss, 419 U.S 565 (No. 73-898), 1974 WL 185915 (citing testimony from trial record).

\(^{38}\) Lopez v. Williams, 372 F. Supp. 1279, 1292 (S.D. Ohio 1973) (finding, as a matter of fact, that “[m]ost students respond [to suspensions] in one or more of the following ways: (1) The suspension is a blow to the student’s self-esteem. (2) The
appeal, the student-appellees cited numerous scholarly articles further demonstrating the educational harms associated with suspensions, and amicus briefs filed by the NAACP, the Children’s Defense Fund, and the ACLU likewise cited studies describing the negative repercussions of school suspensions on children, including reputational harm to the student, loss of instructional time, exacerbation of deviant behavior, lower high school graduation rates, and fewer future employment opportunities. 39 Yet the majority ignored these facts altogether. Justice Powell’s dissent acknowledged them, but dismissed them with little discussion as “generalized opinion evidence.”40 With no mention of the clearly erroneous standard applicable to the trial court’s findings of fact,41 Justice Powell summarily reached the contrary factual conclusion, that “[f]or average, normal children—the vast majority—suspension for a few days is simply not a detriment; it is a commonplace occurrence . . . it leaves no scars; affects no reputations; indeed, it often may be viewed by the young as a badge of some distinction and a welcome holiday.”42 In this way, restrictions on procedural due process rights in public schools rested on the unsupported factual contention that school discipline furthers the educational interests of the suspended student.

Two years later in Ingraham v. Wright43 the Court again invoked the perceived educational value of school discipline, this time to reject a constitutional challenge to abuse in the administration of corporal punishment. In Ingraham, a student who was slow to respond to a teacher’s instructions was hit
twenty times with a paddle, resulting in a hematoma requiring medical attention, while another student lost use of his arm for a week because of a teacher's paddling. Rejecting the students’ Eighth Amendment and due process claims, the Court reasoned that “since before the American Revolution,” corporal punishment has been viewed as necessary for the “moderate correction” of a child’s misbehavior and “for the proper education of the child.” The Court acknowledged that its holding meant that youth in schools enjoyed fewer protections than convicted criminals, as criminals subjected to corporal punishment would clearly be entitled to raise a constitutional challenge to that punishment. Nonetheless, the Court imposed a categorical distinction between punishment in the law enforcement context and punishment in the school discipline context, concluding that the “prisoner and the schoolchild stand in wholly different circumstances” with respect to the constitutional rights to which they are entitled.

As in Goss, the Ingraham Court made little effort to garner factual support for its assumptions about corporal punishment. At the trial level, plaintiff-schoolchildren submitted evidence of repeated physical abuse in the administration of corporal punishment, which would destroy whatever educational value might otherwise inhere in its use. The district court did not enter any factual findings about the credibility of testimony regarding these allegations, but concluded that—even if the testimony were credible—no relief would be granted. The Supreme Court, rather than engaging with the evidence of abuse, simply assumed this abuse was infrequent, thereby obviating the need for the requested procedural protections. Instead of invoking factual support, the majority relied on its “common-sense judgment that excessive corporal punishment is exceedingly rare in the public schools.” Criticizing the lack of evidentiary support for the majority’s claims, the dissenting four Justices accused the majority of relying on “mere armchair speculation” to justify denials of constitutional rights to schoolchildren.

44 Id. at 657.
45 Id. at 661-62, 664.
46 Id. at 669.
47 Id. at 664-71.
48 Id. at 658.
49 Id. at 677 n.45.
50 Id. at 690 n.7 (White, J., dissenting).
Finally, in New Jersey v. T.L.O., the Court relied on the purported educational value of school discipline and consequent alignment of interests between student and school official to limit the scope of Fourth Amendment protections available in public schools.\(^{51}\) In T.L.O., a school principal searched the purse of a student accused of smoking cigarettes; smoking was a violation of school rules but not of any criminal law. During the course of the search, the principal found items implicating the student in drug dealing, which he turned over to the police to be used against the student in subsequent delinquency proceedings.\(^{52}\)

Holding that school officials may search a student’s person or belongings absent the warrant or probable cause that would be required outside of the school context, the Court reasoned that to hold otherwise would compromise “the value of preserving the informality of the student-teacher relationship.”\(^{53}\) Justice Powell’s concurrence repeated his insistence from Goss that the alignment of interests presented in school discipline cases “make[s] it unnecessary to afford students the same constitutional protections granted adults and youths in a nonschool setting.”\(^{54}\) This alignment, he reasoned, sharply distinguished the teacher-student relationship from that of citizens and law enforcement officers:

Law enforcement officers function as adversaries of criminal suspects. These officers have the responsibility to investigate criminal activity, to locate and arrest those who violate our laws, and to facilitate the charging and bringing of such persons to trial. Rarely does this type of adversarial relationship exist between school authorities and pupils.\(^{55}\)

Similarly, Justice Blackmun’s concurrence reasoned that the educational role of teachers excused them from ordinary Fourth Amendment standards: “A teacher’s focus is, and should be, on teaching and helping students, rather than on developing evidence against a particular troublemaker.”\(^{56}\)

Again, the Court in T.L.O. appeared to rely on commonsense intuitions about what is good for the child, rather than engaging in a fact-based inquiry. Indeed, the contention

---

\(^{51}\) 469 U.S. 325 (1985).
\(^{52}\) Id. at 328-29.
\(^{53}\) Id. at 340.
\(^{54}\) Id. at 348 (Powell, J., concurring).
\(^{55}\) Id. at 349-50.
\(^{56}\) Id. at 353 (Blackmun, J., concurring).
regarding an alignment of interests between student and school official, purported to distinguish the relationship from that between police officer and suspect, was undercut by the facts of *T.L.O.* itself: the school official ultimately referred the student to law enforcement and the juvenile court. Yet, rather than performing any empirical inquiry into the frequency with which the interests of accused students conflict with those of school officials, the Court deemed that rate to be sufficiently “rare” to justify restricting students’ rights. Underscoring the absence of evidentiary support for the majority’s claims, Justice Brennan’s dissenting opinion characterized the majority’s rationales as “brief nods by the Court in the direction of a neutral utilitarian calculus while the Court in fact engages in an unanalyzed exercise of judicial will” designed to “reach[] a predetermined conclusion acceptable to this Court’s impressions of what authority teachers need.”

In these foundational cases, then, the Court has relied on the assumed educational value of school discipline and the purported convergence of interests between school official and student to conclude that ordinary procedural protections are inapplicable in schools.

It is true, however, that these cases did not rely exclusively on these premises to justify limits on students’ procedural rights. Rather, the Court has suggested an additional justification for such restrictions—the weighty interests of other students in maintaining an environment conducive to learning. *Goss* emphasized that the maintenance of order and discipline is “essential if the educational function is to be performed,” and *T.L.O.* underscored the heavy weight

---

57 Id. at 372 (Stevens, J., concurring in part and dissenting in part) (challenging the assumption that law enforcement and school discipline categorically differ by pointing out that T.L.O. herself was subject to prosecution as a delinquent as a result of the search); see also Tribe, supra note 7, at 312 (“[E]ven if one conceives that there is no inherent clash in the interest of teacher and student, does it not remain possible for them to clash in fact . . . ?”).


59 See Ryan, supra note 6, at 1341, 1411-14 (interpreting limits to constitutional rights in school discipline cases as resting on the view that they are necessary to “preserve an atmosphere that is safe and conducive to learning” and “to maintain discipline in order to transmit academic knowledge”); see also Tribe, supra note 7, at 312 (“[E]ven if in general the teachers’ and child’s interest truly converge, at the moment of suspension convergence must surely turn into clash: the teacher is saying that the best interests of other students will be served by this particular student’s suspension.”).

of the “[s]chool’s interest in maintaining an environment where learning can take place.”

As a purely doctrinal matter, however, this alternative justification proves less than satisfactory. While few would contest that the state interest in providing a functional educational environment is significant, the Court has not attempted to explain why this interest would outweigh the individual student’s interest if those interests are in fact in conflict. Outside the school context, the state interest in preventing violent street crimes is of course significant, yet it does not outweigh the individual interest in receiving full procedural protections. It is not at all clear why this calculus balancing competing interests between state and individual should not apply in the school context as well. Perhaps recognizing this deficiency, the Court has relied on this alternative argument only to buttress its more primary assumptions regarding the perceived educational value to the investigated or punished student and the purported absence of adversarial interests in the context of school discipline.

B. Recent Cases and the Consideration of Evidentiary Support

Notwithstanding the Supreme Court’s earlier reluctance to engage with empirical evidence in this area, more recent cases suggest an increased willingness to assess the actual operation of school discipline in particular jurisdictions to determine whether the traditional deference remains warranted. In these opinions, the Court has begun to consider the purpose and impact of the investigation and punishment of students. Moreover, where evidence suggested that a particular disciplinary practice negatively impacts the education of students, the Court has concluded that judicial interference is warranted.

*Vernonia School District v. Acton* and *Board of Education v. Earls* examined suspicionless drug testing in

---

61 469 U.S. at 326. More recently, in sustaining the suspicionless drug testing of student-athletes, the Court in *Vernonia School District v. Acton* emphasized that schools routinely require students to submit to invasions of their privacy not only “for their own good,” but also for “that of their classmates.” 515 U.S. 646, 656 (1994).


public schools. In Acton, the Court sustained such tests for student athletes, but only after conducting a factual assessment of how the search policy operated in the particular school at issue and the actual impact of the policy on students. Rather than simply assume, as it had done in prior cases, that the investigation of students for drug use benefited students, the Court analyzed empirical evidence of the dangers of drug use to students, and then sustained the policy only after it satisfied itself that the particular drug tests at issue were “undertaken for prophylactic and distinctly non-punitive purposes,” as school policy dictated that the results of any positive tests would not be turned over to law enforcement authorities. Thus, the Court engaged in a factual, school-specific inquiry before concluding that the interests of the student and school officials were aligned, rendering unnecessary the individualized suspicion that would be required outside of the school context. Similarly, in Earls, the Court sustained suspicionless drug testing of students participating in extracurricular activities only after engaging in a factual inquiry regarding the operation of the drug testing policy in the particular school and concluding that the drug tests benefited rather than harmed the tested student in large part because under school policy the results were not turned over to law enforcement or used to punish the student.

Even more explicitly, in Safford v. Redding, the Court engaged in a factual inquiry to test the long-held categorical assumption that the investigation and punishment of student wrongdoing benefits the student. In Redding, school officials strip searched a thirteen-year-old student accused of bringing prescription-strength ibuprofen to school. Although the Court affirmed T.L.O.’s holding that the school setting requires only a reasonable suspicion to justify a student search, it nonetheless considered the actual impact of the search on the student and concluded that the student’s Fourth Amendment rights had been violated. Rather than assuming, as it had in T.L.O., that school searches categorically advance the educational interests

---

64 536 U.S. 822 (2002).
65 515 U.S. at 658 & n.2.
66 By focusing on the operation of a school discipline practice in a particular school, the Court appeared to treat the impact of school discipline on youth as an adjudicative fact rather than a legislative one. See Kenneth Culp Davis, An Approach to Problems of Evidence in the Administrative Process, 55 Harv. L. Rev. 364, 402-03 (1942).
67 536 U.S. at 833.
of the searched student, the Court cited the amicus brief for the National Association of Social Workers and an article from the *Journal of School Psychology* to emphasize the negative psychological impact of a strip search on youth. Redding thus suggests an increased willingness by the Court to evaluate—rather than simply assume—the impact of investigations and punishments on students in determining whether the rights of those students warrant restriction.

II. **EMERGING MODES OF SCHOOL DISCIPLINE**

As set forth in the previous part, doctrinal restrictions on the rights of youth accused of school-based misconduct rest on a series of factual assumptions about the manner in which school discipline operates and the educational value of this discipline. Specifically, the Supreme Court has limited these rights in public schools on the ground that the investigation and punishment of students is intended for the students' educational benefit. Thus, it has reasoned, the adversarial relationship characteristic of law enforcement encounters outside of the school context simply does not apply to investigations of student misconduct to necessitate the same level of procedural protections. Recently, the Court has appeared willing to assess the facts underlying these assumptions. Based on this development, this part evaluates empirical evidence of contemporary school discipline practices and their educational impact on students.

As others have observed, the past two decades witnessed a dramatic shift in public discourse, with an increasing focus on school safety and crime prevention. This part evaluates how that shift has manifested in the operation of school discipline. It first provides an overview of policy developments that led to a convergence of school discipline and law enforcement. Next, it reviews a series of recently published studies measuring the extent to which school officials rely on law enforcement to maintain order. Finally, it analyzes social science research

---

69 *Id.* at 2641-42 (citations omitted).


examining the likely impact of law enforcement referrals on educational outcomes. Based on these findings, this part concludes that, at least in some jurisdictions, school discipline no longer serves the educational interests that traditionally justified insulating it from ordinary constitutional requirements.

A. The Shift Toward School Crime Control

Following a series of school shootings in the 1990s, a widespread sense of panic descended on public schools across the nation. A year after the shooting at Columbine High School in 1999, almost two-thirds of Americans reported feeling it was somewhat likely or very likely that a school-shooting spree would occur in their community.\(^72\) As one scholar put it, “policy makers reacted abruptly to what they perceived to be a huge swing in public opinion: a moral panic swept the country as parents and children suddenly feared for their safety at school.”\(^73\) Importantly, this fear extended to predominantly white suburban and rural areas; school violence was no longer contained in the predominantly minority, low-income, inner-city neighborhoods traditionally associated with crime. Republican Senator Ben Nighthorse Campbell of Colorado, for example, stated,

> These recent school shootings have occurred in suburbs, small towns, and major metropolitan areas all across our nation. They have shattered the myth that school violence is a problem solely confined to the inner cities. Events now clearly show that the potential for serious and deadly school violence is everywhere.\(^74\)

Indeed, a poll conducted in 1999 found that suburban and rural parents were more likely than minority parents to feel a school shooting was somewhat likely or very likely to occur in their communities.\(^75\)

This fear resulted in the deployment of large numbers of police officers to patrol public school hallways. Today, nearly

---


half of all public schools have assigned police officers,\textsuperscript{76} and 60 percent of high school teachers report armed police officers stationed on school grounds.\textsuperscript{77} Often with the help of federal funding, school-based police officers, frequently referred to as “school resource officers,” are the fastest-growing segment of law enforcement.\textsuperscript{78} These officers’ roles vary significantly across schools, with some charged primarily with enforcement of criminal laws, while others are focused on mentoring, counseling, and teaching duties.\textsuperscript{79}

The reliance on law enforcement to maintain school order is not limited to jurisdictions with school resource officers. Jurisdictions lacking the resources to hire full-time police personnel nonetheless may regularly summon the local police department through calls for service. Indeed, rapidly spreading “zero-tolerance” policies mandate that school officials call the police any time certain predetermined infractions are committed. In Rhode Island, a statewide policy requires school principals to report all school fights to the police for criminal prosecution.\textsuperscript{80} Alabama requires all principals to notify law enforcement any time a person violates district policies regarding physical harm or threats of harm.\textsuperscript{81}

Similar mandates have been adopted at the local district level as well. The Atlanta Public School System, for example, maintains a zero-tolerance policy requiring school officials to immediately report to the police any student involved in drug-related offenses or gang activity.\textsuperscript{82} Chicago Public Schools began requiring school officials to notify police of all burglary, aggravated assault, and gang activity offenses, while providing administrators with discretion to refer students to the police for

\textsuperscript{76} Barbara Raymond, U.S. DEP’T OF JUSTICE, OFFICE OF CMTY. POLICING SERVS., ASSIGNING POLICE OFFICERS TO SCHOOLS (2010).
\textsuperscript{77} Hirschfield, supra note 9, at 82.
\textsuperscript{79} Peter Finn & Jack McDevitt, NATIONAL ASSESSMENT OF SCHOOL RESOURCE OFFICER PROGRAMS 43 (2005).
\textsuperscript{80} 140 Cong. Rec. 10281 (1994).
lesser offenses such as gambling, forgery, or petty theft.\textsuperscript{81} The Houston Independent School District requires school principals to notify the police any time there are reasonable grounds to believe that a student has engaged in any criminal offense at school.\textsuperscript{84} The East Carroll Parish School System in Louisiana, a small, rural district, requires that law enforcement remove and file charges against any student age twelve or over who is an aggressor in a fight.\textsuperscript{85} Guilford County Schools system in North Carolina requires that school officials call the police every time an aggravated assault, sexual offense, weapons offense, or drug possession is suspected.\textsuperscript{86} Nelson County Public Schools system in Virginia requires schools to refer to the police all instances of drug offenses, violence, interference with school authorities, and driving without a license on campus.\textsuperscript{87}

As the scope of these zero-tolerance policies suggests, the infractions for which students are referred to law enforcement have expanded considerably. Numerous states criminalize the offense of disrupting school activities\textsuperscript{88} or

talking back to teachers.” In 1994, the South Carolina Attorney General issued an opinion stating that students who fight in school, fail to leave school grounds upon request, or use foul or offensive language toward a principal or teacher are subject to criminal prosecution.

As a result of these policy developments, schoolchildren today are more likely to be arrested and prosecuted for school-based misconduct than they were a generation ago. According to the Federal Advisory Committee on Juvenile Justice, the number of referrals to the juvenile justice system for relatively minor school-based conduct is on the rise. Given this shift toward criminalization, the administration of school discipline appears to be increasingly adversarial.

**B. Rates of School-Based Referrals to Law Enforcement**

Simply acknowledging that law enforcement intersects with school discipline more often than it did in the past does not fully resolve the issues posed by current doctrine. Even *T.L.O.* conceded that law enforcement sometimes overlaps with school discipline; it concluded, however, that the incidence of convergence is sufficiently rare to warrant treating the two institutions distinctly. The key issue, then, is the scope of the convergence. This section examines empirical studies measuring school-based student referrals to law enforcement to assess the extent to which school discipline remains discrete from law enforcement. A school-based referral to law enforcement may take several forms. I use the term “school-based law enforcement referral” to refer to incidents in which a youth is arrested at school or for school-related conduct, which sometimes but not always results in the youth being processed through the juvenile or criminal court system. It also includes incidents in which the youth is not arrested, but is processed...
through the juvenile or criminal justice systems or is required to respond to a criminal citation.

The analysis draws from several recently published studies. None of these studies purports to determine why some schools rely on law enforcement to maintain discipline while others do not. A wide range of causal factors may be at play, ranging from attitudes of school officials to limits on funding for traditional classroom management techniques. Determining the underlying causes for differing rates of referrals is beyond the scope of this article. Rather, this article seeks to provide a descriptive assessment of school officials’ reliance on law enforcement.

The studies examined here differ somewhat in methodology and jurisdictions examined. Nonetheless, collectively, they provide useful guides to assess the degree of intersection between school discipline and law enforcement. Specifically, the studies provide data on three important indicators, namely: (1) the share of juvenile law enforcement referrals that stem from school-based misconduct, (2) the number of school-based law enforcement referrals per one thousand enrolled students per year, and (3) the types of offenses for which students are referred to law enforcement.

**Percentage of law enforcement referrals resulting from school-based misconduct:** The percentage of youth referrals to law enforcement stemming from school-based misconduct provides a useful empirical measure of the extent to which school discipline and law enforcement converge. A finding that school-based referrals represent a large share of the overall number of law enforcement referrals would challenge the doctrinal view that school discipline is categorically discrete from law enforcement. In addition, it would suggest that a relatively large share of juvenile arrests and investigations occurs without the guarantees of *Miranda* warnings and probable cause, since these protections generally are not constitutionally required for school-based investigations.93

According to a recent assessment of the National Incident Based Reporting System, which maintains records of crime incidents from 20 percent of the nation’s police agencies, approximately one in six (17 percent) juvenile arrests stems from school-based misbehavior.94 This figure casts doubt on the doctrinal contention that the investigation and punishment of

---

93 See *supra* notes 4-5 and accompanying text.
94 Cook, Gottfredson & Na, *supra* note 9, at 319, 332.
school misconduct categorically differ from investigation and punishment for law enforcement purposes. Moreover, data from jurisdiction-specific studies suggest an extremely high degree of variance across jurisdictions. State-level data show that the share of juvenile court cases that originate from school-based misconduct ranges from a low of 4 percent to a high of 43 percent. These data, limited to formal referrals to juvenile court, do not provide a precise measure of the share of youth law enforcement referrals that are school-based. They omit incidents in which a youth is referred to law enforcement through an arrest at school, but charges are dropped before a case is filed in juvenile court. They also omit school-based law enforcement referrals that result in charges being filed in adult criminal court rather than in juvenile court.

The Annual Report for the North Carolina Department of Juvenile Justice indicates that 43 percent (16,140 out of 37,584) of offenses that result in referral to the juvenile justice system are school based. A recently published survey by education scholar Michael Krezmien and colleagues finds far lower rates in the five states for which it was able to obtain data: Arizona, Hawaii, Missouri, South Carolina, and West Virginia. Among those states, West Virginia exhibited the highest share of court referrals that were school based, with approximately 17 percent of juvenile cases originating in schools, while Hawaii exhibited the lowest share, with only 4 percent originating in schools. While it is not clear that the North Carolina report and the multistate study employed a sufficiently similar methodology to permit precise comparison, the results of the two reports suggest wide variance across states in the extent to which law enforcement is used to maintain school order.

---

95 See infra notes 97-100 and accompanying text.
96 The number of school-based law enforcement referrals that result in charges being filed in adult criminal court is likely to be particularly large in jurisdictions such as North Carolina, where youth aged sixteen or older are automatically processed through the adult criminal courts, regardless of the offense. N.C. GEN. STAT. ANN. § 7B-1604 (West 2004).
98 Krezmien et al., supra note 9, at 283.
99 See id.
100 Similarly, according to a recent study by the Council for State Governments, only 6 percent of youth referrals to the juvenile court system in Texas (5349 out of 85,548 formal referrals) came directly from schools. COUNCIL OF STATE GOVT'S JUSTICE CTR. & PUB. POLICY RESEARCH INST., BREAKING SCHOOLS' RULES: A STATEWIDE STUDY OF HOW
A study of county-level data in Florida suggests that even within a given state, the extent to which law enforcement converges with school discipline varies considerably.\footnote{Breaking Schools’ Rules, supra note 97.} According to the Florida Department of Juvenile Justice, the statewide share of juvenile court referrals that stem from conduct on public school grounds, at a school bus stop, or at a school event is approximately 15 percent.\footnote{FLA. DEP’T OF JUVENILE JUSTICE, OFFICE FOR PROGRAM ACCOUNTABILITY, DELINQUENCY IN FLORIDA’S SCHOOLS: A SIX-YEAR STUDY (2010) [hereinafter DELINQUENCY IN FLORIDA’S SCHOOLS], available at http://www.djj.state.fl.us/research/School_Referrals/FY-2009-10-Delinquency-in-Schools-Analysis.pdf.} The rates for individual counties, however, diverge significantly from this baseline. In Gulf and Dixie Counties the percentage of delinquency referrals for school-based misconduct was only 8 percent, while in other counties more than a quarter of delinquency cases came from schools: Okeechobee (29 percent), St. Lucie (27 percent), Hamilton (27 percent), Jackson (26 percent), and Marion Counties (26 percent).\footnote{Id. at 4.} Again, these data are limited to cases resulting in a referral to juvenile court; they omit cases in which students are arrested but released before juvenile court charges are filed, and they omit cases in which youth are processed through the adult criminal justice system.\footnote{Id. at 2.}

Although differences in methodology across the studies limit to some degree the comparability of these data, the studies suggest that school discipline practices vary widely across jurisdictions. At the low end, in some jurisdictions as few as 4 percent of juvenile court cases may originate in schools; at the high end, as many as 43 percent may originate from schools.\footnote{See supra notes 97-100 and accompanying text.} Nonetheless, the data demonstrate that in at least some jurisdictions it has become difficult to defend the claim that school discipline differs categorically from law enforcement, or that school discipline serves educational rather than police purposes. The heavy reliance on law enforcement to maintain school order suggests that one can no longer assume a nonadversarial, benevolent relationship between school disciplinarian and student.
Number of school-based law enforcement referrals per one thousand enrolled students per year: Data showing rates of school-based law enforcement referrals per one thousand students per year likewise permit an evaluation of the claim that school discipline differs categorically from law enforcement. Where the rates of school-based arrest and referral are relatively high, the claim that the investigation and punishment of school misconduct furthers educational rather than law enforcement purposes becomes harder to defend. A number of recently published studies provide data on this measure. These data show that, as with the proportion of juvenile court referrals stemming from school-based misconduct, rates of school-based law enforcement referrals per one thousand enrolled students vary significantly across jurisdictions.\footnote{These figures do not purport to estimate the likelihood that a given student will be subject to school-based law enforcement, because some students may have been referred to law enforcement more than once.}

The Florida Department of Juvenile Justice’s Annual Report calculates the number of school-related delinquency referrals per one thousand students enrolled in grades six through twelve; statewide, there were thirteen (13) school-related delinquency referrals per one thousand middle and high school students.\footnote{\textit{Delinquency in Florida’s Schools}, supra note 101, at 5.} Rates of school-based delinquency referrals varied dramatically across counties within Florida, though, from a low of only four (4) school-related delinquency referrals per one thousand students in Lafayette and Nassau Counties, to a high of forty-two (42) and fifty (50) referrals per one thousand students in Hamilton and Putnam Counties, respectively.\footnote{The multistate study found a smaller range, from a low of two (2) school-based juvenile court cases per one thousand students enrolled in Hawaii, to a high of nine (9) cases per one thousand students in Missouri. These data, which unlike the Florida report include elementary school students, understate the arrest rate for middle and high school students who are arrested at higher rates than their younger counterparts. Krezmien, supra note 9, at 279.} These figures understate the actual incidence of school-based law enforcement referral because they omit instances in which a student is arrested at school but released before charges are filed in juvenile court, and they omit cases referred to adult criminal court.

Even greater variability exists in Texas. The advocacy organization Texas Appleseed recently published its analysis of rates of school-based arrests in seventeen districts where data
were available. The study, analyzing data from seventeen districts representing 13 percent of the state’s student body, found that at the low end of the spectrum there were one-and-a-half (1.5) and two (2) school-based arrests per one thousand enrolled students in Castleberry and Wichita Falls districts, respectively. At the high end of the spectrum, East Central reported fifty-one (51) arrests for every one thousand students. These data also understate the actual incidence of school-based law enforcement referral, as they omit instances in which a student is referred to juvenile court without being arrested at school, such as when the student receives a summons to appear in lieu of arrest. Unlike the figures for Florida, these figures include elementary school students as well as middle and high school students; had the data excluded elementary school students, rates of school-based arrests per one-thousand enrolled middle and high school students would be higher.

The Texas Appleseed study also provides data on the issuance of misdemeanor tickets in public schools. These citations require the recipient to appear in municipal court and may result in fines of up to $500; a failure to appear subjects the individual to a bench warrant for arrest. Data from the twenty-six districts for which data were available again show wide disparities in the administration of punishment. At the lower end of the scale, United and Humble districts each issued fourteen (14) criminal citations per one thousand enrolled students, while at the high end, Galveston issued 109 citations per one thousand enrolled students. In Texas alone, then, there are as few as one-and-a-half (1.5) to as many as fifty-one (51) school-based arrests per one thousand enrolled students.

---

110 Id. at 104-05.
111 Id.
112 Id. at 114 (documenting data showing that majority of school-based arrests are for high school students).
113 Id. at 67-96.
114 Id. at 69.
115 Id. at 67-96.
116 Id. at 77-78 (indicating 522 tickets issued in United Independent School District, which enrolls 37,671 students, and 431 tickets issued in Humble Independent School District, which enrolls 31,144 students).
117 Id. at 77 (indicating 921 tickets issued in the Galveston Independent School District, which enrolls 8,430 students).
students, and as few as fourteen (14) and as many as 109 misdemeanor citations per one thousand of these students.\footnote{118 See supra notes 111-19 and accompanying text.}

**Types of behaviors for which students are subject to law enforcement referral:** An examination of the types of behaviors for which students are referred to law enforcement provides another useful indicator to evaluate the factual supposition that school discipline differs categorically from the criminal process. Where law enforcement is being deployed to address student behavior that traditionally would have been handled more informally—through the imposition of after-school detention or suspension—the resulting adversarial relationship belies the claim that the intervention pedagogically benefits the punished student.

The South Carolina Department of Juvenile Justice’s Annual Report indicates that “disturbance of schools” represents the single most frequent offense resulting in a referral to juvenile court.\footnote{119 S.C. DEP’T OF JUVENILE JUSTICE, 2009-2010 ANNUAL STATISTICAL REPORT 4 (2010), available at http://www.state.sc.us/djj/pdfs/2010%20Annual%20Statistical%20Report.pdf.} Conduct resulting in a charge of “disturbance of schools” may not amount to conduct that would result in an assault charge or other more serious offenses. These data indicate that South Carolina schools heavily rely on law enforcement and juvenile courts to handle conduct that would not amount to a crime outside of the school context.

According to the Florida Department of Juvenile Justice, two-thirds of all school-related delinquency referrals involved misdemeanors, while one-third involved felonies.\footnote{120 DELINQUENCY IN FLORIDA’S SCHOOLS, supra note 101, at 8.} Misdemeanor assault and battery and disorderly conduct violations represented the largest segment of school-based delinquency referrals, at 21 percent and 15 percent, respectively.\footnote{121 Id.} Weapons offenses counted for approximately 5 percent of all school-related delinquency referrals.\footnote{122 Id.}

The Texas Appleseed study found that among the twenty-two districts that disaggregated criminal citations by offense, more than half the tickets were for disorderly conduct (e.g., profanity, offensive gesture, or fighting) or disruption of class or transportation.\footnote{123 TEXAS APPLESEED, supra note 109, at 82.} An additional 10 percent of the tickets...
were for violations of curfew or the Student Code of Conduct. Among the eleven school districts that disaggregated data on school-based arrests by offense, 24 percent of school-based arrests were for disorderly conduct.

Anecdotal newspaper accounts provide further corroboration that some schools rely on police to handle relatively minor forms of student misbehavior. In Lucas County, Ohio, the majority of school-related referrals to juvenile court were for disruptive conduct, while only approximately 2 percent were for more serious incidents such as assaulting a teacher or taking a gun to school. In Lafayette Parish, Louisiana, 46 percent of school-based arrests were for disturbing the peace or simple assault or battery, while 4 percent were for weapons or drug offenses. According to the presiding family court judge in Birmingham, Alabama, only approximately 7 percent of school-based arrests involved offenses that actually warranted arrest, such as weapons offenses or other felonies. In some jurisdictions, then, school officials appear to have delegated their traditional authority to handle common forms of student misconduct—such as those involving disruptive behavior or fights—to law enforcement.

Data from these studies suggest that in some, but not all, jurisdictions it has become more difficult to claim that students accused of misconduct are investigated and punished for their own educational benefit, or that efforts to maintain school order are justified by pedagogical goals, not law enforcement ones. Rather, at the high end of the range, up to 43 percent of youth referrals to law enforcement involve school misconduct, and there are up to fifty (50) school-based referrals to juvenile court, fifty-one (51) school-based arrests, and 109 criminal citations for every one thousand students enrolled in public schools per year. Nonetheless, these figures are not typical of all jurisdictions. Indeed, in some jurisdictions, as few as 4 percent of juvenile court cases stem from school

124 Id.
125 Id. at 107.
129 See supra notes 97, 109, 113, 119 and accompanying text.
misconduct, and there may only be four (4) school-based court referrals, one (1) or two (2) school-based arrests, and fourteen (14) criminal citations per one thousand students per year.\textsuperscript{130}

C. Educational Impact of Law Enforcement Referrals

The increasing reliance on law enforcement referrals to maintain school order challenges the assumption that school discipline furthers educational interests. Whatever might be said about the pedagogical value of suspensions or other more traditional forms of school discipline, the available social science shows that referring a student to law enforcement has

\textsuperscript{130} See supra notes 100, 109, 111, 118 and accompanying text. An examination of the causes for these disparities is beyond the scope of this article. Nonetheless, it is worth noting that the disparities are not entirely randomized. As sociologist Paul Hirschfield has observed, “criminalization in middle class schools is less intense and more fluid than in the inner-city, where proximate or immediate crime threats are overriding concerns.” Hirschfield, supra note 9, at 84; see also Maureen Carroll, Comment, Educating Expelled Students After No Child Left Behind: Mending an Incentive Structure that Discourages Alternative Education and Reinstatement, 55 UCLA L. REV. 1909, 1934-37 (2008) (discussing racial disparities and discrimination in school discipline); Heath Cobb, Separate and Unequal: The Disparate Impact of School-Based Referrals to Juvenile Court, 44 HARV. C.R.-C.L. L. REV. 581, 581 (2009) (same); Elizabeth E. Hall, Criminalizing Our Youth: The School-to-Prison Pipeline v. the Constitution, 4 S. REG’L BLACK L. STUDENTS ASS’N L.J. 75 (2010) (same). A qualitative study conducting interviews with law enforcement officers deployed to public schools across Massachusetts concluded that larger, urban school districts rely on arrests to maintain school discipline more heavily than suburban and rural school districts; officers stated that in urban school districts, school officials prioritized sending a “get tough” message, while those in affluent suburban schools with predominantly white populations prioritized preserving the reputation of their students and the school. Thurau & Wald, supra note 8, at 988, 1010. Given the disproportionate representation of minority students in urban schools, it may come as no surprise that students of color bear the brunt of the trend toward increased criminalization. According to the multistate study described in the text, the likelihood of school-based law enforcement referrals for Latino students in Arizona is three times higher than for white students; the rate for black students is twice as high as for white students. Krezmien et al., supra note 9, at 14. The Advancement Project reports similar trends, finding that black youth are more than two times more likely to be referred to law enforcement at school than white students in Colorado, two-and-a-half times more likely in Florida, and three-and-a-half times more likely in Philadelphia. ADVANCEMENT PROJECT, TEST, PUNISH, AND PUSH OUT: HOW “ZERO TOLERANCE” AND HIGH-STAKES TESTING FUNNEL YOUTH INTO THE SCHOOL-TO-PRISON PIPELINE 19 (2010), available at http://www.advancementproject.org/sites/default/files/publications/rev_fin.pdf. Moreover, a report by the American Civil Liberties Union suggests these racial disparities cannot be blamed exclusively on differences across school districts or differences in students’ behavior. It found that black students involved in physical altercations are twice as likely to be arrested at school in the same district as white students who commit the same acts; likewise, black and Latino students who commit drug offenses at school are ten times more likely to be arrested than white students who commit drug offenses in the same district. ACLU, HARD LESSONS: SCHOOL RESOURCE OFFICER PROGRAMS AND SCHOOL-BASED ARRESTS IN THREE CONNECTICUT TOWNS 26 (2008), available at http://www.aclu.org/files/pdfs/racialjustice/hardlessons_november2008.pdf.
negative educational consequences not only on the youth referred, but also likely on the larger student body.

1. Educational Impact on the Punished Student

Behavioral theories posit three competing models of the relationship between harsh punishments and youth outcomes: deterrence theory, propensity theory, and labeling theory.\textsuperscript{131} Deterrence theory, as its name suggests, posits that formal behavioral interventions deter youth from future deviant behavior. Propensity theory suggests that harsh punishments neither encourage nor discourage future deviant conduct; any relationship between the punishment and future behavior is correlative rather than causative, because the same inherent traits that led to the first instance of misconduct will lead to future deviance. Conversely, labeling theory suggests that harsh punishments for youth will actually increase future misconduct by labeling the youth as a deviant and creating a deviant self-concept with potentially life-altering consequences; others may also come to identify the youth as a deviant, thereby foreclosing opportunities that would otherwise have been available. Among the three competing theories, only deterrence theory supports the doctrinal assumption that punitive forms of school discipline such as a school-based arrest or law enforcement referral further the educational interests of the punished youth.

In fact, the available empirical evidence lends no support for the deterrence theory with respect to law enforcement referrals for school-based offenses. On the contrary, social science consistently shows that a law enforcement referral has significant negative consequences on youth educational outcomes. Among the more recent research, a 2006 study by criminologist Gary Sweeten assessed the relationship between law enforcement referral and educational attainment. Using data from the National Longitudinal Survey of Youth, a nationally representative sample, the study found that a first-time arrest during high school years nearly doubles the likelihood of dropping out of high school; an arrest coupled with a court appearance quadruples the likelihood.\textsuperscript{132} The magnitude of this effect holds, even after controlling for other factors thought to influence dropout rates including being held

\textsuperscript{132} Id. at 473.
back a grade, living in a single-parent household, poor prior academic performance, and rates of delinquent conduct.\textsuperscript{133} Similarly, a 2009 study by sociologist Paul Hirschfield assessed the impact of a first-time arrest on high school dropout rates in Chicago.\textsuperscript{134} Drawing a sample of students in Chicago Public Schools with high concentrations of low-income and minority students, it found that those who were arrested in ninth or tenth grade were six to eight times more likely to drop out of high school as classmates who were not arrested, even after controlling for variables including prior delinquency, peer delinquency, truancy, academic achievement, and anger control.\textsuperscript{135} A number of other studies have drawn similar conclusions regarding the negative impact of arrest on high school graduation rates.\textsuperscript{136}

None of these studies specifically examines the impact of an arrest when it occurs on school grounds or for school-related conduct. Yet, there is no reason to think that the negative educational impact of school-based arrest would be any less than the negative educational impact of arrest generally. On the contrary, given the importance of school officials and students maintaining positive relationships, one would expect that the negative impact of an arrest would be exacerbated when the arrest occurs at school. Further research should be conducted on this question. Nonetheless, in light of the findings to date, the use of school-based law enforcement referrals cannot currently be defended on the ground that it educationally benefits the referred youth.

2. Educational Impact on Other Students

Moreover, there is little empirical support for the claim that the use of law enforcement to maintain school order accrues educational benefits to the larger student population. It may well be true that if one student persistently disrupts the classroom, removal of that student enhances the remaining students’ ability to learn.\textsuperscript{137} However, there is no evidence

\textsuperscript{133} Id. at 478.
\textsuperscript{135} Id. at 368.
\textsuperscript{136} Id. at 370.
\textsuperscript{137} See Cook et al., supra note 9, at 372 (“Clearly, removing troublemakers from school helps maintain an environment more suitable for learning for the remaining students.”).
suggesting that referring the student to law enforcement specifically—in lieu of or in addition to some other mechanism such as traditional suspension—improves the educational climate for the remaining students. Indeed, a recent meta-analysis of 178 individual studies assessing the effectiveness of different school-based disciplinary interventions found no evidence that the use of arrest and juvenile courts to handle school disorder reduces the occurrence of problem behavior in schools. Some scholars have reasoned that, by creating adversarial and distrustful relationships between law enforcement and school authorities on the one hand, and the student body on the other, coercive police-like interventions may actually increase school disorder. Education scholars Matthew Mayer and Peter Leone analyzed data from the National Crime Victimization Survey to assess the relationship between coercive school security measures and educational climate and found that restrictive measures such as the use of security personnel, metal detectors, and locker searches were not only associated with higher levels of school disorder, but also possibly caused that disorder. Based on these findings they concluded, “creating an unwelcoming, almost jail-like, heavily scrutinized environment, may foster the violence and disorder school administrators hope to avoid.” Similarly, one criminologist recently expressed concern that “aggressive security measures produce alienation and mistrust among students” and such measures “can disrupt the learning environment and create an adversarial relationship between school officials and students,” while another criminologist suggested that the use of aggressive law enforcement tactics in schools “may cause students to distrust educational and law enforcement authorities which could motivate students to engage in greater delinquency.” Far from suggesting that law enforcement referrals improve the educational climate for remaining students, the limited evidence to date has led experts to conclude that such referrals likely compromise educational goals.

The primary doctrinal justifications for restricting the procedural rights of youth when they are investigated or

138 See id. at 369.
141 Brown, supra note 9, at 599.
punished for misconduct—that such investigations and punishments serve the youth’s educational interests and that they differ categorically from law enforcement—prove considerably less persuasive in light of evidence of contemporary school discipline practices and their likely educational impact on students. Whatever might be said about the more traditional school discipline practices of suspension or paddling, it can hardly be argued that school-based arrest is “a valuable educational device” or has “long been an accepted method of promoting good behavior and instilling notions of responsibility . . . into the mischievous heads of school children.” In jurisdictions where school officials frequently remove students from schools through formal arrest or the filing of a delinquency petition, one can no longer claim that the interests of the investigating school official and the student are aligned rather than adversarial. The next part explores the implications of these findings.

III. INCORPORATING CONSIDERATION OF THE EDUCATIONAL IMPACT OF SCHOOL DISCIPLINE

The developing body of empirical evidence in the preceding part challenges the doctrinal justification for denying procedural protections to youth who are accused of misconduct in schools. To the extent school discipline increasingly takes the form of law enforcement referrals, it can no longer be justified by the educational benefits it confers on the child or its purportedly nonadversarial nature. Those rationales for insulating traditional forms of school discipline from constitutional protections simply no longer apply in jurisdictions that rely on law enforcement to maintain order in schools.

Changes in the operation of school discipline parallel the evolution of the juvenile justice system decades ago. Until the 1960s, youth in juvenile court were denied the constitutional procedural protections afforded to adults in criminal court on the ground that the juvenile court, unlike the adult criminal court, was assumed to act in a nonadversarial manner in furtherance of the accused youth’s interests—to rehabilitate rather than punish the youth. By the 1960s,

144 In re Gault, 387 U.S. 1, 25-26 (1967) (“It is urged that the juvenile benefits from informal proceedings in the court. The early conception of the Juvenile
however, emerging social science evidence on actual juvenile court practices and their impact on the emotional and social development of youth cast doubt on those earlier premises:

While there can be no doubt of the original laudable purpose of juvenile courts, studies and critiques in recent years raise serious questions as to whether actual performance measures well enough against theoretical purpose to make tolerable the immunity of the process from the reach of constitutional guarantees available to adults.\textsuperscript{145}

The Supreme Court reasoned that “neither sentiment nor folklore should cause us to shut our eyes” to these studies.\textsuperscript{146} Evaluating this evidence, it concluded first in \textit{Kent v. United States} that juveniles had the “worst of both worlds”—neither the nurturing benefit of a nonadversarial system, nor the procedural protections of adult criminal court.\textsuperscript{147} One year later, in \textit{In re Gault}, it found that “[t]he rhetoric of the juvenile court movement has developed without any necessarily close correspondence to the realities of court and institutional routines.”\textsuperscript{148} In light of empirical developments, the Court reversed decades of precedent to extend to juveniles many of the constitutional procedural protections previously reserved for individuals in the adult criminal system.

Similarly, recent empirical evidence suggests that school discipline practices may no longer advance the beneficent, nonadversarial goals that once insulated them from ordinary judicial scrutiny. In light of the growing divergence between stated goals and actual practices in school discipline, courts should reconsider the validity of doctrinal restrictions on procedural rights in the school discipline context, just as the Supreme Court did for juvenile courts in \textit{Gault}. This part sets forth a framework for courts to do so. In addition, it discusses the potential role of nonjudicial actors in reform efforts.

---

\textsuperscript{146} In re Gault, 387 U.S. at 21.
\textsuperscript{147} 383 U.S. at 556.
\textsuperscript{148} 387 U.S. at 30 (citation omitted).
A. A Context-Specific Approach to Procedural Protections in Courts

In light of the empirical evidence showing an increased reliance on law enforcement to maintain public school order and the educational harms associated with this increased reliance, courts should critically evaluate the operation of school discipline to ensure that it actually advances the educational interests that previously justified insulating school discipline from closer judicial scrutiny. Given how significantly school discipline practices vary across jurisdictions, however, this analysis should be location-specific.

The framework proposed here preserves the presumption that school discipline generally serves pedagogical goals, but permits the youth to rebut this presumption by showing that discipline practices in the youth's particular school or district do not further educational interests. Relevant evidence might include, for example, data showing high rates of school-based arrests in the school or district, the frequent use of school-based arrests to handle relatively minor misconduct, or expert testimony regarding the educational impact of particular disciplinary practices employed in the school or district.

Reviewing this evidence, the court would render an interpretive judgment as to whether the particular discipline practices in a school or district primarily further a law enforcement goal rather than an educational one. Where the court finds the evidence persuasive, investigations of student misconduct would be presumed adversarial and thus subject to the full scope of constitutional protections that would be available to youth outside of the school context. By contrast, where school discipline practices adhere to the traditional model of discipline furthering pedagogical goals, doctrinal restrictions on constitutional rights would remain in place. Thus, the availability of procedural protections for youth in public schools would depend on a jurisdiction-specific assessment, rather than a categorical assumption, of the educational benefit of school

---

149 See Eve Brensike Primus, Disentangling Administrative Searches, 111 COLUM. L. REV. 254, 310-11 (2011) (critiquing current “one-size-permits-everything” categorical approach to administrative searches such as those conducted in public schools).

150 See also Anthony V. Alfieri, Post-Racialism in the Inner City: Structure and Culture in Lawyering, 98 GEO. L.J. 921, 959-60 (2010) (urging development of “local, school-specific fact investigation” to examine impact of school-based law enforcement referrals on youth).
discipline practices. This framework ensures that the education of youth is the paramount interest served.

There are a number of potential objections to this contextualized approach. First, one might argue that even if school discipline no longer serves the student’s educational interests, procedural rights should remain limited to protect the interests of other students. Second, a critic might contend that even though jurisdictions differ in their reliance on law enforcement, a categorical rule that assumes that school discipline always furthers the educational interests of youth is preferable because it is easier to administer. Third, one might argue that a better approach would be to employ an individualized analysis to determine whether the investigation of misconduct in a particular case furthers educational goals. Fourth, one might object to this type of contextualized rule on the ground that it would improperly result in constitutional protections varying by geography (e.g., probable cause required for searches in schools in district A, but not in district B). Fifth and finally, the framework is subject to the criticism that courts should not be in the business of second-guessing the educational value of school discipline. This section addresses each in turn.

**Rejecting reliance on the interests of other students.**

One might argue that even if the relationship between investigating official and accused student is recognized as adversarial and in service of law enforcement, procedural restrictions may remain warranted because the interests of the individual student are outweighed by the countervailing interests of other students in learning without disruption. The Supreme Court has noted this concern about preserving the rights of other students in some cases. This view does not necessarily hold that the ordinary calculus weighing individual interests against collective interests does not apply, but rather that the collective interest in school order weighs so heavily as to overcome the competing individual student’s interest in securing the full scope of procedural protections.

---

151 See Ryan, supra note 6, at 1341, 1411-14 (interpreting limits on constitutional rights in school discipline cases as resting on the view that they are necessary to “preserve an atmosphere that is safe and conducive to learning” and “to maintain discipline in order to transmit academic knowledge”); see also Tribe, supra note 7, at 314 n.128 (“[E]ven if in general the teachers’ and child’s interest truly converge, at the moment of suspension convergence must surely turn into clash: the teacher is saying that the best interests of other students will be served by this particular student’s suspension.”).

152 See supra notes 58-61 and accompanying text.
As a purely doctrinal matter, this approach is not entirely satisfactory. Relying on the interests of other students fails to articulate why the collective interest in school order should outweigh the collective interest in, say, reducing violent crimes in neighborhoods. It sets up the same tension between individual interests and collective interests that exists for all criminal procedural protections. Perhaps for this reason, the Court has never relied exclusively on this rationale, instead using it to buttress its primary rationale—that the collective interests and the individual student’s interests are aligned.

Moreover, as an empirical matter, there is little evidentiary support in the literature for the claim that the use of law enforcement, rather than other forms of discipline, in fact advances the educational interests of other students. On the contrary, as set forth in the preceding part, social scientists from related disciplines in education, criminology, and sociology suggest that heavy reliance on policing measures in public schools generates an adversarial atmosphere that may compromise the educational environment for all students.153

Rejecting a categorical rule. The proposed context-specific approach to determining the scope of procedural rights is preferable to an alternative categorical regime notwithstanding concerns of administrability. Concededly, the current categorical rule that exempts school searches from probable cause requirements and presumes school officials are not agents of the police for Miranda purposes is easy for courts to apply consistently; courts and litigants are not required to engage in a potentially costly fact-specific assessment of actual school discipline practices.154 A defense of the current categorical rule might reason that schools in general continue to use school discipline to further educational interests, even if some schools depart from this norm; thus, the factual premise on which the doctrinal restrictions rest—that school discipline and law enforcement generally remain discrete—remains accurate most of the time.

Such reliance on generalities is explicitly permissible pursuant to the Mathews v. Eldridge framework for determining procedural rights in the civil context, which assesses the risk of error associated with the denial of

153 See supra notes 138-41 and accompanying text.
procedural protections in the “generality of cases, not the rare exceptions.” Under this view, restrictions on procedural rights are warranted so long as school discipline “generally” is discrete from law enforcement, and disciplinarians “rarely” stand in an adversarial position vis a vis students.

The increasing availability of relevant data mitigates this concern about administrability. Federal agencies have amended the biannual federal Civil Rights Data Collection to require school districts to maintain and publicly report data on the total number of student referrals to law enforcement and the total number of school-related arrests for each school. These mandatory data-collection and reporting requirements reduce the litigation costs associated with obtaining and analyzing this information.

More importantly, the primary purpose of uniform, categorical rules is to ensure that “similarly situated litigants are treated equally.” Where the facts show that litigants are not, in fact, similarly situated in a legally significant way, the application of the same rule to them is no longer appropriate. In addition, reliance on generalities in the name of efficiency is inappropriate where, as here, criminal—as opposed to civil—procedures are implicated. As Jerry L. Mashaw has pointed

155 Mathews v. Eldridge, 424 U.S. 319, 335, 344 (1976) (stating that determination of what process is due in the administrative context requires balancing of three factors: “First, the private interest that will be affected by the official action; second, the risk of an erroneous deprivation of such interest through the procedures used, and the probable value, if any, of additional or substitute procedural safeguards; and finally, the Government’s interest, including the function involved and the fiscal and administrative burdens that the additional or substitute procedural requirement would entail.”); see, e.g., Carey v. Piphus, 435 U.S. 247, 259 (1978) (applying Mathews framework to determine procedural due process rights to challenge school suspension); Ingraham v. Wright, 430 U.S. 651, 675 (1977) (same for corporal punishment); Doe ex rel. Doe v. Todd Cnty. Sch. Dist., 625 F.3d 459, 462-64 (8th Cir. 2010) (same for placement in an alternative high school setting); Watson ex rel. Watson v. Beckel, 242 F.3d 1237, 1240 (10th Cir. 2001) (same for expulsion); Palmer v. Merluzzi, 868 F.2d 90, 95 (3d Cir. 1989) (same for suspension from interscholastic athletics); Newsome v. Batavia Local Sch. Dist., 842 F.2d 920, 923-24 (6th Cir. 1988) (same for expulsion); In re Expulsion of E.J.W. from Indep. Sch. Dist. No. 500, 632 N.W.2d 775, 780 (Minn. Ct. App. 2001) (same); Hinds Cnty. Sch. Dist. Bd. of Trs. v. R.B. ex rel. D.L.B., 10 So. 3d 387, 399-402 (Miss. 2008) (same).

156 OFFICE FOR CIVIL RIGHTS, U.S. DEP’T OF EDUC., 2009-10 CIVIL RIGHTS DATA COLLECTION, OMB 1875-0240 (on file with author).


158 See generally Alexes Harris, Diverting and Abdicating Judicial Discretion: Cultural, Political, and Procedural Dynamics in California Juvenile Justice, 41 LAW & SOC’Y REV. 387 (2007) (discussing competing values of predictability and uniformity while describing tension between individual justice versus equal justice).

out, criminal procedural rules prohibit coerced confessions not because they will yield inaccurate conclusions about guilt—indeed, they are often all too accurate—but rather because such confessions offend our basic sense of personal autonomy.\textsuperscript{160} Our juvenile and criminal justice systems require a more granular assessment of facts than the current categorical rule provides.

**Rejecting an individualized rule.** In light of the need for a more granular rule, one might argue in favor of an individualized case-by-case assessment of whether procedural rules should be extended to a particular student, instead of the jurisdiction-specific rule proposed here. This alternative approach might determine procedural rules depending on, for example, whether a school resource officer was present or participated in the particular search or questioning, or the subjective intent of the school official in conducting the search or questioning.

This individualized alternative, however, would not be workable. The presence or absence of a school resource officer does not indicate whether a search or questioning was carried out for law enforcement purposes. One recent study found that the presence of school resource officers was not correlated with the number of school-based arrests in a school.\textsuperscript{161} The particular roles played by school resource officers differ significantly across schools. In some schools they are charged with enforcement of criminal laws, while in others they focus on mentoring, counseling, and teaching.\textsuperscript{162} Because of this variability in roles, school resource officers cannot neatly be categorized as law enforcement or school official. The better approach to determine whether school investigations serve law enforcement purposes is one that examines aggregate data of rates of law enforcement referrals in the particular school or district.

Nor would it be desirable for a court to attempt to discern the subjective intent of individual school officials to determine

\begin{itemize}

\item[\textsuperscript{161}] Theriot, *supra* note 9, at 284-85 (finding that the presence of school resource officers does not predict more total arrests, but does predict more arrests for disorderly conduct).

\item[\textsuperscript{162}] Peter Finn & Jack McDevitt, *National Assessment of School Resource Officer Programs* 43 (2005).
\end{itemize}
whether a particular search or interrogation was for law enforcement purposes; this approach would present significant evidentiary difficulties and potentially encourage false statements from school officials. More importantly, such an individualized rule would chill school principals, who might feel pressured to provide Miranda warnings any time they questioned a student about anything—even about not having a hall pass—in case the student responded with incriminating statements that might subsequently be excluded from court because of a Miranda violation. The context-specific rule, by contrast, would encourage school officials to provide procedural protections only in schools or districts that routinely rely on law enforcement, and not in those schools or districts that do not.

Accepting geographically contingent rules. A fourth objection would challenge the proposed context-specific rule because it would result in constitutional protections varying by geography: probable cause and Miranda warnings would be required for student investigations in district A, but not in district B. There is precedent for these location-specific constitutional protections, however.

In Illinois v. Wardlow, the Supreme Court concluded that the question of whether flight from the police provides reasonable suspicion to justify a stop-and-frisk depends on background facts regarding the particular area—specifically, the level of crime in that particular area. Here, courts would employ a similar approach to reject the categorical rule denying procedural protections in schools and instead determine the entitlement to such protections based on the background facts of student criminalization in the particular school or district. In those schools or districts in which law enforcement goals were shown to predominate over educational ones, courts would demand probable cause and Miranda warnings for school-based investigations.

Accepting educational assessments by courts: Finally, one might object to the proposed context-specific approach on the ground that courts should not be in the business of second-

---

164 See also Livingston, supra note 7, at 286 (arguing that courts should consider “character and social meaning” of encounters between officers and individuals to determine whether the encounter furthers ordinary law enforcement goals and should thus be subject to ordinary criminal procedural protections, or whether, instead, the encounter lacks the adversarial relationship characteristic of law enforcement encounters and should thus benefit from more relaxed procedural protections).
guessing the decisions of school officials. As the Supreme Court emphasized in the student speech case *Hazelwood School District v. Kuhlmeier*, “the education of the Nation’s youth is primarily the responsibility of parents, teachers, and state and local school officials, and not of federal judges.”

Notwithstanding the deference afforded to school officials, however, the *Kuhlmeier* decision affirmed the propriety of judicial scrutiny over the educational value of school officials’ decisions. In *Kuhlmeier*, which involved censorship of a high school newspaper published by students in a journalism class, the Court held that even in the context of classroom activities, where the deference afforded to school officials is greatest, it would review a school official’s conduct to determine whether it is “reasonably related to legitimate pedagogical concerns.” If courts are trusted with reviewing the educational goals of in-classroom decisions in the First Amendment context—albeit pursuant to a forgiving standard of review—it is difficult to understand why they should not be trusted with determining whether particular discipline practices further educational goals in the Fourth and Fifth Amendment contexts, which involve the procedural rights of criminal suspects, an area in which courts have particular expertise.

**B. The Role of Nonjudicial Actors in Preserving the Educational Value of School Discipline**

Ultimately, courts have limited authority in shaping school discipline. They may determine which procedural protections will extend to youth who are investigated or punished at school, but they lack the authority to determine what kind of conduct warrants punishment, and what kind of punishment should be imposed. They are poorly situated institutionally to prevent criminal charges from being filed against a student for engaging in a schoolyard shoving match or cursing loudly in class. Therefore, a critical role exists for

---

165 Buss, *supra* note 3, at 570 (discussing institutional disadvantages of judicial determinations of school policies).
167 *Id.*
168 Buss, *supra* note 3, at 571 (noting that, although “the public school context may require special latitude for educational or administrative judgment,” procedures involving alleged student misconduct are within “the field in which courts are most competent”).
nonjudicial actors to ensure that these decisions are informed by the educational impact of discipline practices.\footnote{Michael Heise, \textit{Litigated Learning and the Limits of Law}, 57 VAND. L. REV. 2417 (2004) (discussing limits of litigation to create social change in education).} Even with robust judicial procedural protections, nonjudicial actors including school boards, principals, teachers, and individual police officers will always retain a great deal of discretion in determining how to handle student misconduct. In light of the emerging empirical evidence, those charged with developing discipline practices should reduce reliance on law enforcement and instead institute practices that improve educational outcomes for youth.

Policy makers in several jurisdictions have already taken the lead in examining the intersection between school discipline and law enforcement. For example, school officials in Clayton County, Georgia, a part of the Atlanta Metro region, convened a Blue Ribbon Commission to study school discipline issues.\footnote{\textit{CLAYTON CNTY. PUB. SCH., BLUE RIBBON COMMISSION ON SCHOOL DISCIPLINE: A WRITTEN REPORT PRESENTED TO THE SUPERINTENDENT AND BOARD OF EDUCATION} (1997).} In its report, the Commission found that in the span of a few years, the number of student referrals to law enforcement per year grew from eighty-nine to 1,400.\footnote{\textit{Id.}} The Commission further found that most of the offenses involved minor incidents such as fights or disorderly conduct that “have traditionally been handled by the school and are not deemed the type of matters appropriate for juvenile court.”\footnote{\textit{Id.}} Based on these empirical findings, the chief judge of the local juvenile court convened a group of local stakeholders including parents, police officers, school officials, and juvenile public defenders to discuss the use of the juvenile justice system to maintain student discipline. After a series of roundtable meetings, participants reached a resolution that would further school safety and at the same time reduce the number of youth referred to juvenile court for school-based misconduct. The resulting cooperative agreement imposed a three-strikes policy for disciplinary infractions.\footnote{Clayton County Cooperative Agreement (2007), available at http://www.juveniledefender.org/files/resources-juvenile-cooperative_agreement_070804.pdf.} The first time a child commits certain offenses identified as “focus acts”—affray, disruption of school, disorderly conduct, minor obstruction of the police, and protections mean little when there are no substantive limits to the government’s ability to define impermissible behavior, as in the situation of a school principal)
criminal trespass—the student receives a warning rather than being referred to law enforcement, as was the prior practice. If the youth commits one of these offenses a second time, the child is referred to a school conflict-diversion program, mediation program, or other court-sponsored program. It is only if the student commits the offense a third time that he or she may be referred to law enforcement.

Importantly, the Cooperative Agreement has succeeded not only in reducing the number of school-based arrests, but also in improving school order and, according to its advocates, educational outcomes. Since the agreement was implemented, the number of dangerous weapons incidents decreased by 70 percent, fighting offenses decreased by 87 percent, and other focus acts decreased by 36 percent. Advocates for the reform effort maintain that the reduced reliance on school-based arrests furthers safety goals by facilitating nonadversarial relationships between students and authority figures. At the same time, graduation rates increased by 20 percent, although it is not clear that this can be attributed to the reduced reliance on law enforcement. Similar community reform efforts are underway in Denver, Baltimore, Raleigh, San Francisco, Atlanta, and Birmingham. These efforts suggest the possibility of effective reform by nonjudicial actors to ensure that mechanisms for maintaining school order actually benefit the educational interests of students.

CONCLUSION

Emerging empirical evidence casts significant doubt on the ongoing validity of doctrinal justifications for denying procedural protections to youth accused of misconduct in schools. Where a growing number of jurisdictions are relying on

175 Id.
176 Id.
177 Id.
179 STOP THE SCH. HOUSE TO JAIL HOUSE TRACK, supra note178.
180 Id.
law enforcement to maintain school order, the investigation and punishment of youth can no longer categorically be insulated from judicial scrutiny on the ground that it furthers the educational interests of the suspect-youth. Just as the Supreme Court in Gault confronted emerging factual evidence regarding the operation of the juvenile justice system to extend fuller procedural protections to youth in juvenile court, courts should consider how school discipline actually operates in today’s society and revisit the scope of procedural protections available to youth in public schools accordingly.

This article provides a descriptive assessment of the increased criminalization of school discipline and its impact on youth and sets forth a means by which courts and policymakers should respond to these factual developments. The underlying causes of such criminalization and the reasons for the disparities among jurisdictions are beyond the scope of this article. Such factors may include the availability of federal funding for school resource officers, reduction of resources for classroom management training, overcrowded classrooms, income levels, racial demographics, and dismantling of desegregation decrees, among many others. Future research in this area will be important to ensure that youth benefit from the model of school discipline as educational tool idealized by the Court.
INTRODUCTION

The United States food-safety regulatory program is a behemoth. It is overseen by at least five federal agencies administering at least six statutes. Yet almost 17 percent of the American population (approximately forty-eight million people) gets sick from food each year, 128,000 of these people are hospitalized, and three thousand die. Foodborne illness costs the United States over $150 billion per year. Food recalls are massive and frequent—like that of cantaloupe in 2011, which, with approximately thirty deaths from listeria, was the deadliest outbreak in almost a century; eggs in 2010, where...
more than a thousand people were sickened with salmonellosis; peanut butter in 2009, with over four hundred hospitalizations, and at least six deaths; and spinach in 2006, where 131 people were sickened and over sixty hospitalized. The nation’s food regulatory system is inefficient, underenforced, and underfunded.

Although multiple agencies are responsible for food safety, and have overlapping duties, there are wide gaps in oversight and regulation. The agencies must negotiate the competing goals of protecting the public health, marketing the nation’s commodities, and appeasing the interests of regulated entities. Compounding this problem is a severe shortage of resources allocated to food-safety enforcement. Given the regulatory failures, and their

---


7 Broadly conceived, food safety covers three areas: (1) the prevention of foodborne illness; (2) nutrition monitoring; and (3) the prevention of fraud in the marketplace. This article only addresses the first category, although there is overlap amongst the three. For example, I do not discuss cases regarding deceptive claims or fraudulent labeling, such as those brought in recent years by the Center for Science in the Public Interest (CSPI), although they certainly concern food safety. See, e.g., Amended Complaint at 1-2, 29, Parham v. McDonald's Corp., No. CGC-10-506178 (Cal. Super. Ct. Jan. 5, 2011), available at http://cspinet.org/new/pdf/mcdonald_scandalplaint.pdf (CSPI's McDonald's litigation charges McDonald's with unfair and deceptive marketing for including toys in Happy Meals); Complaint at 2-3, Ackerman v. Coca-Cola Co., No. 09-CV-0395 (E.D.N.Y. May 20, 2009), available at http://cspinet.org/new/pdf/vitaminwater Filed_complaint.pdf (CSPI's VitaminWater litigation charges Coca-Cola with fraudulent marketing for including claims of healthfulness on its VitaminWater products).


9 The Food Safety Modernization Act (FSMA), signed by the President on January 4, 2011, addresses some of the problems with the regulatory system. It provides more power, and, theoretically, more funding to the Food and Drug Administration for its food-safety duties, and changes the food-safety focus of the agency from responsive to preventative. See Food Safety Modernization Act, Pub. L. No. 111-353, 124 Stat. 3885 (2011). It does not, however, address the inefficiencies or inconsistencies caused by the fact that multiple agencies have authority over food regulation, nor does it alter the fact that the agencies have political imperatives, see, e.g., James T. O'Reilly, Losing Deference in the FDA's Second Century: Judicial Review, Politics, and a Diminished Legacy of Expertise, 93 CORNELL L. REV. 939, 962-72 (2008), and, in some instances, must balance the competing goals of economic viability of food production with the public health. See, e.g., Mission Statement, U.S. DEPT AGRIC., http://www.usda.gov/wps/portal/usda/usdahome/navid=MISSION_STATEMENT (last modified Oct. 29, 2009).
significant public health consequences, it is imperative to discuss alternatives to regulation as a means to improve the food-safety system in this country.

Advocacy group litigation can complement agency regulation in the field of food safety. And, just as such litigation enhances the enforcement of environmental laws, so too can advocacy group litigation add an element of “attentive monitoring” to food-safety statutes. This paper provides, for the first time, a comprehensive examination of advocacy group litigation, or food-safety-impact litigation, in the food-safety context, assessing its viability and utility.

In evaluating the potential efficacy of food-safety-impact litigation, this paper puts particular focus on the barriers to justiciability that such litigation faces, including standing challenges and justiciability issues under the Administrative Procedure Act. These threshold issues are particularly important here because plaintiffs in advocacy group litigation are not the direct objects of the regulation they target and because the injury to which they point is often an increased risk of future harm. Although the environmental laws contain structural legal elements, such as citizen-suit provisions and consultative arrangements, which food-safety laws do not, the absence of such

---

10 William H. Rodgers, Jr., The Seven Statutory Wonders of U.S. Environmental Law: Origins and Morphology, 27 LOY. L.A. L. REV. 1009, 1019-21 (1994). Rodgers explains that “attentive monitoring,” which contributes to the success of certain seminal environmental laws, comprises “personal activities such as face-to-face observation, emotions such as shame and pride, and group sanctions such as ostracism and citizen lawsuits.” Id. at 1020. He notes that these mechanisms are encouraged by structural legal changes. Id.

11 This article defines food-safety-impact litigation as suits brought against food-safety regulatory agencies to improve the regulatory scheme for the purpose of protecting the public health in general and the plaintiff's own health in particular. Impact litigation may be brought, for example, to contest an allegedly arbitrary and capricious denial of an administrative petition, or to argue that a final agency action was unreasonable.

12 I confine my discussion here to cases brought by individuals or entities against regulatory agencies for the purpose of forcing the agency to comply with its statutory mandate by either passing regulation, or interpreting existing regulations differently.

13 Food-safety litigants suing private parties under state law or suing state government to enforce or strengthen state regulation must also negotiate federal preemption doctrine. Preemption issues, however, do not arise in litigation against the federal government, for obvious reasons, and I therefore leave a discussion of preemption issues to another paper.


15 Impact litigation in the field of food safety is brought before an outbreak of foodborne illness, and seeks stronger regulation to prevent or minimize the chance of such an outbreak. For that reason, such litigation must be based on probabilities—the plaintiffs argue that it is x percent more likely that an outbreak of foodborne illness will take place with the current state of regulation than with the requested regulation.
provisions is not an insurmountable barrier to impact litigation.\textsuperscript{16} Thus despite these justiciability challenges, food-safety-impact suits are possible to bring, and possible to win.

Why, then, is there so little food-safety-impact litigation? Essentially, the answer is that there is no culture of citizen food-safety litigation, as there is for environmental-impact litigation. The environmental bar has spent decades learning to litigate around the justiciability barriers discussed in this paper, and courts have adapted many of these doctrines for the environmental context.\textsuperscript{17} But there is no food protection community practiced in the art of litigation against agency action and inaction.

For food-safety-impact litigation to be successful it must be modeled after environmental litigation. The doctrines developed by environmental-impact litigation provide an avenue for successful pre-illness food-safety litigation. Both food-safety and environmental-impact litigation depend on probabilities and involve great uncertainty. Governmental regulation of this uncertainty involves a negotiation of risk, cost-benefit analysis, public perception, and political reality. Both the environmental laws and the food-safety laws envision a joint state-federal system of enforcement. Moreover, the injury stemming from environmental or food-safety harms is likely to be widely shared, yet particularized. For this reason, the few courts that have actually dealt with food-safety-impact litigation look to environmental litigation as a backdrop,\textsuperscript{18} and mechanisms developed by litigants to maneuver environmental

\textsuperscript{16} I have spoken with several lawyers at public interest organizations who confirmed my inclination that food-safety cases can be successful. For example, Allison Zieve, the director of Public Citizen’s Litigation Group, told me that these cases were no longer generated by anyone at her organization, and that she would be willing to consider such suits if litigable issues were brought to her attention. Telephone Interview with Allison Zieve, Dir., Pub. Citizen Litig. Grp. (Jan. 28, 2011) (notes on file with the author). I spoke with a former Associate Chief Counsel of the FDA, who asked not to be named, who told me that during his tenure at the FDA, which was during George W. Bush’s administration, several of his colleagues and he wished there had been citizen litigation against the FDA. During these years, he explained, the agency was market oriented instead of food-safety oriented, and he felt that citizen litigation could have forced the agency’s hand as to certain regulatory issues in the food-safety field. Indeed the food-safety-impact litigation that does exist has a more than respectable success rate—out of thirteen cases, plaintiffs achieved at least some of their requested relief in three of the cases, and a fourth case was mooted when the agency adopted plaintiff’s position during pendency of the suit.


litigation past justiciability barriers are both necessary and useful for food-safety litigation.

This article begins a discussion of the parameters and implications of food-safety citizen and advocacy group litigation. Moreover, this article provides an added perspective on the need for courts to adapt traditional barriers to justiciability to the realities of litigation that involves injuries based on the possibility of future harm and the increase in risk associated thereof. Food safety involves issues on the frontier of regulation, including the need to address the role of lifestyle choice in conjunction with public health, and the need to regulate a massive system involving minutely local as well as global elements. Moreover, issues about food are increasingly in the forefront of public awareness. The understanding that advocacy-group litigation is possible and can be successful may be a valuable tool for this burgeoning movement.

Any expansion of the regulatory state carries challenges to its scope and authority as a counterpart to its growth. These challenges come both from proponents and opponents of regulation. As noted above, public challenges to the regulatory system will confront obstacles to litigation, as courts attempt to negotiate their role vis-à-vis this public litigation. Justiciability issues will be at the forefront of much public litigation in the coming years.

This article proceeds as follows: Part I summarizes the structure of the United States food regulatory system, and looks at some of the problems with agency enforcement and nonlegal sanctions. Part II discusses the absence of a culture of food-safety-impact litigation, addressing the nature of food and the divergent paths that environmental protection and food-safety regulation have taken in the United States. Part III turns to the justiciability barriers that food-safety-impact plaintiffs will face, and have faced in the few already litigated cases, focusing on constitutional and prudential standing issues, and reviewability challenges under the Administrative Procedure Act. This section concludes with an analysis of how food-safety citizen plaintiffs can successfully navigate the justiciability challenges they are sure to face. Finally, the

---

19 See, e.g., Pub. Citizen Health Research Grp. v. U.S. Dep’t of Labor, 557 F.3d 165, 169 (3d Cir. 2009) (public interest group and union challenge Occupational Safety and Health Administration (OSHA) regulation for its underregulation of hexavalent chromium; industry group challenges the same regulation for its overregulation of same chemical).
article concludes that, under certain circumstances, citizen litigation can be a valuable counterpart to government regulation and spur regulatory agencies to fulfill their statutory mandates. Food-safety regulation, like environmental regulation, is a massive endeavor involving many agencies, statutes, regulated entities, and beneficiaries, and requiring vast resources. As citizen advocacy groups have acted as a counterpart to government regulation in the environmental arena, so could they be useful in the field of food safety.

I. AN OVERVIEW OF THE UNITED STATES FOOD REGULATORY SYSTEM

A. The Regulatory Structure of Food Safety

Food safety is mainly overseen by the Food and Drug Administration (FDA) and the U.S. Department of Agriculture (USDA), although several other agencies play a part as well.\textsuperscript{20} Government regulation of food has historically had three main purposes: (1) to protect the integrity of the market; (2) to regulate the nutritional content of food; and (3) to protect the safety of the food supply.\textsuperscript{21} Although there is overlap among the three goals, the topic of this paper—impact litigation to minimize the threat of foodborne illness—is mostly contained within the third.\textsuperscript{22}

The FDA has authority under the Federal Food, Drug, and Cosmetic Act (FFDCA) over most food products except meat, poultry, and processed egg products.\textsuperscript{23} The FFDCA, passed in

\textsuperscript{20} The Environmental Protection Agency regulates drinking water and pesticide residues; the Federal Trade Commission shares jurisdiction over advertising; the Alcohol and Tobacco Tax and Trade Bureau regulates alcohol; the Centers for Disease Control and Prevention tracks foodborne illness; the National Marine Fisheries Service helps regulate fish and seafood products; the Customs Service regulates imported foods; and the Department of Justice prosecutes individuals and companies for violations of food-safety statutes. See NEAL D. FORTIN, FOOD REGULATION 24-27 (John Wiley & Sons, Inc. 2009).

\textsuperscript{21} Peter Barton Hutt, Government Regulation of the Integrity of the Food Supply, 4 ANN. REV. NUTRITION 1, 2 (1984).

\textsuperscript{22} Protecting the food supply includes more than the prevention of foodborne illness. It can also involve measures to prevent external security threats, for example, intentional poisoning of the food supply, which I do not discuss in this article. See, e.g., Armen Keteyian, Latest Terror Threat in U.S. Aimed to Poison Food, CBS NEWS (Dec. 20, 2010, 11:48 PM), http://www.cbsnews.com/stories/2010/12/20/eveningnews/main7169266.shtml. Additionally, steps taken to protect the food supply may also protect the integrity of the market, and vice versa.

\textsuperscript{23} 21 U.S.C. §§ 301-2252 (2006). The FDA does have authority, however, over imported wild game. See Fact Sheets: Meat Preparation, U.S. DEP'T AGRIC.,
1938, updated the 1906 Pure Food and Drug Act. It has been substantially amended since 1938, but still retains its basic structure. The 1938 FFDCA has been described as “a catalogue of definitions elaborating two basic concepts: ‘adulteration’ and ‘misbranding.’” The Act specifies when a food (or drug device or cosmetic) is adulterated or misbranded, and prohibits the distribution or sale of any such food.

To enforce the FFDCA, the FDA has authority to institute various administrative, civil, or criminal actions. It can issue warning letters, request a voluntary recall of an adulterated product, order recalls under certain circumstances, seize products that violate the Act, assess civil penalties, and work with the Department of Justice to take court action.

The FDA Food Safety Modernization Act, signed into law in January 2011, is the biggest reform of the FDA’s food regulatory powers since 1938, and its implementation will reorient the FDA to take a preventative rather than a responsive role regarding food safety. The Act amends the FFDCA to, among other things, give the FDA: (a) mandatory recall authority; (b) expanded authority to inspect records; and (c) the authority to suspend the registration of a food facility. The Act also requires owners and operators of food

24 Hutt, supra note 21, at 7.
25 PETER BARTON HUTT, RICHARD A. MERRILL & LEWIS A. GROSSMAN, FOOD AND DRUG LAW 14 (Foundation Press, 3d ed. 2007).
26 Id. at 13.
30 Id. § 337.
31 Certain provisions in the Act, such as the mandatory recall provisions, take effect immediately, while the FDA must write rules to implement other provisions of the Act. See Helena Bottemiller, Q & A with Michael Taylor, Part I: Implementing FSMA, FOOD SAFETY NEWS (Jan. 23, 2012), http://www.foodsafetynews.com/2012/01/q-a-with-michael-taylor-part-1-implementing-fsma/. Implementation of the Act’s provisions and the FDA’s expanded powers also rests on whether adequate funding is provided by Congress, an issue that is in question. See Lyndsey Layton, Overhaul of Food Safety Laws May Not Be to GOP’s Taste, WASH. POST (Dec. 25, 2010), http://www.washingtonpost.com/wp-dyn/content/article/2010/12/24/AR2010122402795.html.
33 Id. § 101 (amending 21 U.S.C. § 350e(a)).
34 Id. § 102(b) (amending 21 U.S.C. § 350d(a)).
facilities to evaluate the hazards that could affect food, and implement and monitor preventative controls. Imported food will have to meet the same standards. The FDA is also directed to increase inspections of domestic and foreign facilities, directing resources to the riskiest facilities.

The USDA shares authority over food safety with the FDA. Its Food Safety Inspection Service (FSIS) has authority over meat, poultry, processed egg products, and egg grading, and it regulates products related to meat and poultry, including stews, pizzas, and frozen foods. The major food-safety statutes administered by the USDA are the Federal Meat Inspection Act (FMIA), passed in 1906 and amended in 1967 by the Wholesome Meat Act; the Poultry Products Inspection Act (PPIA); and the Egg Products Inspection Act (EPIA). The FSIS, like the FDA, can issue warning letters and seize products under the FMIA and the PPIA.

Unlike food producers under the authority of the FDA, however, meat-and-poultry producing establishments must have a USDA inspector present whenever they are operating. Much of the FSIS’s power over the food supply for which it is responsible stems from its ability to take a regulatory control action in relation to its inspection authority. Such actions include “the retention of product, rejection of equipment or facilities, slowing or stopping of lines, or refusal to allow the processing of specifically-identified product.” FSIS can also withhold a mark of federal inspection (without which a product

---

35 Facilities will have to implement Hazard Analysis and Critical Control Point (HACCP) protocols, which the USDA has required of meat and poultry facilities since 2003. See, e.g., Control of Listeria Monocytogenes in Ready-to-Eat Meat and Poultry Products, 9 C.F.R. § 430.4 (2003).
37 Id. §§ 301-309 (amending scattered sections of 21 U.S.C.).
38 Id. §§ 201-211 (amending scattered sections of 21 U.S.C.).
40 FORTIN, supra note 20, at 6.
43 Id. §§ 1031-1056.
44 Id. §§ 673, 467b.
45 This may change, in that the FSMA directs the Secretary of the Department of Health and Human Services (HHS) to establish regulations concerning inspections for facilities under its authority. S. 510, 111th Cong. § 201 (2010).
46 9 C.F.R. § 302.3 (2010); FORTIN, supra note 20, at 567 (“FSIS inspects meat and poultry under a ‘continuous inspection’ system, which means that an inspector is assigned to every FSIS-regulated establishment and is required to be present when the establishment is in operation.”).
47 FORTIN, supra note 20, at 568.
cannot be distributed or sold, or suspend inspection at a certain facility.\textsuperscript{48} If inspections are suspended at a facility, the facility must stop operating until the USDA reinstitutes inspections.\textsuperscript{49} If a grant of federal inspection is withdrawn, a facility must cease operations completely.\textsuperscript{50}

B. Enforcement Shortcomings

Notwithstanding the imposing regulatory structure built to oversee food safety, and the recent enactment of the FSMA, our food regulatory system is falling short for several reasons. First, the bifurcation of major food regulatory duties between the FDA and the USDA results in inefficiencies, inconsistencies, and even some absurdities.\textsuperscript{51} For an oft-repeated example, pizza is regulated by the FDA unless it has a topping of more than 2 percent of cooked meat or poultry, in which case the USDA is in charge.\textsuperscript{52} For this reason, pizza production facilities are often regulated by both agencies.\textsuperscript{53}

Redundant oversight is clearly inefficient. But more dangerous to the consuming public is the possibility of inconsistency in the agencies’ regulatory regimes. For example, eggs are subject to a baffling array of regulations and regulatory oversight, the result of which leaves gaps in food-safety enforcement.\textsuperscript{54} Eggs were responsible for approximately 75

\textsuperscript{48} 9 C.F.R. § 500.1-2.
\textsuperscript{49} FORTIN, supra note 20, at 513.
\textsuperscript{50} Id. at 514.
\textsuperscript{51} Consolidation of agency responsibilities, an often touted solution, is not only politically inviable at this time, but also may not actually fix the problems. See, e.g., Note, Reforming the Food Safety System: What if Consolidation Isn’t Enough?, 120 HARV. L. REV. 1345, 1347 (2007) [hereinafter Reforming the Food Safety System].
\textsuperscript{53} Reforming the Food Safety System, supra note 51, at 1350 (citing INST. OF MED. NAT’L RESEARCH COUNCIL, ENSURING SAFE FOOD: FROM PRODUCTION TO CONSUMPTION 85 (1998)).
\textsuperscript{54} See U.S. GEN. ACCOUNTING OFFICE, REPORT TO CONGRESSIONAL COMMITTEES, FEDERAL FOOD SAFETY OVERSIGHT (2011), available at http://www.gao.gov/new.items/d11289.pdf (“FDA is generally responsible for ensuring that eggs in their shells—referred to as shell eggs—including eggs at farms such as those where the outbreak occurred, are safe, wholesome, and properly labeled. FSIS, on the other hand, is responsible for the safety of eggs processed into egg products. In addition, USDA’s Agricultural Marketing Service (AMS) sets quality and grade standards for shell eggs, such as Grade A, but does not test the eggs for bacteria such as Salmonella. Further, while USDA’s Animal and Plant Health Inspection Service manages the program that helps ensure laying hens are free from Salmonella at birth, FDA oversees the safety of the feed they eat.”).
percent of all salmonella outbreaks between 1985 and 1998—although this percentage may be dropping due to a new egg-safety rule promulgated by the FDA in 2009—and the country was reminded of their potential dangerousness during the massive 2010 egg recall. There is also the possibility of inconsistency within one agency’s regulation of different products. For example, in one of the suits discussed below, poultry consumers sued the USDA over what they contended was irrationally inconsistent treatment of meat and poultry.

Funding for food safety is also a problem. The Center for Science in the Public Interest has noted that “since 1972, inspections conducted by the FDA declined 81 percent. Since 2003, the number of FDA field staff dropped by 12 percent, and between 2003 and 2006 federal inspections dropped by 47 percent.” Although meat and poultry account for less foodborne illness than do seafood and fresh produce, which are regulated by the FDA, the USDA spends significantly more money on food safety than does the FDA. Moreover, the FDA will not be able to implement the inspections mandated by the Food Safety Modernization Act if Congress withholds funding, as it has threatened to do.

Finally, there is a perception that both the FDA and the USDA are subject to agency capture, whereby regulated entities exert such an influence over their regulators that they essentially control the agencies, at the expense of the intended beneficiaries of the regulatory system. The prevalence of such

---

57 See infra notes 245-47 and accompanying text.
61 See, e.g., Nicholas Bagley, Agency Hygiene, 89 TEX. L. REV. SEE ALSO 1, 2, 8 n.32 (2010); William W. Buzbee, Asymmetrical Regulation: Risk, Preemption, and the Floor/Ceiling Distinction, 82 N.Y.U. L. REV. 1547, 1590-91 (2007) (“Agencies may be able to secure expanded budgets or even engage in outright favoritism to affected industry in exchange for the usual rewards of regulatory capture—electoral support for the
an industry viewpoint in the FDA and the USDA may produce lax enforcement and a reliance on industry self-regulation, even when it may not be the best approach for the public.\footnote{For example, the mission statement of the USDA contains a provision that makes such industry pressure a mandate, rather than an eventuality—the agency is charged both with expanding and protecting agricultural markets, as well as enhancing food safety. Mission Statement, U.S. DEP'T AGRIC., supra note 9. While in theory prevention of foodborne illness is entirely consistent with improving the economics of agriculture, the USDA's main priority is split between economic development and public health. For another example, the New York Times recently reported that the USDA was assisting in the marketing of cheese, while simultaneously discouraging its consumption in an anti-obesity campaign. Michael Moss, While Warning About Fat, U.S. Pushes Cheese Sales, N.Y. TIMES, Nov. 6, 2010, at A1, available at http://www.nytimes.com/2010/11/07/us/07fat.html?scp=1&sq=domino+cheese&st=nyt.} This phenomenon can be explicit, where there is an actual flow of individuals between industry and decision-making regulatory positions, or implicit, which involves more attenuated but no less real connections between decision makers and industry.\footnote{See Nicholas Bagley & Richard L. Revesz, Centralized Oversight of the Regulatory State, 106 COLUM. L. REV. 1260, 1285 (2006). The question of whether agency capture is actually a problem, or more a perceived problem, and the effects thereof, has been amply discussed in legal scholarship. See, e.g., Thomas W. Merrill, Capture Theory and the Courts, 72 CHI.-KENT L. REV. 1039 (1997); Cass R. Sunstein, What's Standing After Lujan? Of Citizen Suits, "Injuries," and Article III, 91 Mich. L. Rev. 163, 183-84 (1992-93); see also Bagley & Revesz, supra, at 1284-92 (arguing that the theory of regulatory capture does not adequately explain the reality of governmental agency processes).} Under some circumstances, citizen-impact litigation can address some of these regulatory shortcomings and be a useful counterpart to government regulation.
II. The Absence of an Impact Litigation Culture in the Context of Food Safety

Food-safety-impact litigation, however, is rare. Over the last forty years, fewer than twenty published cases fall into this category.\textsuperscript{64}

\textsuperscript{64} I searched for cases by citizens or organizational plaintiffs against the government seeking to change regulation for the stated purpose of improving public health, as well as state and federal cases brought against the USDA, HHS, or the FDA since 1970 that concerned food safety, and found the following fourteen cases: Levine v. Vilsack, 587 F.3d 986 (9th Cir. 2009) (citizens brought suit against Secretary of Agriculture challenging rule excluding poultry from Humane Methods of Slaughter Act; Ninth Circuit dismissed for lack of Article III standing; discussed, infra Part III.B.2); Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. USDA, 499 F.3d 1108 (9th Cir. 2007) (trade organization challenged USDA regulation of the importation of Canadian cattle; court deferred to agency; discussed, infra Part III.B.3); Baur v. Veneman, 352 F.3d 625 (2d Cir. 2003) (citizen sued USDA to challenge refusal to prohibit downed cattle from entering food supply; Second Circuit found standing; discussed, infra Part III.B.1); Am. Fed’n of Gov’t Emps., AFL-CIO v. Veneman, 284 F.3d 125 (D.C. Cir. 2002) (organizations representing government employees challenged a new provisional inspection system instituted by the USDA of meat and poultry carcasses; court ultimately found that this new, but provisional and temporary system was adequate; standing not discussed); Kenney v. Glickman, 96 F.3d 1118 (8th Cir. 1996) (poultry consumers and red meat producers challenged alleged inconsistencies between regulation of meat and poultry; court found case reviewable under APA; discussed in detail, infra Part III.C); Arent v. Shalala, 70 F.3d 610 (D.C. Cir. 1995) (consumer and public interest groups challenged labeling of raw fish and produce; court found FDA’s industry standard to be reasonable); Simpson v. Young, 854 F.2d 1429 (D.C. Cir. 1988) (consumer advocacy public interest group challenged FDA conclusion that a certain color additive was safe; court deferred to agency decision); Cmty. Nutrition Inst. v. Young, 818 F.2d 943 (D.C. Cir. 1987) (public interest groups and consumers sued the FDA for its promulgation of “action levels” instead of formal tolerance levels for aflatoxins, a carcinogen, in corn; after the United States Supreme Court decided tolerance levels were not necessary, Court of Appeals found the action levels nevertheless needed to be promulgated pursuant to notice and comment rulemaking; standing conceded “under the broad grant of standing” found in the FFDCA and the APA); Nat’l Pork Producers Council v. Bergland, 631 F.2d 1353 (8th Cir. 1980) (trade organizations representing pork producers and meat packers challenged new USDA regulations allowing certain products not cured with nitrates (such as hot dogs) to be sold under their traditional names; court found regulation valid); Pub. Citizen v. Foreman, 631 F.2d 969 (D.C. Cir. 1980) (consumer advocacy public interest group sought declaratory judgment that nitrates were an unsafe food additive; court held that nitrites fell under the prior sanction exception to FDA responsibility; discussed infra note 162); Am. Pub. Health Assoc. v. Butz, 511 F.2d 331 (D.C. Cir. 1975) (public health advocacy organization sued Secretary of Agriculture for allegedly violating the Wholesome Meat Act and the Wholesome Poultry Products Act by refusing to affix safe handling instructions to raw meat and poultry; court held that labels as currently written were not false and misleading); Schuck v. Butz, 500 F.2d 810 (D.C. Cir. 1974) (citizen sued the Secretary of Agriculture seeking repeal of regulations permitting the use of nitrates and nitrites in meat products; court held that appellants had to petition for a rulemaking); Pub. Citizen v. Heckler, 653 F. Supp. 1229 (D.D.C. 1987) (consumer advocacy public interest group petitioned HHS for rule banning interstate sales of raw milk; court agreed with petitioners; discussed in detail, infra Part III.C); Pub. Citizen v. Dep’t of Health and Human Servs., 632 F. Supp. 220 (D.D.C. 1986) (consumer advocacy public interest group challenged provisional listing of nine color additives as safe for use; court held the listing to be consistent with
The paucity of citizen suits is not explained by the lack of citizen-suit provisions in the statutes regulating food safety, nor by the absence of a private right of action under the FFDCA, the FMIA, or the PPIA. A viable alternative for plaintiffs is to sue under the Administrative Procedure Act (APA), which prescribes procedural safeguards and establishes judicial review over federal regulatory agencies. And although such litigation will, and does, confront justiciability barriers—including challenges to standing and to the suits’ justiciability under the APA—these barriers are surmountable. Food-safety citizen litigation can be brought, and can be won, as demonstrated by prior cases.

This is not to minimize the advantage that a statutory citizen-suit provision provides a plaintiff. Citizen-suit provisions generally permit “any person” to sue certain persons, including government officials, who violate certain legal obligations, or who fail to carry out nondiscretionary duties. Judicial review over agency decision making is...
limited, and courts may be reluctant to delve too deeply into an agency's decision-making processes, especially if the decision at issue contains technical or scientific aspects. Whereas a plaintiff suing under a citizen-suit provision must show a violation of legal obligations under the relevant statute to win, a plaintiff suing pursuant to the APA must show that the agency's action was arbitrary and capricious. But, as the presence of a citizen-suit provision does not guarantee a plaintiff's ability to bring suit, in that he or she must still satisfy the constitutional and prudential requirements of Article III, neither does the absence of such a provision foreclose food-safety-impact litigation.

Citizen-suit provisions also provide for attorney's fees to the prevailing party. Although it may be possible for litigants against the federal government to be awarded fees and costs under other statutory provisions, it is unrealistic to think that organizations dedicated to food safety and committed to impact litigation would rely on attorney-fee provisions for their survival in any event. Consider environmental organizations dedicated to impact litigation. Although the availability of attorney's fees may assist in perpetuating organizations funding environmental-impact litigation, these fees are not necessary to the survival of these groups, nor are they the focus of litigation. These organizations are instead sustained through private donations and grants. Notably, environmental-impact litigation began before the passage of environmental legislation containing citizen-suit provisions.

74 See, e.g., 42 U.S.C. § 7604(d) (citizen suit provision of the Clean Air Act).
75 See Equal Access to Justice Act, 28 U.S.C. § 2412(d)(1)(A) (2006) (allowing fees against the federal government "unless the court finds that the position of the United States was substantially justified or that special circumstances make an award unjust.”; see also Hanover Potato Prods., Inc. v. Shalala, 989 F.2d 123 (3d Cir. 1993) (holding that a coalition of fresh potato producers, which was the prevailing party in its litigation against the FDA regarding a regulation, was entitled to fees under the Equal Access to Justice Act).
Food safety significantly overlaps with environmental protection. Beyond the actual commonalities between the two subjects—i.e., more responsible stewardship of land will lead to a safer food supply; cleaner agriculture and husbanding practices result in cleaner land—both food safety and environmental protection involve compulsory and comprehensive participatory systems. Everyone has to breathe the air, drink the water, live on the earth, and somehow nourish themselves. Beyond that, of course, individual discretion exists as to where, how, and what, although the extent of this individual discretion varies according to numerous demographic factors.

There are certain key differences between food safety and environmental protection, however, which have led to a vigorous impact litigation culture in the context of environmental law, and a virtual absence of one in the context of food safety. These differences include: (a) the nature of food versus that of the environment, and (b) the disparate historical development of (i) the regulatory structures overlaying the food-safety system and the management of the environment, and (ii) the advocacy movements concerned with issues touching on food safety and with issues regarding environmental protection.

A. The Nature of Food

Although both food and the environment involve compulsory systems, an individual’s relationship with each system is quite different. To begin with, there is no seemingly helpless entity in the food-safety context that needs an advocate on its behalf. Since its proposal in a law review article by Christopher Stone in 1972, and its citation in Justice Douglas’s dissent in *Sierra Club v. Morton*, the concept of trees and rocks having standing to sue in American law is a concept that, although repeatedly mocked, has had remarkable resilience and has even been recently revived. It is argued that, if inanimate objects and spaces cannot protect themselves in court, it stands to reason that advocates are needed to

---

represent them. No one, however, has argued that a tomato needs standing. Food itself is not perceived as requiring protection outside of its relationship to the human consumer. A consumer may choose to protect his consumption based on numerous factors—religious, cultural, economic, and health—but the food supply has no independent moral content. An exception to this is the prevention of animal cruelty, which is invested with a moral value beyond the consumption of animals by humans, and there is a vigorous tradition of advocacy surrounding the prevention of animal cruelty. The American Society for Prevention of Cruelty to Animals was founded in 1866, a quarter century before the founding of the Sierra Club. Animal welfare organizations have, in fact, brought a significant amount of the small body of existing food-safety impact cases.

Moreover, environmental protection is about the protection of physical space. Pre-harm environmental litigation is focused on protecting a particular space, whether it is an individual’s home, or somewhere imbued with environmental value. Geographic proximity to the area at issue allows a court a measurable index to assess actual injury.

Real property holds a unique and exalted place in American history and the American psyche, and consequently, in American law. The narrative of development in this country is one of territorial conquest, and the closing of the United States frontier in 1890 was a significant event. Property ownership has long been an American success symbol, fostered by governmental policies supporting homeownership. The law also

---


81 See, e.g., FREDERICK JACKSON TURNER, HISTORY, FRONTIER, AND SECTION (University of New Mexico Press 1993). Turner claimed that the frontier was the most important force in American society, culture, and politics, and that its disappearance would gravely affect the national character and mindset.

treats property uniquely, above and beyond necessity. For example, if a court, either state or federal, takes jurisdiction over real property, it may be able to enjoin any other court from continuing adjudication of the same property.\textsuperscript{83} Although the underlying policy reason behind this—to avoid inconsistent determinations over the same piece of property—is logical, it does not explain the absence of such safeguards over other adjudicatory subjects, such as bank accounts, where inconsistent judgments would carry the same detrimental consequences.\textsuperscript{84}

On the other hand, food is largely divorced from its origin—distributed and packaged as uniformly as possible so as to disguise any indication of where it is from. One of the hallmarks of our country’s food distribution system is its centralization, and the inability to trace the origins of much of the domestic food supply is one of the obstacles to minimizing the effect of foodborne illness outbreaks.\textsuperscript{85} Acknowledging this, the Food Safety Modernization Act includes a provision that enhances the capacity of the Secretary of Health and Human Services to trace food items for the purpose of improving the capacity of the government to detect and respond to food-safety problems.\textsuperscript{86}

The significance of food’s ubiquity and non-fixedness is threefold. First, it may be difficult for courts to analogize to environmental law in finding particularized injury to fulfill Article III standing requirements for food-safety impact litigants because there is no analog to the category of geographic proximity. Second, food is seen as more fungible than pieces of land. If a consumer believes that the regulatory scheme overlaying a particular food product is unlawfully inadequate, he may choose another comestible, whereas a homeowner who believes an environmental agency is acting unlawfully may have no recourse but to sell her property. This may lead to decreased incentive to litigate for stronger


\textsuperscript{84} See ERWIN CHEMERINSKY, FEDERAL JURISDICTION 746-47 (Wolters Kluwer, 5th ed. 2007).

\textsuperscript{85} For example, cantaloupe grown on a Colorado farm caused a multi-state outbreak of listeria poisoning in 2011. Authorities had difficulty, however, determining to which states exactly the tainted cantaloupe were shipped, thus complicating the attempt to stem the outbreak. See, e.g., Michael Booth, \textit{More States May Have Received Listeria-Contaminated Cantaloupe}, \textit{DENVER POST} (Sept. 22, 2011, 3:35 PM), http://www.denverpost.com/breakingnews/ct_18955276.

regulation, because if an individual thinks chicken inspection is inadequate, that individual can choose not to buy chicken.\textsuperscript{87}

Third, because of the difficulty of knowing where one’s food comes from, and the invisibility of the processes that led to its packaging in its current consumable form, citizens have less personal control over determining the source of any foodborne illness. One cannot see the pathogens that spread foodborne illness, nor is the consumer exposed to the production facility where the pathogen originated. By contrast, even if the pollutant itself targeted by environmental litigants is invisible, it is most likely that its origin is not.\textsuperscript{88} This inability to self-trace in the food-safety context leads to the necessity for a reliance on experts and scientific evidence at an early stage of litigation, and is compounded in the case of impact litigation.

B. The Disparate Historical Development of Food Safety and Environmental Protection

Major milestones in American environmentalism, including developments in the governmental management of natural resources, as well as the ferment of citizen activity, coincide roughly with major milestones in the developments of a national food-safety regulatory regime. These milestones generally took place at times of urbanization and technological advancement, which led to actual threats on the country’s natural resources and food supply, as well as to national anxiety regarding threats to traditional ways of life, and consequentially, to human health.

However, the regulatory and legislative structures developed very differently in the two fields. In the context of environmental protection, the government organized entities to manage natural resources but major environmental legislation was not passed until the early 1970s, whereas in the context of food-safety regulation, we see the amendment and supplementation of several major statutes that originated at the beginning of the twentieth century. We also see that, in environmental protection, citizen advocacy groups were always an important counterpart to governmental regulation, a situation that did not exist in the context of food safety.

\textsuperscript{87} Of course, consumer choice is constrained by numerous factors, including cultural norms, economic realities, and governmental policy.

1. A Brief History of Environmental Protection in the United States

The first wave of American environmentalism took place at the turn of the twentieth century in reaction to a massive increase in immigration and urbanization. From 1860 to 1890 the United States population increased from thirty-one million to seventy-five million people, and between 1860 and 1900 the number of people living in urban areas doubled from 20 percent to 40 percent. In 1890, the United States Census Bureau declared the United States frontier officially closed, and in 1893, Frederick Jackson Turner articulated his famous thesis as to the effect of the frontier, and its closing, on American history. Urbanization resulted in the need for the American people to refashion their relationship with the natural resources of the country.

A small group of prominent intellectuals and public figures brought the perceived necessity for the conservation of the American wilderness to the nation’s attention. The Sierra Club was founded in 1892, and the National Audubon Society in 1905. Theodore Roosevelt’s administration embarked on a concerted campaign to forward the regulated use of resources towards the goal of the fullest use for the present generation. Before Roosevelt left office in 1909, he worked with the head of the Forest Service and the secretary of the interior to withdraw over four million acres of the public domain from consideration for private sale.

In the 1930s, a series of natural disasters began to make apparent the consequences that could ensue from the unregulated exploitation of the nation’s resources. For

---


See Turner, supra note 81.

See Stephen Fox, The American Conservation Movement: John Muir and His Legacy 110 (University of Wisconsin Press 1981). Historians of the American environmental protection movement divide the early twentieth century environmental movement into two strains, the conservationist—which measured nature’s value according to its worth for mankind—and the preservationist—which was ecologic and biocentric. See id. at 108; Joseph M. Petulla, American Environmentalism: Values, Tactics, Priorities 26 (Texas A & M Press 1980).


McGeary, supra note 93, at 116-17.
example, the dust storms of the 1930s, which rendered nine million acres of formerly arable land unusable by 1938, were partly the result of haphazard farming practices.\textsuperscript{95} After World War II, several organizations concerned with conserving the country's natural resources became involved in major, and public, environmental battles,\textsuperscript{96} and these organizations, including the Sierra Club and the Wilderness Society, had over three hundred thousand members by 1960.\textsuperscript{97}

It was not until the publication of Rachel Carson's monumentally influential book \textit{Silent Spring} in 1962, however, that the American populace was galvanized to the cause of environmental protection. \textit{Silent Spring} addressed the effects of DDT and other pesticides on human health and the environment.\textsuperscript{98} The book was on the \textit{New York Times} bestseller list for thirty-one weeks and sold over a half million hardcover copies.\textsuperscript{99}

The 1960s also saw the advent of strategic environmental litigation brought by public interest groups as administrative law challenges. Beginning with the 1965 case \textit{Scenic Hudson Preservation Conference v. Federal Power Commission}, in which the Second Circuit allowed citizens standing to sue under the Federal Power Act to overturn three orders of the Federal Power Commission based on their "interest in the aesthetic, conservational, and recreational aspects of power development,"\textsuperscript{100} newly formed environmental groups aggressively sought judicial review of administrative action.\textsuperscript{101} The organizations bringing these challenges faced hurdles, especially regarding their standing to sue (although \textit{Scenic Hudson} was a milestone in that regard), but pursued suits nonetheless.\textsuperscript{102}

Inspired by this movement, a flood of environmental legislation was passed in the early 1970s.\textsuperscript{103} Congress intended

\textsuperscript{95} DONALD WORSTER, NATURE'S ECONOMY: A HISTORY OF ECOLOGICAL IDEAS 221-53 (Cambridge University Press 2d ed. 1994).
\textsuperscript{97} Id. at 5-6.
\textsuperscript{98} RACHEL CARSON, SILENT SPRING (Houghton Mifflin Co. 1994) (1962).
\textsuperscript{99} SALE, supra note 96, at 4.
\textsuperscript{100} 354 F.2d 608, 616 (2d Cir. 1965).
\textsuperscript{102} A. Dan Tarlock, \textit{The Story of Calvert Cliffs}, in \textit{ENVIRONMENTAL LAW STORIES} 81 (Foundation Press 2005).
\textsuperscript{103} This legislation includes the National Environmental Policy Act of 1969, 42 U.S.C. §§ 4321-4370f (2006); the Clean Air Act of 1970, 42 U.S.C. §§ 7401-7671q (2006);
citizens and advocacy groups to have a role in enforcing the new environmental legislation,\textsuperscript{104} and these statutes contain citizen-suit provisions, a legal mechanism encouraging citizen actions against polluters. These powerful provisions permit “any person” to sue certain persons, including government officials, who violate certain legal obligations, or who fail to carry out nondiscretionary duties.\textsuperscript{105} The remedies available in such suits are injunctive relief, civil penalties—which go to the federal treasury, and the recovery of attorney’s fees and costs.\textsuperscript{106} The citizen-suit provisions were written into environmental laws partly to encourage the very litigation that was already taking place.\textsuperscript{107}

2. A Brief History of Food-Safety Regulation in the United States

The urbanization of the country that took place at the turn of the twentieth century also resulted in massive changes to the relationship between individuals and the food supply. Consolidation of the food supply and the creation of distribution networks became necessary to feed large populations of people who could not grow their own food. The end of the nineteenth century also saw extensive development in food science, especially in the creation of new food additives.\textsuperscript{108} At that time, only certain imported foods were subject to federal regulation—all other food regulation was state and local.\textsuperscript{109}

Public awareness of the need for national regulation was raised during the late nineteenth century by the head of the U.S. Bureau of Chemistry, Harvey W. Wiley, M.D., who campaigned

\begin{footnotes}
\item[104] Indeed, environmental organizations assisted in the drafting of certain of the new statutes. See Adams et al., supra note 77, at 51, 53 (explaining how two members of the Natural Resources Defense Council helped to draft the 1977 amendments to the Clean Air Act); Tarlock, supra note 102, at 92 n.51 (recounting the influence of Myron Cherry, an important anti-nuclear lawyer in the 1970s, on the drafting of NEPA and its legislative history).
\item[106] It is important to keep in mind, however, that the existence of a citizen-suit provision does not satisfy Article III standing requirements; a litigant must still show injury-in-fact, causation, and redressability before her suit may proceed. Lujan v. Defenders of Wildlife, 504 U.S. 555, 562, 571-78 (1992).
\item[107] Adams et al., supra note 77, at 27.
\item[108] Fortin, supra note 20, at 5.
\item[109] Hutt, supra note 21, at 6.
\end{footnotes}
for a federal food and drug law. The press also began to expose some safety problems with commonly used food preservatives and dyes.\textsuperscript{110} A series of damaging newspaper articles about the food industry as well as the publication in 1905 of Upton Sinclair's \textit{The Jungle} garnered public support for the national regulation of the food supply, and in 1906, Congress passed the Pure Food and Drug Act and the Meat Inspection Act.\textsuperscript{111} The Food and Drug Act was to be administered by the Bureau of Chemistry within the USDA, and the FMIA by the Bureau of Animal Industry, also within the USDA.\textsuperscript{112}

The focus of each of the 1906 acts was to prevent the adulteration of food, and the Food and Drugs Act included the prohibition of additives that would be deleterious to human health as well as of substances that would dilute the product for the purpose of making the food cheaper to produce.\textsuperscript{113} The Meat Inspection Act prohibited adulterated or misbranded livestock products to be sold as food, and also mandated the improvement of sanitation at slaughtering facilities, in response to Sinclair's book.\textsuperscript{114}

The Bureau of Chemistry was split in 1927, with regulatory functions, including responsibility for the 1906 Act, taken over by the Food, Drug, and Insecticide Administration (which became the FDA in 1930).\textsuperscript{115} Responsibility for the Meat Inspection Act was not transferred, however, and authority for its enforcement remained within the Department of Agriculture. FDA officials began advocating for the modernization of the 1906 Act in 1933,\textsuperscript{116} but passage of a new bill was stalled until 1938.\textsuperscript{117} Precipitating passage of the 1938 Federal Food, Drug, and Cosmetics Act (FFDCA) was the death of over one hundred people in 1937 who had taken an antibiotic that had been mixed with a sweet substance to improve its

\textsuperscript{110} \textit{Fortin, supra} note 20, at 5-6. Wiley established a “Poison Squad”—a group of volunteers who consumed food additives, including boric acid and formaldehyde, to assess their effects on the human body. \textit{Id.} at 5; see also \textit{About FDA: Milestones in U.S. Food and Drug Law History}, U.S. FOOD & DRUG ADMIN., \url{www.fda.gov/AboutFDA/WhatWeDo/History/Milestones/default.htm} (last visited Jan. 13, 2012).

\textsuperscript{111} \textit{Fortin, supra} note 20, at 6.


\textsuperscript{113} Hutt, \textit{supra} note 21, at 6.

\textsuperscript{114} \textit{Fortin, supra} note 20, at 6.

\textsuperscript{115} \textit{Fortin, supra} note 20, at 6.

\textsuperscript{116} \textit{About FDA: Significant Dates in U.S. Food and Drug Law History}, U.S. FOOD & DRUG ADMIN., \url{http://www.fda.gov/AboutFDA/WhatWeDo/History/Milestones/ucm128305.htm} (last visited Jan. 13, 2012).

\textsuperscript{117} Hutt, \textit{supra} note 21, at 7.

\textsuperscript{118} HUTT, MERRILL & GROSSMAN, \textit{supra} note 25, at 13.
taste, but had never been tested for safety. In 1940, the FDA was transferred from the USDA to the Federal Security Agency, and in 1988, to the Department of Health and Human Services, with a commissioner of food and drugs appointed by the President with the advice and consent of the Senate.

Since 1938, the FFDCA has been amended hundreds of times. Significantly, during World War II, food processing technology was developed to preserve and transport food for war, and this massive alteration in the general food supply led to public concern about the addition of synthetic ingredients and potential carcinogens to the food supply. The FDA reacted by passing the 1958 Food Additives Amendment and the 1960 Color Additive Amendment. The 1970s and 1980s saw such food-safety amendments to the FFDCA as the Low-Acid Food Processing Regulations (1973). The Nutritional Labeling and Education Act was passed in 1990.

As to meat and poultry, the Poultry Products Inspection Act (PPIA) was passed by Congress in 1957 in response to a rapidly expanding poultry industry, which was also developing poultry processing techniques. In 1967, the FMIA was amended as the Wholesome Meat Act, which requires the inspection of all meat, and the PPIA was amended in 1968. In 1971, the Animal and Plant Health Inspection Service, a division of the USDA, was charged with meat and poultry inspection, and this was assigned to the Food Safety and Quality Service in 1977. The Food Safety and Quality Service became the Food Safety and Inspection Service (FSIS) in 1981.

Food-safety advocates have never embraced impact litigation as a way to achieve food-safety goals. Even food-safety organizations claiming to use litigation as a strategy to achieve their goals use it sparingly. For example, the Center for Food Safety, which was established in 1997, and is a “non-profit public interest and environmental advocacy membership

---

118 Id.; FORTIN, supra note 20, at 6.
119 About FDA: Significant Dates in U.S. Food and Drug Law History, supra note 115.
120 See HUTT, MERRILL & GROSSMAN, supra note 25, at 14-15.
121 FORTIN, supra note 20, at 7; HUTT, MERRILL & GROSSMAN, supra note 25, at 393-94. Concern about carcinogens and pesticides was also, at this time, galvanizing the environmental movement.
123 FORTIN, supra note 20, at 8.
organization [that] combines multiple tools and strategies in pursuing its goals, including litigation and legal petitions for rulemaking, is listed as a party in fewer than forty cases since its founding. The Center for Science in the Public Interest, an influential consumer advocacy organization focused on improving public health, nutrition, and food safety, is listed as a party in under twenty cases since its founding in 1971. By comparison, since its founding in 1967, the Environmental Defense Fund is listed as a party in approximately 280, and since 1970, the year of its founding, NRDC is listed as a party in over 750 cases. Since 1997, when the Center for Food Safety was founded, the NRDC is listed as a party in over 400 cases.

This comparison of impact litigation in the realms of environmental protection and food safety serves to illustrate how environmental advocates and food-safety advocates have different approaches to advocacy and effective means of change. It is crucial to emphasize that the difference cannot be explained by the lack of citizen-suit provisions in the food-safety laws. As mentioned above, the passage of environmental legislation in the 1970s was not the beginning of environmental-impact litigation. Although environmental-impact litigation accelerated greatly after the passage of environmental legislation, strategic litigation had been imagined and implemented in the decade prior, as environmental advocacy organizations, both old and new, began to litigate to protect the environment.

---

127 In 2005, CSPI hired Stephen Gardner to direct its food-safety litigation, which is focused on private litigation against companies that refuse to take action to improve food safety. This strategy has been remarkably successful in improving the food safety of market-leading companies, as well as improving relationships between these companies and advocacy organizations, such as CSPI. Telephone Interview with Stephen Gardner (Jan. 5, 2011).
129 See, e.g., Sive, supra note 101, at 729 (“In no other political and social movement has litigation played such an important and dominant role [as in the environmental movement].” (citation omitted)).
130 ADAMS & ET AL., supra note 77, at 53; Tarlock, supra note 102, at 82 n.15; Sive, supra note 101, at 731.
Nor can the disparity be explained by pointing to the superior funding of the environmental advocacy groups—although the NRDC was founded in part with money from the Ford Foundation, this money was granted after the idea to focus on litigation was formed, not before.\footnote{131} It may be impossible to actually identify why food-safety litigation is not a prevalent form of advocacy and activism, but it is important to note that the absence of the development of such a tradition was not dictated by legal constraints, nor prescribed by social circumstances.

III. JUSTICIABILITY BARRIERS TO FOOD-SAFETY-IMPACT LITIGATION

Food-safety-impact litigation faces difficulties getting into court because it involves multiple—including associational—parties, probabilistic harms, widely-shared harms, and requests for prospective relief. The judicial system is still struggling to adapt traditional doctrines of justiciability to such litigation, even though such litigation has been ongoing in various contexts since at least the middle of last century.\footnote{132} Although it is relatively uncontroversial that suits based on uncertain injury \textit{can} be heard by federal courts,\footnote{133} the questions of who can bring such suits, when they are ready for suit, and where the suits can be brought are still vigorously contested, with the answers changing by jurisdiction, and over time. These questions of who, when, and where must be determined before courts can reach the merits.

The justiciability barriers most likely to be faced by food-safety impact litigants are standing challenges and challenges to justiciability under the APA.\footnote{134}

A. \textit{Standing}

Standing determines whether a federal court litigant is the proper party to bring the suit before the court.\footnote{135} The notion

\footnote{131} Adams & Et Al., \textit{supra} note 77, at 17-24.

\footnote{132} All social justice litigation falls into this model, including litigation in the fields of civil rights and environmental protection.

\footnote{133} Abbott Labs. v. Gardner, 387 U.S. 136 (1967); Brandt v. Vill. of Winnetka, 612 F.3d 647 (7th Cir. 2010).

\footnote{134} Although ripeness and mootness challenges in this context are conceivable, indeed likely, I focus here on the actual cases and the barriers with which courts have actually grappled when dealing with this type of litigation.
of standing has both constitutional and prudential dimensions. A court must first determine if the litigant meets the criteria of Article III: (1) whether the litigant has suffered an injury-in-fact, which must be “concrete and particularized,” and “actual or imminent”; (2) whether the alleged injury is fairly traceable to the harm targeted; and (3) whether the remedy sought is likely to redress the alleged injury. Prudential concerns include prohibitions against the litigation of generalized grievances and litigating the rights of third parties, and determining whether the litigant is within the zone of interests of the statute at issue.

In 1992, the Supreme Court made it more difficult for impact plaintiffs to establish standing, even when they are suing pursuant to a citizen-suit provision. In Lujan v. Defenders of Wildlife, the Court denied standing to several environmental organizations challenging a regulation as violating the Endangered Species Act. The Court found that the organizations did not adequately allege injury in fact. Justice Scalia, writing for the majority, wrote, “when the plaintiff is not himself the object of the government action or inaction he challenges, standing is not precluded, but it is ordinarily ‘substantially more difficult’ to establish.”

Most significantly, the decision prohibited Congress from statutorily creating a legal injury and thereby bestowing standing upon citizens. The Court held that a citizen-suit provision did not eliminate the need for a plaintiff to show that she has sustained direct and personal injury. In Lujan, the

---

135 For a comprehensive discussion and critique of the separation of powers rationale for the standing doctrine, see Heather Elliott, The Functions of Standing, 61 STAN. L. REV. 459 (2008).
136 See, e.g., Lujan v. Defenders of Wildlife, 504 U.S. 555, 560 (1992). Whether these three criteria actually stem directly from Article III, or are themselves judge-made and perhaps misguided, is a matter of discussion among commentators. See, e.g., William A. Fletcher, The Structure of Standing, 98 YALE L.J. 221, 229 (1988); Sunstein, supra note 63, at 185-86. Nevertheless, used as they have been as Article III criteria for the last four decades, they are, at this point, accepted as constitutional.
138 Lujan, 504 U.S. at 578.
139 A plurality of the Court also found that the plaintiffs had not established redressability. Id. at 568-71.
140 Id. at 562 (citation omitted).
141 The Court explained that Congress could still create legally cognizable injuries, i.e., it could “elevat[e] to the status of legally cognizable injuries concrete, de facto injuries that were previously inadequate in law.” Id. at 578. What Congress could not do was “abandon . . . the requirement that the party seeking review must himself have suffered an injury.” Id. (citation omitted).
142 Id. at 573.
Court of Appeals had held that the organizational petitioners had adequately alleged injury in fact because they alleged a procedural injury—that the Secretary of the Interior had failed to consult as required by the ESA—and that this procedural injury was adequate because of the citizen-suit provision, even if the petitioners failed to allege a personal injury. The *Lujan* Court rejected this view, explaining that

> [t]o permit Congress to convert the undifferentiated public interest in executive officers' compliance with the law into an “individual right” vindicable in the courts is to permit Congress to transfer from the president to the courts the Chief Executive's most important constitutional duty, to “take Care that the Laws be faithfully executed.”

The Court held that a litigant must show that she herself had suffered “injury in fact,” and had a tangible and concrete stake, beyond the vindication of a right as a citizen, in the outcome of the case.

By requiring a showing of personalized injury-in-fact even when alleging procedural injury, the Court in *Lujan* made it significantly more difficult for the beneficiaries of regulation to protect their interests. Commentators predicted that citizen-suit environmental litigation would be severely restricted, or even eliminated, and the environmental advocacy community began to plan different tactical routes toward the enforcement and strengthening of environmental litigation. The Court reinforced its strict interpretation of Article III standing in *Steel Co. v. Citizens for a Better Environment*, which held that plaintiffs did not have Article III standing based on a lack of redressability when the violations for which the plaintiffs sued had occurred solely in the past.

The standing cases of the late 1980s and early 1990s demonstrate a general tightening of standing law and a trend toward the restriction of access to courts for regulatory beneficiaries in public law litigation. This trend did not, however, reflect a unified court, nor consensus among the justices, and in the late 1990s the Court began to relax its approach to standing.

---

143 *Id.* at 571, 572.
144 *Id.* at 577.
145 *Id.* at 575-78.
146 *Buzbee, supra* note 76, at 214-20.
In *Friends of the Earth, Inc. v. Laidlaw Environmental Services*, the Court, with a vigorous dissent by Justice Scalia, the author of the *Lujan* and *Steel Co.* decisions, backed away from the broadest implications of the *Lujan* decision, and granted standing to several environmental organizations that had instituted a citizen suit against a company for alleged Clean Water Act violations.\(^{148}\) The Court held that the availability of civil penalties against the company was adequate to provide redressability for the citizen plaintiffs because of the potential that such penalties would deter future violations, although injunctive relief against the company was not available because the company had ceased violations since the commencement of the litigation.\(^{149}\) Distinguishing *Steel Co.*, the Court explained that in *Laidlaw* the violations had not ceased prior to suit as in *Steel Co.*, but had been ongoing at the time suit was commenced.\(^{150}\) Moreover, the Court found that the absence of evidence of injury to the environment was irrelevant to the injury-in-fact analysis because the plaintiffs were harmed by the lessening of the aesthetic and recreational values of the area.\(^{151}\)

After the Court's movement away from its extremely restrictive standing decisions of the early 1990s, it appears that citizen suits were once again viable. *Laidlaw* and other contemporaneous cases\(^{152}\) signaled the Court's return to deference to legislatively defined injuries, and thereby wedged the doors to the courtroom back open for legislative beneficiaries.\(^{153}\)

The Court's decision in *Massachusetts v. EPA* further affirmed this trend. In *Massachusetts v. EPA*, the Court found that Massachusetts had standing to challenge the EPA's refusal to regulate greenhouse gas emissions from motor

---

149 Id. at 173-74.
150 Id. at 187-88.
151 Id. at 181.
153 In *Akins*, a group of voters had challenged a Federal Election Commission's determination that the American Israel Public Affairs Committee (AIPAC) was not a "political committee" as defined by statute, and, consequentially, that AIPAC did not have to disclose certain information. See *id.* at 13-14. The Court held that the petitioners had standing to challenge the FEC's determination because Congress had specifically granted such parties the right to sue; the statute in question contained a provision that "any person who believes a violation of this Act . . . has occurred, may file a Complaint with the Commission." *Id.* at 19 (citing 2 U.S.C. § 437g(a)(1) (1971)). The Court wrote that "[t]he 'injury in fact' that respondents have suffered consists of their inability to obtain information . . . that, on respondents' view of the law, the statute requires that AIPAC make public." *Id.* at 21.
vehicles under the Clean Air Act.\textsuperscript{154} The Court provided Massachusetts with “special solicitude in [the] standing analysis” based on its sovereign status,\textsuperscript{155} but also found, with a potentially wide reach, that climate-change risks pose a concrete and particularized injury to Massachusetts as a landowner, even though the harm may be widely shared, and that the alleged injury (global warming) would be lessened by the requested remedy (a reduction in motor vehicles emissions), if not eliminated.\textsuperscript{156} The full reach of the Court’s standing and redressability analyses remains to be seen.\textsuperscript{157}

The zone-of-interests test is also significant in any discussion of food-safety-impact litigation. This test is a prudential standing requirement fashioned by the Supreme Court providing that the interest alleged by the plaintiff must arguably be within the zone of interests protected by the statute or constitutional provision at issue.\textsuperscript{158} To satisfy this test, congressional intent to benefit the plaintiff is not required.\textsuperscript{159} In 1997, the Supreme Court found that the citizen-suit provision of the Endangered Species Act negated the zone-of-interests test in relation to that statute.\textsuperscript{160} In the same case, the Court also noted the “generous review provisions” of the APA, and clarified that the zone-of-interests test should be assessed, not in relation to the overall purpose of the statute at issue, but rather in relation to the specific provision relied upon.\textsuperscript{161}

B. Standing in Food-Safety-Impact Litigation

Standing is a critical issue in food-safety-impact cases. Three of the six post-1992 decisions in food-safety impact cases

\textsuperscript{155} For a discussion of the potential implications of the Court’s reliance on Massachusetts’ sovereign status in its standing analysis, see Amy J. Wildermuth, Why State Standing in Massachusetts v. EPA Matters, 27 J. LAND RESOURCES & ENVTL. L. 273 (2007).
\textsuperscript{156} Massachusetts, 549 U.S. at 518-23, 525.
\textsuperscript{157} Maxwell L. Stearns argues that the Roberts Court will continue to expand standing doctrine in Standing at the Crossroads: The Roberts Court in Historical Perspective, 83 NOTRE DAME L. REV. 875 (2008).
\textsuperscript{160} Bennett v. Spear, 520 U.S. 154, 166 (1997). It is unclear whether this holding applies to all statutes with citizen-suit provisions or is confined to the ESA. See, e.g., Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dep’t of Agric., 415 F.3d 1078, 1102 n.17 (9th Cir. 2005).
\textsuperscript{161} Bennett, 520 U.S. at 163 (citation omitted).
that I found involved standing issues. Each of these cases was a challenge to agency action or inaction brought for the ostensible purpose of minimizing the risk of foodborne illness in the United States. One survived a motion to dismiss by demonstrating injury in fact, the second survived an injury in fact challenge only to be dismissed for a lack of redressability, and the third had one of its claims dismissed because the plaintiffs failed to show that they were within the zone of interests protected by the statute.

1. Using Increased Risk of Harm as a Means to Show Injury-in-Fact in Food-Safety-Impact Cases

Plaintiffs in food-safety-impact suits must find a way to show that they have suffered concrete and particularized injury from the challenged regulation although they have not contracted a foodborne illness. Although they may represent themselves as a consumer of the regulated food, they must show that their grievance is more than a generalized one.

Several courts have held that an allegation of increased risk of injury due to a challenged agency action suffices to show such a concrete and particularized injury, but this trend has mainly, although not always, been confined to the environmental context. In *Baur v. Veneman*, however, a 2003 case regarding mad cow disease, the Second Circuit reversed the district court’s dismissal of the case for lack of Article III standing, finding that increased risk of harm—based on the increased...
risk of contracting a foodborne disease—satisfied the injury in fact requirement in the context of food and drug safety suits.¹⁶⁵

The plaintiff in this case, Michael Baur, challenged the USDA’s and the FDA’s allowance of “downed” cattle into the food supply.¹⁶⁶ “Downed” cattle are cattle that are too sick to stand or walk before slaughter, and, at the time, USDA regulations allowed downed cattle to enter the food supply after inspection. Baur alleged that downed cattle were more likely to carry transmissible spongiform encephalopathies (TSEs), which are progressive neurological diseases.¹⁶⁷ The most common of these is bovine spongiform encephalopathy (BSE), which is known as “mad cow disease.”¹⁶⁸ Baur claimed that the downed cattle policy violated both the FMIA and the FFDCA.

After Baur’s petition to the USDA and the FDA was denied, he brought suit in district court under the APA seeking judicial review of the FSIS’s decision.¹⁶⁹ Baur claimed standing

¹⁶⁵ Baur v. Veneman, 352 F.3d 625, 628 (2d Cir. 2003).
¹⁶⁶ In 1998, Michael Baur and Farm Sanctuary, Inc., an animal protection organization, filed a petition with the USDA and the FDA requesting that the agencies “label all downed cattle as adulterated,” under the Federal Food, Drug, and Cosmetic Act (FFDCA), Section 342(a)(5). Baur, 352 F.3d at 628. This section of the FFDCA provides that any food that is “the product of a diseased animal” is adulterated. 21 U.S.C. § 342 (a)(5) (2006). The FFDCA prohibits the manufacture, delivery, receipt, or introduction of adulterated food “into interstate commerce.” Id. § 331.
¹⁶⁷ Baur, 352 F.3d at 627-28.
¹⁶⁸ Id. at 627.
¹⁶⁹ Michael Baur originally brought suit with Farm Sanctuary, Inc., an animal welfare organization. Farm Sanctuary claimed that its members were injured when they observed the treatment of animals at slaughterhouses. The district court dismissed Farm Sanctuary’s claims because it had failed to state an interest within the zone of interests of the FMIA. Farm Sanctuary did not appeal its dismissal on standing grounds, and the Second Circuit opinion only discusses Baur. For that reason, I refer only to Baur, although he was joined by Farm Sanctuary at early stages of the litigation. See Farm Sanctuary, Inc. v. Veneman, 221 F. Supp. 2d 280, 284-85 (S.D.N.Y. 2002).

Note that there is no requirement in the Article III injury-in-fact standing inquiry that the plaintiff’s alleged reason for bringing suit is genuine. The court does not address whether Baur’s alleged injury—increased risk of foodborne illness—is genuine, or, in other words, if that is Baur’s real motive for being before the court, which in this case, was open to question. Farm Sanctuary is a prominent animal protection organization, founded in 1986 by Gene Baur “to combat the abuses of factory farming and to encourage a new awareness and understanding about ‘farm animals.’” About Us, FARM SANCTUARY, http://farmsanctuary.org/about/ (last visited Jan. 13, 2012). The organization opposed what it saw as the unusually cruel practice of dragging a cow that has collapsed on the way to the slaughterhouse to be killed. Robert Terenzi, Jr., When Cows Fly: Expanding Cognizable Injury-in-Fact and Interest Group Litigation, 78 FORDHAM L. REV. 1559, 1561 (2009). Many of the members of Farm Sanctuary, however, were vegan. For this reason, Michael Baur, Gene Baur’s brother and a Fordham Law professor, joined the case to provide a meat-eating plaintiff. Id. A court may delve deeper into a plaintiff’s asserted reason for bringing suit if there is a disjunction between an association’s stated reason and the main purposes for the association’s existence. If the association’s purpose for existence does not fall into the zone of interests protected by the statute at issue, the association’s suit may be
based on his status as a consumer of meat who was at increased risk of contracting a foodborne illness because of the USDA's policies regarding downed cattle. The district court dismissed Baur’s claims for lack of Article III standing, finding that because there was, as yet, no evidence of BSE in the United States, the harm that he alleged was too speculative and not sufficiently particularized to support standing.

The Second Circuit reversed. First, the court held that Baur’s increased risk of harm claim was capable of satisfying the injury in fact requirement. Baur sued under the FMIA and the FFDCA, and the court recognized that the purpose of these statutes is, in part, to protect the nation’s food supply and minimize the risk from dangerous food. There was a “tight connection” between the injury alleged and the allegedly violated statutes. The court explained that

> Although this type of injury has been most commonly recognized in environmental cases, the reasons for treating enhanced risk as sufficient injury-in-fact in the environmental context extend by analogy to consumer food and drug safety suits. Like threatened environmental harm, the potential harm from exposure to dangerous food products or drugs “is by nature probabilistic,” yet an unreasonable exposure to risk may itself cause cognizable injury.

The court also found that Baur had shown that he himself faced a credible harm because “the probability of harm which a plaintiff must demonstrate in order to allege a cognizable injury-in-fact logically varies with the severity of the probable harm.” Thus because of the severity of contracting dismissed for lack of standing. See Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dept of Agric., 415 F.3d 1078, 1103-04 (9th Cir. 2005).

170 Baur, 352 F.3d at 630.
171 Id. at 631.
172 Id. at 636.
173 Id. at 634-35.
174 See William A. Fletcher, The Structure of Standing, 98 YALE L.J. 221 (1988). Fletcher advocates a return to the “legal interest” test for standing, which asks only whether the plaintiff has a legal right bestowed by the legal provision under which he is suing, and if so, does not require an injury in fact. He writes that the APA was meant to provide a flexible standing rule, and that if a litigant bases suit on a statute, the court should look to the relevant statute to determine whether the litigant should have standing, not whether the litigant has suffered an “injury in fact.” Id. at 255-65. Although not throwing over the injury in fact requirement, the Second Circuit’s decision in Baur incorporates Fletcher’s viewpoint by looking at the degree of connection between the injury alleged and the statutes implicated.

175 Baur, 352 F.3d at 634 (citing Friends of the Earth, Inc. v. Gaston Copper Recycling, Corp., 204 F.3d 149, 160 (4th Cir. 2000)).
176 Id. at 637. The court explained that although the standard was lenient at the pleading stage, the plaintiff could still not rely on conclusory allegations to show standing. Id. This analysis may have been different had the case been decided after the
mad cow disease, which is fatal and has no known cure, the
court held that the increase in risk may be moderate for
standing purposes. The court found the fact that Baur’s
allegations were supported by government studies supported
his claim for standing, as did the fact that his alleged increased
risk of harm resulted from an “established governmental
policy.” Incidentally, after the Second Circuit decision in Baur
was filed, on December 23, 2003, a cow in Washington State
was diagnosed with BSE. Soon thereafter, the USDA passed a
regulation banning downed cattle from the food supply, and
the case became moot.

While increased risk of harm is a widely recognized
basis for injury in fact, it is not entirely uncontroversial. The
Baur Court commented that “the courts of appeals have
generally recognized that threatened harm in the form of an
increased risk of future injury may serve as injury-in-fact for
Article III standing purposes,” and that “[w]ithout questioning
standing, the Supreme Court has decided cases in which it
appeared to assume that enhanced risk may cause real
injury.” However, as mentioned above, increased risk of harm
has rarely been used outside of environmental law cases.
Indeed, the Second Circuit refused to sanction the doctrine
generally, but instead held only that “[i]n the specific context of

United States Supreme Court’s decisions in Ashcroft v. Iqbal, 556 U.S. 662 (2009), and
Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007), which instructed district courts
that a plaintiff needed factual allegations that “raise a right to relief above the
speculative level,” and must “state a claim to relief that is plausible on its face.”
Twombly, 550 U.S. at 555, 570. Lower courts are now grappling with the implications
of applying the “plausibility standard,” to determinations regarding challenges to
subject matter jurisdiction, including standing challenges. See, e.g., Coal. for a
2010); Eugster v. Wash. State Bar Ass’n, No. CV 09-357-SMM, 2010 WL 2926237 (W.D.

Baur, 352 F.3d at 637. The dissent in Baur argued that Baur had not
shown that he himself faced a credible harm of contracting BSE. The dissenting Judge
found the absence of evidence of BSE in the United States to be particularly
significant, and wrote that although Baur may be correct in his allegations that the
USDA should act differently to prevent an outbreak of BSE in the country, he “cannot
properly use this Court as vehicle to advance the claims to proper policy.” Id. at 652.

Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dept
of Agric., No. 05-35264, 2005 U.S. App. LEXIS 17360, at *18 (9th Cir. Aug. 17, 2005).
9 C.F.R. § 309.2(b) (2010). For a discussion of the USDA’s regulatory
response to mad cow disease, and an argument that the agency has acted
incompetently and inefficiently, see Jason R. Odehoo, No Brainer? The USDA’s
Regulatory Response to the Discovery of Mad Cow Disease in the United States, 16

For a discussion of whether risk itself is a harm, see Claire Finkelstein, Is

See Baur, 352 F.3d at 633 & n.7 (listing cases).
food and drug safety suits... we conclude that such injuries are cognizable for standing purposes, where the plaintiff alleges exposure to potentially harmful products.\textsuperscript{182}

The Second Circuit cited cases from the Fourth Circuit, the Seventh Circuit, the Ninth Circuit, and the D.C. Circuit to show the general acceptance of increased risk of harm as a basis for Article III standing. All of the cited cases, except those in the Seventh Circuit, were in the environmental context.\textsuperscript{183} And although it has recognized that increased risk can be a basis for standing, the D.C. Circuit has taken a strict view of whether increased risk of harm constitutes injury in fact. In \textit{Natural Resources Defense Council v. Environmental Protection Agency}, the court vacated its earlier decision which had dismissed NRDC’s petition for lack of standing.\textsuperscript{184} In this case, NRDC sued the EPA, charging that the agency’s issuance of a rule establishing exemptions from an international treaty that mandated the reduction of the use of methyl bromide—a substance that degrades the ozone layer—violated both the treaty and the Clean Air Act.\textsuperscript{185} The court initially held that NRDC’s claim that its members faced a greater chance of contracting skin cancer and other illnesses under the EPA rule was too hypothetical to constitute injury-in-fact.\textsuperscript{186} NRDC moved for rehearing, and both NRDC and EPA, in its opposition to the petition for rehearing, presented new

\textsuperscript{182} Id. at 634.

\textsuperscript{183} Id. at 633. Courts have occasionally recognized increased risk of harm as a basis for Article III standing in contexts outside of environmental law. For example, in \textit{Sutton v. St. Jude Medical S.C., Inc.}, 419 F.3d 568 (6th Cir. 2005), the Sixth Circuit permitted a plaintiff’s allegation of an increased risk of harm from the implantation of a medical device that required current medical monitoring to constitute injury in fact. The court accepted plaintiff’s analogy to cases where plaintiffs have been exposed to toxins (i.e., nuclear emissions or asbestos) and have an increased risk of disease. \textit{Id.} at 571. Courts in the Second Circuit have also permitted an allegation of increased risk of harm to satisfy the injury in fact requirement when a plaintiff claims an “increased future risk of identity theft,” see \textit{Caudle v. Towers, Perrin, Forster & Crosby, Inc.}, 580 F. Supp. 2d 273, 279 (S.D.N.Y. 2008), an increased risk of being assessed penalties because of reliance on fraudulent tax advice, see \textit{Denney v. Deutsche Bank AG}, 443 F.3d 253, 264-65 (2d Cir. 2006) (“An injury-in-fact may simply be the fear or anxiety of future harm. For example, exposure to toxic or harmful substances has been held sufficient to satisfy the Article III injury-in-fact requirement even without physical symptoms of injury caused by the exposure, and even though exposure alone may not provide sufficient grounds for a claim under state tort law.”), and an increased risk of injury based on the defendant’s failure to secure plaintiff prisoner’s wheelchair properly when he was being transported. \textit{Shariff v. Goord}, 04-CV-6621 CJS(F), 2006 U.S. Dist. LEXIS 49957, at *10, *20 (W.D.N.Y. July 20, 2006).

\textsuperscript{184} 464 F.3d 1 (D.C. Cir. 2006), \textit{vacating} 440 F.3d 476 (D.C. Cir. 2006).


\textsuperscript{186} \textit{Id.} at 484.
information regarding the risk to NRDC’s members from the EPA’s rule.\textsuperscript{187} In response to this new information, the D.C. Circuit found that NRDC did have standing.\textsuperscript{188} Based on the EPA’s own expert estimate, the court calculated that two to four of NRDC’s five hundred thousand members would develop cancer as a result of the rule—a risk the court considered sufficient to support standing.\textsuperscript{189}

In \textit{NRDC}, the court expressly did not decide whether it was appropriate to take a quantitative approach to determining whether an increased risk of injury constituted injury in fact.\textsuperscript{190} The court repeated its refusal to decide whether any increase in risk was enough for standing shortly thereafter in \textit{Virginia State Corp. Commission v. Federal Energy Regulatory Commission}, but cabinéd this open question to environmental disputes, stating that “[o]utside the realm of environmental disputes . . . we have suggested that a claim of increased risk or probability cannot suffice.”\textsuperscript{191}

The D.C. Circuit clarified its position on increased risk in the non-environmental context a year later, in 2007, in \textit{Public Citizen, Inc. v. National Highway Traffic Safety Administration}.\textsuperscript{192} In \textit{Public Citizen}, petitioners—including a citizens group, tire makers, and a tire industry association—challenged a federal motor vehicle safety standard requiring cars to contain tire pressure monitors that lit up when the tire pressure fell below a set standard.\textsuperscript{193} The court asked for supplemental briefing on whether the challenged standard “creates a substantial increase in the risk of death, physical injury, or property loss,” over the alternative interpretation, and whether the risk of harm, including the alleged increase, was substantial.\textsuperscript{194} The court noted that it had only allowed standing in increased risk of harm cases when both of these factors were present, noting that there were several reasons

\textsuperscript{187} \textit{Natural Res. Def. Council}, 464 F.3d at 3.
\textsuperscript{188} \textit{Id.} at 7.
\textsuperscript{189} \textit{Id.}
\textsuperscript{190} \textit{Id.} at 6-7. After deciding NRDC had standing, the court dismissed the case on the merits. \textit{Id.} at 11.
\textsuperscript{191} \textit{Va. State Corp. Comm’n v. Fed. Energy Regulatory Comm’n}, 468 F.3d 845, 848 (D.C. Cir. 2006). In this case the court found that it “need not face those issues here,” because petitioner, who had argued that its investors faced an increased risk of incorrectly evaluating the company’s financial health, made no showing adequate to explain their position. \textit{Id.}
\textsuperscript{193} \textit{Id.} at 1284.
\textsuperscript{194} \textit{Id.} at 1297.
why standing should not be allowed lightly in probabilistic cases: (1) allowing injury based on speculative injury would allow judicial review of any agency action because almost all agency action slightly increases risk or, according to citizen preference, insufficiently decreases risk; (2) speculative injury standing would eliminate the requirement that an injury be “actual or imminent” from the standing requirements; and (3) such cases would cause the judiciary to infringe on the Executive’s responsibility to “take care” that the laws were faithfully executed by expanding its role beyond the hearing of actual cases or controversies. Not surprisingly, the court found that Public Citizen did not meet its burden in its supplemental briefing and dismissed its claims.

The United States Supreme Court appeared to accept probabilistic harm, characterized as increased risk of injury, as support for standing in the environmental context in *Massachusetts v. EPA*. Although the Court noted that “rising seas have already begun to swallow Massachusetts' coastal land,” it emphasized that

> [t]he severity of that injury will only increase over the course of the next century: If sea levels continue to rise as predicted, one Massachusetts official believes that a significant fraction of coastal property will be “either permanently lost through inundation or temporarily lost through periodic storm surge and flooding events.” Remediation costs alone, petitioners allege, could run well into the hundreds of millions of dollars.

---

195 Id. at 1295.


The Court’s characterization of Massachusetts’ standing relies heavily, if not exclusively, on these allegations of future injury.\textsuperscript{199}

Increased risk of harm as a basis for standing is a critical tool for food-safety-impact litigants, and whether their claims are successful may depend on whether the courts they are before permit increased risk of harm claims in cases outside of environmental disputes, and whether the court views food-safety issues as either a subset of or analogous to environmental disputes. The possibility of reaching the merits of the case in such litigation will also depend on whether the court applies a quantitative assessment to increased risk, or assumes that any increased risk is adequate.\textsuperscript{200}

2. The Necessity of Third-Party Action May Thwart Redressability

In \textit{Levine v. Vilsack}, another food-safety-impact case, the court found that standing failed on redressability grounds rather than injury-in-fact grounds. Here, the district court accepted plaintiffs’ allegations of increased injury of harm as a basis for Article III standing, but the Court of Appeals found that plaintiffs did not have standing to challenge a USDA Notice because even if the plaintiffs prevailed in court, the requested relief was only available through a series of speculative steps and the actions of third-parties.\textsuperscript{201} For this reason, plaintiffs failed to satisfy the redressability prong of the standing inquiry.

In 2005, several individual plaintiffs and several associational plaintiffs (the “Levine Plaintiffs”) brought suit in federal district court, challenging a USDA Notice (“Notice”) issued earlier that year, which stated that the slaughter of poultry is not governed by any federal standard. The Levine Plaintiffs alleged that the Notice was contrary to law, specifically to the APA, and to the Humane Methods of Slaughter Act (HMSA), a 1958 statute that provided that “cattle, calves, horses, mules, sheep, swine, and other livestock”

\textsuperscript{199} In this regard, see Justice Roberts’ dissent: “[A]cepting a century-long time horizon and a series of compounded estimates renders requirements of imminence and immediacy utterly toothless. . . . ’Allegations of possible future injury do not satisfy the requirements of Art. III. A threatened injury must be certainly impending to constitute injury in fact.” \textit{Id.} at 542 (Roberts, C.J., dissenting) (citations omitted).

\textsuperscript{200} Once again, as mentioned \textit{supra} in note 176, it also remains to be seen how courts negotiate the plausibility standard of \textit{Ashcroft v. Iqbal}, 556 U.S. 662 (2009), and \textit{Bell Atlantic Corp. v. Twombly}, 550 U.S. 544 (2007).

\textsuperscript{201} Levine v. Johanns, 587 F.3d 986, 993-95 (9th Cir. 2009).
must be humanely slaughtered.\textsuperscript{202} In 1978, parts of the HMSA were incorporated into the Federal Meat Inspection Act (FMIA), which had the effect of prohibiting federal inspection of meat that had not been slaughtered in compliance with the humane slaughter methods dictated by the HMSA.\textsuperscript{203} Meat that is not federally inspected cannot enter the marketplace. The 1978 version of the HMSA retained the section of the 1958 HMSA that listed the types of animals that must be humanely slaughtered, which included “other livestock.”\textsuperscript{204} The Levine Plaintiffs argued that because the 1958 HMSA was still in force, the Notice—which said that no federal standard applied to poultry—was construing “other livestock” to exclude poultry, which was an arbitrary and capricious interpretation. Because of the Notice, argued the Levine plaintiffs, poultry was being slaughtered inhumanely.\textsuperscript{205}

The associational plaintiffs were nonprofit organizations that worked to prevent cruelty to animals, and they brought suit challenging the Notice on behalf of their members, who were also listed as individual plaintiffs, and who were characterized as “regular consumers of poultry meat.” The Levine Plaintiffs alleged that inhumane methods of slaughter increased the possibility that the poultry would be contaminated by bacteria, thereby increasing their risk of illness each time they ate inhumanely slaughtered poultry.\textsuperscript{206}

\textsuperscript{205} Levine, 2006 U.S. Dist. LEXIS 63667, at *4-5.
\textsuperscript{206} Id. at *5. These plaintiffs were joined by several workers in poultry processing plants and two organizations that represented workers. They alleged physical and emotional injuries from working in plants where poultry is slaughtered inhumanely. Id. at *6-7.

After this case was filed, the court related it with another case challenging the USDA’s failure to apply humane slaughter requirements to bison and reindeer. Plaintiffs in this case alleged that they regularly ate bison and reindeer meat, and were therefore at increased risk of contracting food poisoning whenever they ate meat that had been inhumanely slaughtered. Id. at *9. The court dismissed the Bison plaintiffs’ complaint with prejudice, concluding that the USDA did not have a non-discretionary duty to apply the humane slaughter requirements to bison and reindeer, and the APA challenge therefore failed. Id. at *46-50. Section 706(1) of the APA provides for a court to “compel agency action unlawfully withheld or unreasonably delayed,” 5 U.S.C. § 706(1) (2006), and to bring an action under this section, a plaintiff must show that “an agency failed to take a discrete agency action that it is required to take.” Levine, 2006 U.S. Dist. LEXIS 63667, at *46 (quoting Norton v. S. Utah Wilderness Alliance, 542 U.S. 55, 64 (2004)).
The district court found that the individual plaintiffs established the requirements of Article III standing: injury-in-fact, traceability, and redressability.\textsuperscript{207} The allegation that certain individual plaintiffs were at an increased risk for illness was neither too generalized, nor too speculative to constitute injury-in-fact.\textsuperscript{208} As long as a harm “is separate from an interest in having the government abide by the law,” it “may be concrete even though it is widely shared,” explained the court.\textsuperscript{209} Moreover, the court found there to be a “credible” threat that the plaintiffs would suffer concrete harm in the future, which it found sufficient to satisfy the imminence prong of the injury-in-fact inquiry.\textsuperscript{210} Analogizing to Baur, the court found the plaintiffs’ claims credible because they relied on the USDA’s own studies showing that bacterial contamination was more likely when inhumane slaughtering methods were used.\textsuperscript{211}

The court dismissed the claims of the Levine associational plaintiffs with leave to amend because they had failed to satisfy the requirements for associational standing. For an organization to have standing to sue on behalf of its members, it must show that: (1) “its members would otherwise have standing to sue in their own right,” (2) “the interests at stake are germane to the organization’s purpose,” and (3) “neither the claim asserted nor the relief requested requires the participation of individual members’ in the lawsuit.”\textsuperscript{212} The organizations in the lawsuit had asserted the interest of protecting their members’ health, although the actual main purpose of the organization was to prevent animal cruelty.\textsuperscript{213}

\textsuperscript{207} Although the court dismissed the Bison plaintiffs’ complaint for lack of jurisdiction under the APA, it held that all of the plaintiffs, including the animals, had Article III standing, but that the animals lacked statutory standing as the APA only applied to “person[s].” Levine, 2006 U.S. Dist. LEXIS 63667, at *45.

\textsuperscript{208} Id. at *14-31. John Does I and II were found to have Article III standing based on their allegations of physical and emotional injuries. Id. at *32.

\textsuperscript{209} Id. at *15 (citing FEC v. Akins, 524 U.S. 11, 24-25 (1998)).

\textsuperscript{210} Id. at *29.

\textsuperscript{211} Id. at *20-21.


\textsuperscript{213} One of the organizational plaintiffs was permitted to continue because it was dedicated partly to consumer protection and human health. Levine, 2006 U.S. Dist. LEXIS 63667, at *37-39. The court also found the challenged Notice to constitute a final agency action which was subject to judicial review. The Levine plaintiffs challenged the Notice under APA section 706(2), which provides that a court can “hold unlawful and set aside agency action, findings and conclusions” that are “arbitrary” or “capricious.” 5 U.S.C. § 706(2)(a) (2006). To be set aside under this section, an agency action must be discrete and final, and to be final, an action must “(1) ‘mark the consummation of the agency’s decision making process’ and not be tentative, and (2) have legal consequences.” Levine, 2006 U.S. Dist. LEXIS 63667, at *52 (citing
Subsequently, the district court granted summary judgment to defendants, ruling that "Congress intended to exclude poultry from the categorical word 'livestock.'"214 Plaintiffs appealed, and the Ninth Circuit vacated the district court’s decision, remanding the case for the district court to dismiss based on the plaintiffs’ lack of Article III standing. The Ninth Circuit did not question the district court’s injury-in-fact analysis,215 but found instead that none of the plaintiffs could show that their alleged injury could be redressed by any ruling of the court. Because the HMSA had no enforcement mechanism, any decision of the court would have to be followed by a series of steps to reach the plaintiffs’ desired result of the use of more humane poultry slaughter methods, all of which steps were speculative. If the court ruled that the Notice was contrary to law and poultry should be included as “other livestock,” the Secretary would have to determine that poultry should fall under the FMIA’s umbrella, and then issue regulations for the humane slaughter of poultry. Furthermore, private processors would then have to follow these regulations. Because of the speculative nature of each of these steps, the court found the likelihood of relief to be too low to satisfy the redressability prong of Article III.216

Bennett v. Spear, 520 U.S. 154, 177-78 (1997)). The court determined that the Notice constituted final agency action. Id. at *61.

The court also ordered that the Levine plaintiffs show cause as to why the claims of the workers, EJC, and WNCWC should not be dismissed for improper venue. Id. at *65.

Levine v. Conner, 540 F. Supp. 2d 1113, 1121 (N.D. Cal. 2008). It appears that the Levine individual plaintiffs—characterized by the court as “poultry eaters concerned about food-borne illness”—and the organizations representing the workers were the only plaintiffs left. Id. at 1113.

The absence of a discussion regarding injury-in-fact most likely shows that the Ninth Circuit accepts that a “credible” increase in risk suffices as injury-in-fact. In this regard, see also Central Delta Water Agency v. United States, 306 F.3d 938, 947 (9th Cir. 2002) (“[T]he possibility of future injury may be sufficient to confer standing on plaintiffs; threatened injury constitutes ‘injury in fact.’”).

Levine v. Vilsack, 587 F.3d 986, 988, 993-95, 997 (9th Cir. 2009). The redressability point was not as clear cut as the Ninth Circuit represented. The court concluded that the chain of events that would have to take place to remedy plaintiffs’ injury (the increased risk of foodborne illness) was too speculative to satisfy the Article III standing requirements, supposing that if the court construed “other livestock” to include poultry, the Secretary of Agriculture would still need to enforce the humane slaughter mechanisms in the FMIA and write regulations to do so. The court also noted that poultry processors would then have to adhere to the regulations. However, if one instead assumes that the Secretary will follow the legislative mechanism, then an inclusion of poultry in the HMSA humane slaughter mandate would lead inevitably to poultry’s inclusion in the FMIA, and the writing of regulations to govern its humane slaughter. Moreover, the inclusion of poultry in the definition of “other livestock” would surely relieve the injury to plaintiffs, which was the failure to include poultry in this definition, even if its ultimate effectiveness in ensuring humane slaughter requirements was delayed. See, e.g.,
In *Levine*, the associational plaintiffs failed to satisfy the requirements for associational standing because the interest that they asserted was not germane to the organization’s purpose. However, had they asserted an interest in animal welfare, they may have failed to satisfy the prudential zone-of-interests standing test, as did the plaintiffs in the next case.

3. The Zone-of-Interests Standing Requirement

The zone-of-interests standing requirement appears to be a sticking point for the existing food-safety impact cases. This is only, however, because a significant portion of these cases thus far have been instigated by organizations such as animal welfare organizations or trade associations which have as their main purpose something other than food safety or consumer protection. The interests germane to these organizations’ purposes are not protected by the food-safety statutes. In *Baur*, Farm Sanctuary, with whom Michael Baur originally brought suit, was dismissed from the case because as an animal welfare organization, it could not assert an interest protected by the Federal Meat Inspection Act or the Federal Food Drug and Cosmetic Act.\(^{217}\) The Levine associational plaintiffs were dismissed because the interest they asserted in the lawsuit, the protection of their members’ health—which presumably they asserted because it fell under the zone of

---

\(^{217}\) *Farm Sanctuary, Inc. v. Veneman, 212 F. Supp. 2d 280, 285 (S.D.N.Y. 2000).*

---
interests of the Federal Meat Inspection Act—was not germane to the purposes of their organization.\textsuperscript{218}

And, in 2005, the Ninth Circuit dismissed a claim by the Ranchers Cattlemen Action Legal Fund United Stockgrowers of America (R-CALF), a nonprofit organization representing United States cattle producers “on domestic and international trade and marketing issues [which] . . . is dedicated to ensuring the continued profitability and viability of the U.S. cattle industry”\textsuperscript{219} under NEPA for a failure to satisfy the zone-of-interests prudential standing requirement. R-CALF had sued the USDA seeking to overturn the USDA’s decision to lift a ban on Canadian imports of most bovine meat for human consumption.\textsuperscript{220}

Until January 4, 2005, the USDA had banned the importation of all ruminants and ruminant products from countries in which BSE had been found, including Canada.\textsuperscript{221} On that date, the USDA published a Final Rule relaxing the ban on the importation of Canadian ruminants. R-CALF sued to block the implementation of the Rule, arguing that it was arbitrary and capricious under the APA; that it violated NEPA because the agency had failed to make its environmental assessment public before publishing the Rule and had failed to prepare an Environmental Impact Statement; and that it violated the Regulatory Flexibility Act (RFA) by failing to assess whether the Rule’s impact on small businesses could have been mitigated.\textsuperscript{222}

The district court agreed with plaintiffs on all counts, writing that the USDA had “preconceived intention, based upon inappropriate considerations, to rush to reopen the border regardless of uncertainties in the agency’s knowledge,” and had “attempted to work backwards to support and justify this


\textsuperscript{219} About R-Calf USA; Working for the U.S. Cattle Industry, R-CALF USA, http://www.r-calfusa.com/about/about.htm (last updated Feb. 11, 2011).

\textsuperscript{220} Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dep’t of Agric., 415 F.3d 1078, 1084 (9th Cir. 2005).

\textsuperscript{221} As the Ninth Circuit explained, “[r]uminants are hoofed mammals generally defined by their four-chambered stomachs and their practice of chewing a cud consisting of regurgitated, partially digested food,” and “include cattle, sheep, goats, deer, giraffes, camels, llamas, and okapi.” Id. at 1084 n.1.

\textsuperscript{222} The RFA dates from 1980, and was passed “to ‘encourage administrative agencies to consider the potential impact of nascent federal regulations on small businesses.’” Id. at 1100 (quoting Assoc. Fisheries of Maine, Inc. v. Daley, 127 F.3d 104, 111 (1st Cir. 1997)).
goal,” but the Ninth Circuit reversed. First, the court found that the district court had improperly substituted its judgment for that of the agency and had failed to defer properly to the agency’s determinations. The court found the USDA’s determination that the risks inherent in the Rule were both small and acceptable to be supported by an adequate administrative record. As to the RFA, the Ninth Circuit held that the USDA met RFA’s purely procedural requirements, by “conduct[ing] a detailed economic assessment of its proposed rule on small businesses.”

Regarding NEPA, the Ninth Circuit held that R-CALF did not have standing to pursue its claim under this statute. NEPA, which prescribes the steps an agency must take before taking an action that will affect the environment, contains no private right of action or citizen-suit provision. A plaintiff suing for a NEPA violation must bring suit under the APA and must fall under the zone of interests protected by NEPA. Under Ninth Circuit law, a party seeking to sue for a NEPA violation must assert an environmental injury. R-CALF, however, exists to protect the economic interests of its members, and the injuries it asserted in its complaint were economic. Economic interests are not protected by NEPA, and R-CALF therefore lacked standing to bring its NEPA challenge. The Ninth Circuit remanded the case, and the district court granted summary judgment to the USDA.

4. Conclusion

These cases, both in the abstract and as tools for future food-safety-impact litigation, teach three lessons. First, the

---

224 Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am., 415 F.3d at 1100.
225 Id. at 1101.
226 Id. at 1103-04.
227 Id. at 1103 (citing Stratford v. FAA, 285 F.3d 84, 88 (D.C. Cir. 2002)).
228 R-CALF also asserted that “R-CALF USA members will also be adversely affected by the increased risk of disease they face when Canadian beef enters the U.S. meat supply.” Id. (internal quotation marks omitted).
229 Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dep’t of Agric., 359 F. Supp. 2d 1058, 1062 (D. Mont. 2005). The Ninth Circuit affirmed, reminding R-CALF that it could not use post-decision evidence to show that the USDA had “re[l]ied on faulty assumptions,” but that it could use this new evidence to petition the USDA to reopen rulemaking under the APA. Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dep’t of Agric., 499 F.3d 1108, 1111, 1117-18 (9th Cir. 2007).
standing challenges are navigable. At least the Ninth Circuit and the Second Circuit have shown themselves amenable to the allegation of an increased risk of contracting foodborne illness as a basis for injury in fact. Such a claim matches the purpose of the statutes that protect the safety of our food supply.

Second, these cases illustrate the absence of a food-safety-impact litigation culture. Two of these three cases were brought by animal welfare organizations, and one by a trade organization. Not one of the organizations spearheading this litigation had as its primary purpose the prevention, or minimization of foodborne illness, although they all alleged the increased risk of such as a basis for standing. Food-safety litigants, especially organizational litigants, must better negotiate the zone-of-interests test.

And finally, in two of the three cases, the court found the plaintiffs’ use of the government’s own studies to show that the agency action would result in the alleged harm to be strong support for standing. The agency’s refusal to act in the face of its own studies lends support to the possibility that the agency’s final action was arbitrary and capricious, and allows a court to remain deferential to the agency, while ruling against it.

C. Challenges to Judicial Review of Agency Action

As described above, food-safety-impact litigation must be brought under the APA because the major food-safety statutes have neither citizen-suit provisions nor provide for private rights of action. Each case discussed above in the section on standing involved a challenge to agency action under the APA, for these very reasons, and in two, defendants challenged whether the agency action was reviewable under the APA. In Levine, the defendants argued that the plaintiffs had not challenged a final agency action, as is necessary for review under the APA, but the court disagreed.230 And in the R-CALF case, the Ninth Circuit ultimately determined that the district court had erred under the APA in its failure to defer to the agency’s expertise, and that the agency action was not, therefore, arbitrary and capricious.231

231 Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am., 415 F.3d at 1100.
Another common challenge to the judicial reviewability of agency action in the food-safety context is the claim that the agency action at issue has been committed to agency discretion by law, and is therefore unreviewable under section 701(a)(2) of the APA. There are several categories of actions that the Supreme Court have found to fall under section 701(a)(2), and thus to be unreviewable, including “an agency’s decision not to institute enforcement proceedings,” “an agency’s refusal to grant reconsideration of an action because of material error,” and “the allocation of funds from a lump-sum appropriation.” In the following two cases, defendants claimed that the challenged agency action was an enforcement decision, and therefore unreviewable under the APA. In each of these cases, the court disagreed. Furthermore, both of these cases were actually successful in achieving their requested relief.

Public Citizen v. Heckler was the impetus behind the federal government’s ban on the interstate sale of raw milk. Public Citizen, a citizen advocacy organization, filed a citizen petition with the FDA in April of 1984, requesting that the agency prohibit the sale of unpasteurized milk. After no ruling was made, and the FDA refused to provide a schedule for a ruling, Public Citizen filed suit in September 1984 seeking a response to its petition. HHS held an informal hearing in October 1984 to solicit information on whether raw milk was a

---

232 “Common” insofar as this challenge was made in two of the fourteen cases found.
235 Pub. Citizen v. Heckler, 653 F. Supp. 1229 (D.D.C. 1987). Raw milk is milk that has not been pasteurized or homogenized. Before 1987, the federal government did not regulate the sale of raw milk, although many local and state governments did in some manner. The city of Chicago passed a mandatory milk pasteurization law in 1908, and in 1947, Michigan was the first state to do so.

The FDA collected information from 1974 to 1982, and in 1982, wrote a proposed regulation banning the interstate sale of all raw milk, based on the evidence that the consumption of raw milk was linked to bacterial disease. High-level officials at HHS and the CDC supported this regulation, and statistical support was provided by the Chief of the Bureau of Foods Epidemiology and Clinical Toxicology Division in 1984. Id. at 1232-33.
public health concern and, if so, whether requiring pasteurization was the best solution. The evidence collected by the FDA, and the evidence introduced at the informal hearing “conclusively show[ed] . . . [that] raw milk is unsafe.”

After the district court ruled that the Department of Health and Human Services’ “justifications for delay were ‘lame at best and irresponsible at worst,’” and ordered the Department to respond, the Department responded by denying Public Citizen’s petition for several reasons, including the following: (1) most raw milk was sold intrastate, (2) illnesses from raw milk stemmed mainly from intrastate commerce, and (3) banning interstate sales of raw milk would therefore have little effect on the public health. Public Citizen returned to district court to challenge the rule as arbitrary and capricious.

Defendant HHS challenged the reviewability of the petition’s denial, claiming that it was an enforcement decision, and thus fell under the section 701(a)(2) exception to reviewability. The court rejected this contention, explaining that “[h]ere the action at issue is not an individual enforcement action, but an agency’s refusal to engage in rulemaking.” Furthermore, in *Heckler v. Chaney*, the case establishing the discretionary enforcement exception, the agency chose not to take an enforcement action against an entity, and there were no clear statutory guidelines for the court to interpret on when such actions should be taken. Here, on the other hand, HHS’s action could be reviewed with the clear statutory guidelines of the FFDCA and the Public Health Act as a guide.

The court then determined that the denial of Public Citizen’s petition was arbitrary and capricious. It found the explanation offered to be inconsistent with the evidence in front of the agency; that the documents before the court showed

---

238 Id. at 1234-35 (citing Pub. Citizen v. Heckler, 602 F. Supp. 611, 613 (D.D.C. 1985)).
239 Id. at 1235. The FDA’s other reasons were that milk sold interstate did not pose a greater risk than milk sold intrastate, that the FDA did not have the authority to prohibit intrastate sales, and that the problem was better dealt with by state and local governments anyway. Id.
240 Id. at 1231.
241 Id. at 1236 (citing Heckler v. Chaney, 470 U.S. 821, 825 n.2 (1985)). This issue was settled in *Massachusetts v. EPA*, 549 U.S. 497, 527 (2007) (“Refusals to promulgate rules are thus susceptible to judicial review, though such review is ‘extremely limited’ and ‘highly deferential’” (quoting Nat’l Customs Brokers & Forwarders Ass’n of Am., Inc. v. United States, 853 F.2d 93, 96 (D.C. Cir. 1989))).
242 Public Citizen, 602 F. Supp. at 1236.
that high level officials at the FDA and the CDC thought a ban on raw milk sales was a good idea, and indicated “a lack of rationality on the part of HHS in the decisionmaking process”; and that the reason given for the decision had “no rational connection to the undisputed facts in the record.” Pursuant to the court’s order, the FDA published a Final Rule banning the interstate sale of all raw milk and all raw milk products.

In *Kenney v. Glickman*, a number of individual plaintiffs, calling themselves “poultry consumers,” sued the USDA for discrepancies in the way that the USDA regulated poultry and meat. Plaintiffs argued that the USDA should either issue the same regulations for poultry and meat, or provide a “legally sufficient reason for treating meat and poultry differently.”

The Eighth Circuit rejected the USDA’s contention that the discrepancies between the meat and poultry inspection standards reflected an enforcement decision—that the agency had merely chosen to use agency resources to enforce meat inspections more rigorously. The standards at issue involved neither a decision about whether there had been a violation nor a refusal to institute proceeding, but were, instead, general

---

243 Id. at 1237, 1241.

244 21 C.F.R. § 1240.61 (2010); Memoranda from Milk Safety Branch on Sale/Consumption of Raw Milk to All Regional Food and Drug Directors (Mar. 19, 2003), available at http://www.fda.gov/Food/FoodSafety/Product-SpecificInformation/MilkSafety/CodedMemoranda/MemorandaofInformation/ucm079103.htm.

245 *Kenney v. Glickman*, 96 F.3d 1118, 1118 (8th Cir. 1996). The original plaintiffs included red meat producers, who brought the case to remedy a perceived inequity in the USDA’s treatment of poultry and of beef, but the district court dismissed these plaintiffs for lack of standing. See Joe Roybal, *Fighting for Fairness*, BEEF MAG. (Nov. 1, 2000, 1:00 PM), http://beefmagazine.com/mag/beef_fighting_fairness/. The red meat producers did not appeal. See id.

Both the PPIA, which regulates poultry products, and the FMIA, which regulates meat, require that carcasses be inspected for the presence of certain contaminants that may cause the carcasses to be termed “adulterated,” and hence not allowed into the food supply. *Kenney*, 96 F.3d at 1121. Individual carcasses are inspected, and any contaminants are removed. *Id.* No contaminants are allowed to remain; there is a “zero tolerance” policy as to these contaminants on individual carcasses. *Id.* The contaminated parts must be removed from meat, while they may be washed off of poultry. *Id.* After the individual carcasses are inspected, an inspector then inspects sample carcasses from a particular lot to determine if there may be any process defects on that lot. *Id.* Until 1993, both the PPIA and the FMIA allowed a tolerance level of slightly more than zero for process defects, but in March 1993, the USDA lowered the tolerance level to zero for meat, but not for poultry. *Id.*

246 Id.
The case did not, therefore, fall into the *Heckler v. Chaney* category of presumptively unreviewable cases. The Court of Appeals also found that the prohibition against allowing adulterated products to enter the food supply provided a “sufficient standard” for the court to evaluate whether the USDA’s policies made sense. In addition, the court looked to the legislative history of the PPIA and the FMIA to determine that Congress intended for the two to be construed consistently. For this reason, there was sufficient law to apply to determine whether the USDA acted arbitrarily and capriciously in implementing differing inspection standards for meat and poultry.

After remand, the district court found the discrepancy between meat and poultry regulation to be arbitrary and capricious, and in direct response to this decision—indeed, noting the decision in the background to the Final Rule—the FSIS harmonized the regulations.

*Kenney* and *Public Citizen* have three main implications. First, *Heckler v. Chaney*’s presumption of unreviewability does not apply to arbitrary and capricious challenges to the refusal to promulgate rules, which, in any event, was settled definitively by *Massachusetts v. EPA*. Second, if a petitioner presents evidence to support its allegations, the statutes governing food safety (including the FMIA, the PPIA, and the FFDCA) provide sufficient guidelines for a court to determine whether the relevant agency has acted reasonably. Third, a court is more likely to permit a case to go forward if petitioner’s evidence was produced by the relevant agency.

---

248 Id. at 1123.

249 Id. One judge dissented in part, finding that the USDA’s regulations regarding tolerance standards for process defects and regarding the methods used to cleanse contaminants were enforcement actions, and were therefore presumptively unreviewable under *Chaney*. Id. at 1126 (McMillian, J., dissenting in part).

250 Id. at 1124 (majority opinion).

251 Id.

252 Roybal, *supra* note 245. The district court decision is unavailable.


254 *Massachusetts v. EPA*, 549 U.S. 497, 527 (2007) (“Refusals to promulgate rules are thus susceptible to judicial review, though such review is ‘extremely limited’ and ‘highly deferential’” (quoting Nat’l Customs Brokers & Forwarders Ass’n of Am., Inc. v. United States, 883 F.2d 93, 96 (D.C. Cir. 1989))).

255 This has been a theme throughout the food-safety cases. The court relied on the government’s own documents to find standing in *Baur*, and in *Levine* as well, although that decision was overturned on other grounds.
D. Paths to Success in Food-Safety-Impact Litigation

The upshot of the food-safety cases described above is that suits brought by food consumers or public health organizations alleging an increased risk of contracting foodborne illness because of an established governmental policy, and relying on the agency’s own documents, are likely to get into court. Specifically, these cases have four main implications for justiciability.

First, courts seem to generally accept that the increase in risk of foodborne illness satisfies the Article III injury in fact standing requirement. This is noted with several caveats. One is that not every court is likely to accept such a claim. As discussed above, the D.C. Circuit would most likely look for a quantifiable increase in risk, and may even prohibit this category of claimant from litigating outside of the environmental context. The second caveat is that the cases discussed above were, for the most part, decided before the Court’s decisions in *Ashcroft v. Iqbal* and *Bell Atlantic Corp. v. Twombly*, which instructed district courts that a plaintiff needed factual allegations that “raise a right to relief above the speculative level,” and must “state a claim to relief that is plausible on its face.” This standard may make it more difficult for impact plaintiffs to progress beyond the motion-to-dismiss stage, although many will have been through the administrative petitioning process before filing a complaint, thus allowing them to gather information and evidence towards their complaint. Those plaintiffs who can accumulate more information before filing a complaint will fare better.

The second lesson learned from the above discussion of justiciability relates to a plaintiff’s use of the relevant agency’s own documents and evidence to show that the potential harm from the agency action was significant. The *Levine, Baur, Kenney*, and *Public Citizen* courts found the fact that the agency’s own documents supported plaintiffs’ contentions (i.e., inhumanly slaughtered poultry are more likely to be carriers of foodborne illness)
communicable illness; downed cattle are more likely to have mad cow disease; the PPIA and the FMIA were meant to be construed similarly; FDA officials themselves acknowledged the danger of drinking raw milk) to be a compelling factor in allowing the case into court, and in the case of Public Citizen, to rule in favor of the plaintiff.

Is this not a question for the merits? Yes, such documents ultimately go to whether the plaintiff has managed to show that the agency's action was arbitrary and capricious. But the presence of such documents indicates to the court that the suit is neither futile nor frivolous; it speaks to an implicit likelihood of success inquiry. A court is required to defer to an agency's reasonable construction of its regulations, and may not substitute its judgment for that of the agency. This is even more important when a high level of technical expertise is implicated. Because judicial deference to agency decision making is so strong, a court must see a role for itself in the dispute, and a way to remain within its institutional competence before it allows plaintiffs into court. Documentary evidence of the relevant harm indicates that, whether through corruption, inefficiency, or sheer irrationality, the agency has acted against its own evidence and, in certain cases, its own directives. Moreover, such documents show that the agency was not acting pursuant to internal management considerations or other factors that a court should have no hand in administering.

The use of an agency's own documents by petitioners does not, however, guarantee success on the merits. For example, in the *R-CALF* case, the Ninth Circuit reversed the district court's grant of a preliminary injunction to plaintiffs, Plaintiff U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 843 (1984). Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am. v. Dep't of Agric., No. 05-35264, 2005 U.S. App. LEXIS 17360, at *32 (9th Cir. Aug. 17, 2005). Similarly, in *Simpson v. Young*, 854 F.2d 1429 (D.C. Cir. 1988), the D.C. Circuit dismissed a citizen-suit challenge to the FDA's decision to list a color additive, Blue No. 2 dye, as safe for human consumption. Public Citizen, the Center for Science in the Public Interest, and a private citizen alleged that the FDA's studies on the dye had been improperly done. The court explained that it was compelled to "uphold the FDA's decision if it reveals that significant evidence on both sides of the question has been considered and that the agency has explained its conclusions in light of significant objections." Id. at 1434. Deference to the agency's judgment was particularly important in cases involving "sophisticated scientific judgment," as was this one. Id. And in *Public Citizen v. Foreman*, 631 F.2d 969, 977 (D.C. Cir. 1980), discussed supra note 162, the D.C. Circuit dismissed Public Citizen's challenge to the FDA's determination that nitrites were sanctioned as a preservative prior to 1958, and therefore qualified for an exemption from the FFDCA, finding that it must defer to the agency's technical expertise.

Kenney v. Glickman, 96 F.3d 1118, 1122-23 (8th Cir. 1996).
explaining that the district court had failed to sufficiently defer to the agency’s interpretation of the statute.\textsuperscript{262} The appeals court noted that the district court had erred by “analyzing each protective component of the regulatory system in isolation,” instead of “evaluat[ing] the cumulative effects of the multiple, interlocking safeguards.”\textsuperscript{263}

Third, an organizational plaintiff must be able to show that the interests at stake in the lawsuit are germane to the organization’s purpose and that the interests asserted fall under the zone of interests protected by the statute. This was an issue with the animal welfare organizations in Levine, which could not show that an interest in consumer health, as put forward in this lawsuit, was germane to their purposes;\textsuperscript{264} with R-CALF, which could not show environmental injury so as to have standing under NEPA;\textsuperscript{265} in Baur, when Farm Sanctuary was dismissed from the case;\textsuperscript{266} and in Kenney, where the red-meat producers did not survive a motion to dismiss.\textsuperscript{267}

This barrier is less likely to stand in the way of environmental plaintiffs. In the first place, the citizen-suit provisions arguably negate the zone-of-interests test altogether.\textsuperscript{266} Moreover, as discussed earlier, there is a well-organized environmental impact plaintiff movement. Nothing similar exists in the food-safety world. The inclusion of citizen-suit provisions in the environmental laws was partially a function of the existence of the environmental protection movement. Legislators, with the help of individuals from these organizations, recognized a beneficial symbiosis between the fledgling EPA and the efforts of these organizations to enforce and strengthen regulation. In a sense, several of these organizations, through a push and pull, became extra eyes, ears, and arms of the government in enforcing environmental protection.\textsuperscript{269}

Because there are fewer litigating public health organizations, food-safety litigation may be brought ostensibly for public health, but actually for other purposes, such as the

\begin{footnotes}
\item[263] Id. at *37.
\item[265] \textit{Ranchers Cattlemen Action Legal Fund United Stockgrowers of Am.,} 415 F.3d at 1100.
\item[266] Farm Sanctuary, Inc. v. Veneman, 221 F. Supp. 2d 280, 284-85 (S.D.N.Y. 2002).
\item[267] See Roybal, supra note 245.
\item[268] See supra note 176.
\item[269] ADAMS \& ET AL., supra note 77, at 27; SALE, supra note 96, at 34-35.
\end{footnotes}
humane treatment of animals, or the economic interests of cattle producers—hence the associational standing and zone-of-interests problem commonly faced by food-safety-impact litigants. Of course, if an organization brings suit alongside a consumer or two, the consumer may have standing even if the organizational plaintiff does not. The suit can therefore go forward, and the stated goal of the organization in the suit may still be reached. Other goals, however, such as publicity for the issue, and for the organization, may not be forthcoming in such a suit. And in certain cases, increased public awareness is more important than achievement of the suit's sought relief. Moreover, it is publicity that raises the profile of the organization bringing the suit, lends legitimacy to its enterprise, and teaches the public that this is an issue worthy of donating money.

Fourth, and finally, courts are more likely to grant access if the stated injury stems from a present governmental policy—for example, the policy of allowing “downed” cattle to enter the food supply. This element speaks to the injury-in-fact prong of the Article III standing analysis as well as the redressability prong—no third party has to act for the injury to take place, nor would a third party need to act for the requested relief to take place.

CONCLUSION

Food-safety-impact cases are few and far between, but there is no compelling reason for this to remain the case. The absence of citizen-suit provisions in the food-safety statutes does not foreclose citizen suits. Food-safety-impact litigation brought by individuals or groups able to show that they or their members are at increased risk of contracting foodborne illness as a result of a final agency action, ideally by pointing to evidence in the relevant agency’s own documents, is likely to make it past constitutional and prudential standing challenges. Moreover, it is entirely within the competence of the judiciary

---

270 Baur v. Veneman, 352 F.3d 625, 637 (2d Cir. 2003).
271 The Baur Court distinguished the harm alleged by plaintiff—increased risk of foodborne illness—from “alleged future injury [that] rested on the independent actions of third-parties not before the court.” Id. at 640. And the appeals court in Levine found the need for actions by third parties—the poultry producers—to make the possibility of redress for the plaintiffs too attenuated and consequently too speculative to satisfy the Article III requirement. Levine v. Vilsack, 587 F.3d 986, 994-95 (9th Cir. 2009).
to assess whether there has been arbitrary and capricious action taken in the food-safety context.

Overseeing food safety in this country is an enormous job, and agency oversight of food safety is, and will likely continue to be, severely underfunded. There are numerous food-safety areas where agency decision making has stalled, either from a lack of resources or from industry pressure. Such inertia is detrimental to the public health. Citizen litigation can act as a counterpart to governmental regulation, pushing agencies to fulfill their statutory mandates, and making judicial review of agency decision making regarding food-safety regulation a valuable tool in reducing the incidence of foodborne illness.
From Bad to Worse

ASSESSING THE LONG-TERM CONSEQUENCES OF FOUR CONTROVERSIAL FCC DECISIONS

Rob Frieden†

I. INTRODUCTION

Far too many major decisions of the Federal Communications Commission (FCC) rely on flawed assumptions about the current and future telecommunications marketplaces. When the FCC incorrectly overstates the current level of competition, it risks exacerbating its mistake going forward if actual competition proves unsustainable or lackluster. In many key decisions, the FCC cited robust competition in current and future markets as the basis for deregulatory decisions that relax restrictions on incumbents, abandon strategies for promoting competition, or apply statutory definitions of services that trigger limited government oversight. If the FCC has confidence in the

† Pioneers Chair and Professor of Telecommunications and Law, Pennsylvania State University; email: rmf5@psu.edu.

1 “[T]here is substantial competition in the provision of Internet access services.” In re AT&T Inc. & BellSouth Corp., Application for Transfer of Control, 22 FCC Rcd. 5662, 5724-25 (Mar. 26, 2007) (memorandum opinion and order). In 2008 the FCC stated that “advanced telecommunications capability is being deployed to all Americans in a reasonable and timely fashion.” In re Inquiry Concerning the Deployment of Advanced Telecommunications Capability to All Americans in a Reasonable and Timely Fashion, and Possible Steps to Accelerate Such Deployment Pursuant to Section 706 of the Telecommunications Act of 1996, 23 FCC Rcd. 9615, 9616 (June 12, 2008) (fifth report), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-88A1.pdf. On the other hand, at about the same time the FCC stated that “[s]tudy after study demonstrates that our nation’s broadband infrastructure lags dramatically behind other industrialized nations. In order to reverse this trend, we must encourage ‘third pipe’ technologies to provide some at least some [sic] challenge to the cable/telco broadband duopoly in our cities.” In re Implementation of the Commercial Spectrum Enhancement Act and Modernization of the Commission’s Competitive Bidding Rules and Procedures, 21 FCC Rcd. 6703, 6727 (June 2, 2006) (order on reconsideration of the second report and order).

2 An FCC conclusion that robust competition exists provides the basis for a reviewing court to affirm the Commission’s decision that it can deregulate. For example, the FCC abandoned rules requiring incumbent carriers to make available local switching and routing services to market entrants based on the determination
viability and permanence of competition, then reviewing courts will likely refrain from second-guessing the Commission and uphold its deregulatory initiative. In its zeal to announce deregulatory decisions and to accrue political dividends, the Commission ignores secondary and tertiary consequences of decisions that deprive it of the jurisdiction and flexibility needed to respond to technological and marketplace changes.¹

Ironically, the FCC has not promoted competition. It has exacerbated the trend toward concentration of ownership generated by technological convergence and the real (or perceived) need for incumbents to grow larger by acquiring competitors. Instead of making sure that this trend does not lead to oligopolistic behavior, which can harm consumers,³ the

---

¹ For example, the FCC has expressed confidence that it can assert its ancillary jurisdiction to achieve consumer protection even if it previously opted to streamline or eliminate regulatory safeguards.

We have a duty to ensure that consumer protection objectives in the Act are met as the industry shifts from narrowband to broadband services. Through this Notice, we thus seek to develop a framework for consumer protection in the broadband age—a framework that ensures that consumer protection needs are met by all providers of broadband Internet access service, regardless of the underlying technology. This framework necessarily will be built on our ancillary jurisdiction under Title I; as we explain in the Order, this jurisdiction is ample to accomplish the consumer protection goals we identify below, and we will not hesitate to exercise it.


“We emphasize that we will not hesitate to adopt any non-economic regulatory obligations that are necessary to ensure consumer protection and network security and reliability in this dynamically changing broadband era.” Id. at 14,915.

³ A duopoly controls the broadband Internet access marketplace in the United States.

Cable and DSL providers currently control almost 98 percent of the residential and small-business broadband market. More than one quarter of consumers have only one choice between cable and DSL, and even in markets with both services available, customers usually face a duopoly, with one choice for each type of service. Under any economic standard “nearly every regional broadband market is very highly concentrated.” The problem this situation generates is really very simple to grasp: in order to “reach” the logical and content layers, one has to “pass through” the physical layer; whoever controls the physical layer, unless restricted by law, becomes a gatekeeper for all other layers; and scarcity of physical layers means more control, and ability to realize that control, for fewer gatekeepers.
FCC has removed still-necessary regulatory safeguards designed to curb market power without robbing ventures of opportunities to operate efficiently. Intentionally or not, the FCC has contributed to market concentration even as it abandoned lawful techniques and policies to monitor and remedy likely marketplace abuses.

The FCC has embraced economic and political theory supporting reliance on marketplace forces without a complete empirical confirmation that industry self-regulation can occur. The Commission infers the existence of adequate competition and concludes that such competition will persist even though economic, technological, and future regulatory decisions might favor industry concentration and unsustainable competition.


It took the FCC over four years to detect and remedy over $52 million of deliberate data-service overcharges imposed by Verizon Wireless. *See In re Verizon Wireless Data Usage Charges, 25 FCC Rcd. 15,105 (Oct. 28, 2010) (order). Because the charges refer to Internet access, Verizon arguably could have claimed the FCC lacked jurisdiction to intervene, based on the assertion that all forms of Internet access constitute information services. See In re Inquiry Concerning High-Speed Access to the Internet over Cable and Other Facilities, 17 FCC Rcd. 4798, 4802 (Mar. 15, 2002) (declaratory ruling and notice of proposed rulemaking), aff’d, Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs., 545 U.S. 967 (2005). The FCC rejected the assertion by Verizon and other wireless carriers that it lacked jurisdiction to compel the carrier to provide data service to subscribers of other carriers.

Because encouraging data roaming serves the public interest by promoting connectivity for, and ubiquitous access to, mobile broadband as well as facilitating consumer access to wireless broadband data coverage nationwide, the obligations set forth above are reasonably ancillary to the Title III provisions to manage spectrum, allocate, assign, and to establish spectrum usage conditions in the public interest as set forth above.

favor large enterprises able to exploit economies of scale and scope. Technological and marketplace convergence supports the ability of large firms to offer bundles of services previously offered on a single, standalone basis. Additionally, the FCC’s willingness to conditionally approve mergers and acquisitions also leads to industry consolidation.

The FCC’s deregulatory decisions operate in one direction—the elimination of regulatory safeguards—typically without reserving any lawful and effective option to reassert safeguards should assumptions prove wrong or circumstances change. For example, the FCC’s decision to classify all Internet access technologies as information services—consequently

---

7 Economies of scale refers to the ability of a single firm to offer goods and services at the lowest cost by increasing its size. “[A]n increase in inputs leads to a proportionally greater increase in outputs (for example, a doubling of inputs would lead to more than a doubling of outputs).” Kevin G. Wilson, Deregulating Telecommunications and the Problem of Natural Monopoly: A Critique of Economics in Telecommunications Policy, 14 MEDIA CULTURE & SOC. 343, 345 (1992). “Declining levels of average cost accompanying greater expansion of product output and optimal use of plant and equipment. Cost advantages associated with the increasing size of firms.” MEDIA ECONOMICS THEORY AND PRACTICE, Glossary 286 (Alison Alexander et al. eds., 3d ed. 2004).

Scale economies refer to lower average costs from producing a larger quantity of output. A more technical definition is that economies of scale exist at a particular range of output when the long run average total cost decreases as output expands. Scale economies can be a barrier to entry if entrants are likely to acquire fewer customers and sell less output than the incumbent, and the resulting higher average cost for the entrants makes it difficult for them to compete with the incumbent, particularly if retail prices are close to the incumbent’s average cost.

Review of the Section 251 Unbundling Obligations of Incumbent Local Exchange Carriers, 18 FCC Rcd. 16,978, 17,028, n.245 (2003) (report and order and order on remand and further notice of proposed rulemaking) (citations omitted).

Economies of scope exist when one firm can produce two or more products at a lower total cost than if each product were produced separately by different firms. Scope economies can be a barrier to entry if entrants are unable to produce and sell all of the products the incumbent produces, and the resulting higher cost makes it unprofitable to enter the market.

Id. at 17,029, n.246.

8 The FCC has determined that various broadband technologies for accessing the Internet all qualify for limited regulatory oversight. See, e.g., In re Inquiry Concerning High-Speed Access to the Internet over Cable and their Facilities, 17 FCC Rcd. 4798, 4802 (Mar. 15, 2002) (declaratory ruling and notice of proposed rulemaking), aff’d, Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs., 545 U.S. 967 (2005); Wireline Broadband Classification Order, supra note 3, at 14,855.

9 An information service is defined as

the offering of a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing, or making available information via telecommunications, and includes electronic publishing, but does not include any use of any such capability for the management, control, or operation of a
reducing oversight of Internet Service Providers’ (ISPs) network management—now prevents the Commission from responding to complaints that some ISPs have interfered with subscribers’ Internet traffic. Should an ISP, such as Comcast, deliberately disrupt subscribers’ traffic, which can offer a competitive alternative to the company’s pay-per-view video-programming services, the FCC has no direct statutory authority to sanction the company for engaging in anticompetitive conduct. Worse yet, the decision to treat basic bit transmission as an information service severely restricts the Commission’s ability to impose safeguards on services that combine Internet access with software to provide the functional equivalent of a regulated service (e.g., Voice over the Internet Protocol (VoIP) and Internet Protocol Television (IPTV)). The

---

telecommunications system or the management of a telecommunications service.


12 Peer-to-peer applications, including those relying on BitTorrent, have become a competitive threat to cable operators such as Comcast because Internet users have the opportunity to view high-quality video with BitTorrent that they might otherwise watch (and pay for) on cable television. Such video distribution poses a particular competitive threat to Comcast’s video-on-demand (VOD) service. Comcast Sanction, supra note 11, at 13,030.


14 VoIP is the real-time carriage and delivery of data packets that correspond to voice. VoIP services range in quality, reliability, and price and can link both computers and ordinary telephone handsets. For technical background on how VoIP works, see Susan Spradley & Alan Stoddard, Tutorial on Technical Challenges Associated with the Evolution to VoIP, FCC (Sept. 22, 2003), http://www.fcc.gov/events/tutorial-technical-challenges-associated-evolution-voip. See generally Charles J. Cooper & Brian Stuart Koukoutchos, Federalism and the Telephone: The Case for Preemptive Federal Deregulation in the New World of Intermodal Competition, 6 J. TELECOMM. & HIGH TECH. L. 293 (2008).

15 IPTV offers consumers with broadband connections options to download video files or view (streaming) video content on an immediate “real time” basis. In re Sky Angel U.S., LLC, Emergency Petition for Temporary Standstill, DA 10-679, 25 FCC Rcd. 3879 (2010). Some of the available content duplicates what cable television subscribers receive therein triggering disputes over whether cable operators can secure exclusive distribution agreements and prevent an IPTV service provider from distributing the same content. “Sky Angel has been providing its subscribers with certain Discovery networks for approximately two and a half years, including the Discovery Channel, Animal Planet, Discovery Kids Channel, Planet Green, and the Military Channel. Sky Angel submits that these channels are a significant part of its service offering.” Id. at 3879-80. For background on IPTV, see In-Sung Yoo, The
FCC’s decision to apply the information-service classification to all Internet-access technologies means that the Commission has abandoned direct statutory authority to resolve problems and, in the future, must resort to questionable ancillary jurisdiction\(^{16}\) to resolve legitimate complaints and impose necessary regulatory safeguards.

There are other instances of unintended consequences resulting from decisions based on the FCC’s overly optimistic findings and assumptions about marketplace competition: removing caps on the total spectrum a single wireless carrier can control;\(^{17}\) abandoning local loop unbundling\(^{18}\) and other

---

\(^{16}\) Ancillary jurisdiction refers to an inference of statutory authority to impose rules and regulations based on indirect statutory authority. For example, the FCC asserted jurisdiction over cable television operators because the importation of distant broadcast television signals could have had an adverse financial impact on directly regulated television broadcasters. United States v. Sw. Cable Co., 392 U.S. 157, 178 (1968); see also FCC v. Midwest Video Corp. (Midwest Video II), 440 U.S. 689, 696-709 (1979); United States v. Midwest Video Corp. (Midwest Video I), 406 U.S. 649, 659-70 (1972).


\(^{18}\) Telecommunications carriers have

[the duty to provide, to any requesting telecommunications carrier for the provision of a telecommunications service, nondiscriminatory access to network elements on an unbundled basis at any technically feasible point on rates, terms, and conditions that are just, reasonable, and nondiscriminatory in accordance with the terms and conditions of the agreement and the requirements of this section and section 252 of this title. An incumbent local exchange carrier shall provide such unbundled network elements in a manner that allows requesting carriers to combine such elements in order to provide such telecommunications service.]

structural separation requirements; and concluding that incumbent carriers have no duty to deal with market entrants, even when the incumbent prices the wholesale rate charged to competitors above retail rates. For each of these decisions, the FCC compounded its initial mistakes by foreclosing the option to make necessary and lawful future modifications.

This article will examine the consequences of the FCC’s wishful thinking about the viability of current competition and the sustainability of competition going forward. The article concludes that flawed fact finding and market projections had adverse initial consequences but have even worse future impacts. In response to aggressive incumbent advocacy, impatient lawmakers keen on deregulation, and deferential judges willing to rely on the Commission’s expertise, the FCC has contributed to the development of a telecommunications industry structure that is less competitive, innovative, available, affordable, and responsive than what exists in many

order and order on remand and further notice of proposed rulemaking), corrected by Errata, 18 F.C.C.R. 19,020 (2003), vacated and remanded in part, aff’d in part, United States Telecom Ass’n v. FCC, 359 F.3d 554 (D.C. Cir. 2004).

The FCC eliminated Title II and structural separation requirements applicable to wireline broadband Internet-access services offered by facilities-based providers and gave providers discretion to offer the underlying wireline broadband transmission on a common-carrier basis. Wireline Broadband Classification Order, supra note 3.

See Pac. Bell Tel. Co. v. Linkline Commc’ns, Inc., 555 U.S. 438, 447-57 (2009) (no supplemental antitrust relief available when the FCC determines that a carrier has no duty to deal with a competitor).

To avoid “legislating from the bench” or second guessing the technical expertise of the FCC, reviewing courts typically defer to the Commission:

Our task on review is therefore limited. We review the FCC’s action in this case only to ensure that it is not “arbitrary, capricious, an abuse of discretion, or otherwise not in accordance with law.” 5 U.S.C. § 706(2)(A). That standard is particularly deferential in matters such as this, which implicate competing policy choices, technical expertise, and predictive market judgments.


“Unfortunately, the U.S. is lagging behind much of the rest of the world in terms of broadband service available to its citizens. As we move into a world in which ‘everyone will use the Internet for everything’ this country runs the risk of not being competitive.” RICHARD ADLER, THE ASPEN INSTITUTE, NEWS CITIES: THE NEXT GENERATION OF HEALTHY INFORMED COMMUNITIES 27 (2011), available at http://www.aspeninstitute.org/sites/default/files/content/docs/cands/News_Cities_The__Next_Generation_of_Healthy_Informed_Communities.pdf.

For example, even though the United States has the most broadband lines in use it only ranks 15th in terms of broadband market penetration (subscribers per 100 inhabitants) based on statistics compiled by the Organisation for Economic Co-Operation and Development on national broadband and telecommunications market penetration. See OECD Broadband Portal, OECD, http://www.oecd.org/document/54/ 0,3746,en_2649_33703_38690102_1_1_1_1,00.html (last visited Jan. 6, 2012).
other countries. The FCC’s follies provide a clear warning to other national regulatory authorities: embracing political and economic doctrine at the expense of unbiased fact finding and empirical analysis generates bad decisions that trigger even worse long-term outcomes.

Part II of this article will identify four FCC decisions that started a major deregulatory campaign based on unqualified conclusions about the existence and sustainability of competition. Section A examines the Commission’s decision to treat Internet access as an unregulated information service. Section B tracks the Commission’s deregulatory glide path for common carriers, including decisions to abandon precompetitive interconnection and access-pricing requirements as well as structural safeguards that separate carriers’ telecommunications and information services. This section, emphasizing antitrust and traditional duties to deal, also considers how reviewing courts respond to FCC deregulation and the Commission’s assumptions about market competitiveness. Section C examines the marketplace consequences of the FCC’s decision to allow wireless carriers to acquire unlimited spectrum regardless of the impact on market entry by new competitors. Part III offers conclusions on the short- and long-term consequences of premature deregulation when the marketplace has insufficient competition and market actors do not self-regulate.

II. THE PAST AS PROLOGUE TO THE FUTURE

On numerous occasions spanning several decades, the FCC has decided to abandon or reduce regulatory oversight. Technological innovations, changed circumstances, and a host of legitimate reasons can support selective deregulation. However, a significant number of initiatives, four of which are examined in depth in this article, were wrong at the outset. When the FCC makes a bad call, the normal checks and balances in government are supposed to provide remedies (e.g., judicial review). But well-argued rationales, coupled with shared views on economic doctrine and judicial deference to FCC expertise, can prevent appellate review from reversing

---

25 Supreme Court Justice Scalia shows how the FCC can exploit judicial deference to engage in policymaking outside its lawful jurisdiction:
bad decisions. Once in play, these decisions can trigger secondary and tertiary consequences that the FCC might not have predicted—consequences that, over time, compound the harm caused by the initial decision.

The four decisions examined in this article show how the FCC has engaged in results-driven decision making that lacks empirical support and uses legally unsustainable rationales to bolster preordained results. Authors of these decisions have emphasized stakeholder-submitted data without much close scrutiny by Commission staff or third-party peer review. 26 In the absence of independently generated data, the FCC has had to rely largely on stakeholder-submitted materials that support a particular outcome. Such reliance prevents the Commission from generating a realistic assessment based on a thorough and critical evaluation of all submissions, coupled with in-house fact finding and analysis. The agency has a statutory obligation to compile a complete factual record 27 and to accord interested parties opportunities to participate. 28 However, the Commission

---

This is a wonderful illustration of how an experienced agency can (with some assistance from credulous courts) turn statutory constraints into bureaucratic discretions. The main source of the Commission’s regulatory authority over common carriers is Title II, but the Commission has rendered that inapplicable in this instance by concluding that the definition of “telecommunications service” is ambiguous and does not (in its current view) apply to cable-modem service. It contemplates, however, altering that (unnecessary) outcome, not by changing the law (i.e., its construction of the Title II definitions), but by reserving the right to change the facts. Under its undefined and sparingly used “ancillary” powers, the Commission might conclude that it can order cable companies to “unbundle” the telecommunications component of cable-modem service. And presto, Title II will then apply to them, because they will finally be “offering” telecommunications service! Of course, the Commission will still have the statutory power to forbear from regulating them under § 160 (which it has already tentatively concluded it would do, Declaratory Ruling 4847–4848, ¶¶ 94–95). Such Möbius–strip reasoning mocks the principle that the statute constrains the agency in any meaningful way.


27 “[A] legislative choice is not subject to courtroom factfinding and may be based on rational speculation unsupported by evidence or empirical data.” FCC v. Beach Commc’ns, Inc., 508 U.S. 307, 315 (1993) (holding statutory requirement that satellite master antenna television system operators secure a franchise if they link separately owned buildings or use public rights of way constitutional even though single building service had no such franchising requirement).

28 See, e.g., Am. Radio Relay League, Inc. v. FCC, 524 F.3d 227, 231 (2008) (FCC “failed to satisfy the notice and comment requirement of the Administrative Procedure Act (‘APA’) by redacting studies on which it relied in promulgating the rule for

---

Nat’l Cable & Telecomm. Ass’n v. Brand X Internet Servs., 545 U.S. 967, 1013-14 (2005) (Scalia, J., dissenting) (rejecting the FCC’s rationale for considering cable modem service as lacking a standalone telecommunications service and noting how some reviewing courts fail to scrutinize closely the Commission’s analysis).
has primarily relied on the more comprehensive filings of the parties who have the most to gain or lose in a proceeding. It becomes easy for the FCC to rely on nonempirical data compiled by stakeholders that purport to supply data, but who in reality advocate for a desired outcome regardless of whether the facts support this objective.

A. Unconditional Conclusion that Broadband Access Constitutes an Information Service

The FCC has determined that the legislatively crafted information-service classification applies to Internet access provided via cable modems, digital subscriber line (DSL) service, the electrical power grid, and wireless networks. The Commission accrued short-term political dividends from such determinations because the determinations showed regulatory restraint and endorsed marketplace self-regulation. Whether the result of wishful thinking, inflexible

and failed to provide a reasoned explanation for its choice of the extrapolation factor for predicting how quickly broadband over powerline (BPL) emissions attenuate or weaken; see also Administrative Procedure Act, 5 U.S.C. § 553(b)-(c) (2006).

Information service is defined as,

the offering of a capability for generating, acquiring, storing, transforming, processing, retrieving, utilizing, or making available information via telecommunications, and includes electronic publishing, but does not include any use of any such capability for the management, control, or operation of a telecommunications system or the management of a telecommunications service.


Wireline Broadband Classification Order, supra note 3, at 14,863.


FCC managers are keenly aware of the political consequences resulting from changes in policy.

[In examining rulemaking and transitions in all three branches of government from the agency’s perspective, it may be most helpful to consider how the agency analyzes the costs and benefits of rulemaking. This cost-benefit calculation is quite different than the one typically discussed in administrative law—whether a particular regulation has net benefits to society. Instead, the calculation considers the net benefits of a rulemaking, both in terms of substance and process, to an agency in light of the particular costs to the agency. On the benefit side, the agency may care about the regulatory outcome; budgetary, political, and status rewards; and judicial deference. On the cost side, the agency may worry about
adherence to libertarian economic doctrine, or fair-minded interpretation of applicable statutes, the FCC determined that it must apply a single, mutually exclusive service classification.\textsuperscript{35} Under this either/or doctrine, the FCC opted to abandon any direct statutory foundation for mandating fair and open Internet access. Soon after these decisions, the FCC confronted instances where self-regulation did not prevent anticompetitive practices. Comcast, for example, interfered with the broadband traffic generated by some subscribers in ways that evidenced the incentive and ability to distort competition in the video-programming retail market.\textsuperscript{36} As discussed later in this section, Comcast, lacking effective FCC oversight, unilaterally thwarted subscriber access to competitive alternatives to the company’s pay-per-view video content.\textsuperscript{37}

With an eye toward freeing the Internet of government oversight, the Commission applied the substantially less restrictive information-service classification to all types of Internet-access services based on the view that the

\begin{quote}
We conclude, as the Commission did in the \textit{Universal Service Order}, that the categories of "telecommunications service" and "information service" in the 1996 Act are mutually exclusive. Reading the statute closely, with attention to the legislative history, we conclude that Congress intended these new terms to build upon frameworks established prior to the passage of the 1996 Act. Specifically, we find that Congress intended the categories of "telecommunications service" and "information service" to be mutually exclusive, like the definitions of "basic service" and "enhanced service" developed in our \textit{Computer II} proceeding, and the definitions of "telecommunications" and "information service" developed in the Modification of Final Judgment that divested the Bell Operating Companies from AT&T.

\textit{In re} Federal-State Joint Board on Universal Service, 13 FCC Rcd. 11,501, 11,507-08 (Apr. 10, 1998) (report to Congress) (citations omitted). "Although the Commission has not been entirely consistent on this point, we agree for the wireline broadband Internet access described in this Order with the past Commission pronouncements that the categories of ‘information service’ and ‘telecommunications service’ are mutually exclusive.” Wireline Broadband Classification Order, supra note 3, at 14,862 n.32.\textsuperscript{34} Comcast Sanction, \textit{supra} note 11, at 13,055-56.\textsuperscript{35}

“Comcast’s practice selectively blocks and impedes the use of particular applications, and we believe that such disparate treatment poses significant risks of anticompetitive abuse.” \textit{Id.} at 13,055.
telecommunications component needed to transmit bits and packets is inseparable from the content those bits contain. By treating the telecommunications component as subordinate, the Commission could rationalize a semantic distinction between a carrier providing telecommunications as a component of an information service and a carrier offering retail telecommunications services on a standalone basis. By opting to treat the telecommunications function as wholly

---

38 Telecommunications is defined as “the transmission, between or among points specified by the user, of information of the user’s choosing, without change in the form or content of the information as sent and received.” 47 U.S.C. § 153(43) (2006).

39 To justify its decision to apply the information-service classification to services that combine telecommunications transmission and content, the FCC insisted that the telecommunications component could not be singled out:

[We] reject arguments that companies using their own facilities to provide wireline broadband Internet access service simultaneously provide a telecommunications service to their end user wireline broadband Internet access customers. The record demonstrates that end users of wireline broadband Internet access service receive and pay for a single, functionally integrated service, not two distinct services. This conclusion also is consistent with certain past Commission pronouncements that the categories of “information service” and “telecommunications service” are mutually exclusive. Moreover, the fact that the Commission has, up to now, required facilities-based providers of wireline broadband Internet access service to separate out a telecommunications transmission service and make that service available to competitors on a common carrier basis under the Computer Inquiry regime has no bearing on the nature of the service wireline broadband Internet access service providers offer their end user customers. We conclude now, based on the record before us, that wireline broadband Internet access service is, as discussed above, a functionally integrated, finished product, rather than both an information service and a telecommunications service.

Wireline Broadband Classification Order, supra note 3, at 14,911 (citations omitted).

40 The Supreme Court accepted the FCC’s determination that cable modem Internet access constituted an information service:

Cable modem service is not itself and does not include an offering of telecommunications service to subscribers. We disagree with commenters that urge us to find a telecommunications service inherent in the provision of cable modem service. Consistent with the statutory definition of information service, cable modem service provides the capabilities described above “via telecommunications.” That telecommunications component is not, however, separable from the data-processing capabilities of the service. As provided to the end user the telecommunications is part and parcel of cable modem service and is integral to its other capabilities.

integrated into an information-service composite, the FCC could then abandon conventional common-carrier regulation required by Title II of the Communications Act.\footnote{41}{7 U.S.C. §§ 201-276.}

In the short term, the Commission championed regulatory restraint, a laudable goal that arguably contributed to the Internet’s speedy commercial success.\footnote{42}{Ex Parte Submission in CS Docket No. 02-52 from Tim Wu, Assoc. Professor, Univ. of Va. Sch. of Law, and Lawrence Lessig, Professor of Law, Stanford Law Sch., to Marlene H. Dortch, Sec’y, FCC 4 (August 22, 2003), available at http://www.timwu.org/wu_lessig_fcc.pdf.} However, the Commission soon discovered that—having given up on a direct statutory link—it would experience great difficulty in imposing any lawful safeguards, even when it received complaints of clearly abusive, discriminatory, and anticompetitive practices like those undertaken by Comcast.\footnote{43}{See Comcast Corp. v. FCC, 600 F.3d 642, 661 (D.C. Cir. 2010) (“[T]he allowance of wide latitude in the exercise of delegated powers is not the equivalent of untrammeled freedom to regulate activities over which the statute fails to confer . . . Commission authority.” (quoting Nat’l Ass’n of Regulatory Util. Comm’rs v. FCC, 533 F.2d 601, 618 (D.C. Cir. 1976))); Comcast Sanction, supra note 11, at 13,030-31.}

The FCC appeared quite confident that it could remedy any miscalculations and improper deregulation simply by invoking ancillary jurisdiction to revisit and revise its prior deregulation if consumer protection and other compelling circumstances warranted.\footnote{44}{The FCC overestimated its ability to apply Title I ancillary jurisdiction to re-regulate information services after having previously determined that it lacked statutory authority:}

When faced with instances where it had to remedy a problem (or make another information-service/telecommunications-service determination), the FCC has generated a mixed record. In some instances reviewing courts have deferred to the Commission’s

\textit{The Commission is empowered by statute to weigh these various objectives and craft regulations that specifically target the relevant features of VoIP and other IP-enabled services. Where the Act does not prescribe a particular regulatory treatment, the Commission may have authority to impose requirements under Title I of the Act. Alternatively, the Commission may forbear from applying specific provisions. Finally, of course, the Commission is entitled to amend or revoke its own rules and regulations when the underlying circumstances no longer apply.}

expertise and affirmed the assertion of jurisdiction and rules (e.g., requiring VoIP service providers to comply with many conventional telephone company requirements) despite the absence of direct statutory authority under Title II of the Communications Act. But in other cases, where equally compelling needs existed for the FCC to provide consumer safeguards, courts have deemed the Commission to lack sufficient statutory authority to act (e.g., sanctioning Comcast for deliberately preventing subscribers from transmitting and receiving video content via peer-to-peer traffic streams).

Having made an unconditional determination that the information-service, deregulated “safe harbor” applies to Internet access, the Commission could not subsequently reassert regulatory safeguards—no matter how necessary. When the FCC determined that only the information-service classification would apply, the Commission in effect determined that it had no direct statutory authority to impose regulatory requirements on telecommunications and other noninformation services that constitute a part of the blend of services contained in broadband Internet access. Even if the FCC could belatedly identify legitimate reasons for its

---

45 See, e.g., Vonage Holding Corp. v. FCC, 489 F.3d 1232 (D.C. Cir. 2007).
47 See generally Comcast Corp. v. FCC, 600 F.3d 642 (D.C. Cir. 2010).
48 A safe harbor constitutes “[a]n area or means of protection [or a] provision (as in a statute or regulation) that affords protection from liability or penalty." BLACK'S LAW DICTIONARY 1363 (8th ed. 2004). In light of the lack of a bright line distinction between regulated telecommunications services and largely unregulated information services, ventures can possibly secure a competitive advantage through regulatory arbitrage where ventures seek reduced regulatory oversight by characterizing telecommunications services as information services. The FCC defined regulatory arbitrage as “businesses making decisions based on regulatory classifications rather than on customers' preferences and innovative and sustainable business plans." In re Inquiry Concerning High-Speed Access to the Internet over Cable and Other Facilities, 17 FCC Rcd. 4798, 4846 (Mar. 15, 2002) (declaratory ruling and notice of proposed rulemaking). See generally Rob Frieden, Regulatory Arbitrage Strategies and Tactics in Telecommunications, 5 N.C. J.L. & TECH. 227 (2004).
intervention, the prior determination that cable, DSL, powerline, and wireless services qualified for deregulated safe harbors rendered them effectively off limits.\textsuperscript{49}

The FCC wrongly concluded that the broadband Internet access marketplace was so competitive that no provider would try to engage in anticompetitive practices. In reality the broadband marketplace offers limited options to most U.S. consumers with cable modem and DSL services predominating.\textsuperscript{50} Rather than making a proper deregulatory statutory interpretation, the FCC opined—incorrectly—that industry self-regulation would force carriers to offer low-cost

\textsuperscript{49} The Supreme Court rejected the FCC’s attempt to impose limited regulatory safeguards on information-service providers based on an extension of ancillary jurisdiction:

In this case we must decide whether the Federal Communications Commission has authority to regulate an Internet service provider’s network management practices. Acknowledging that it has no express statutory authority over such practices, the Commission relies on section 4(i) of the Communications Act of 1934, which authorizes the Commission to “perform any and all acts, make such rules and regulations, and issue such orders, not inconsistent with this chapter, as may be necessary in the execution of its functions.” 47 U.S.C. § 154(i). The Commission may exercise this “ancillary” authority only if it demonstrates that its action—here barring Comcast from interfering with its customers’ use of peer-to-peer networking applications—is “reasonably ancillary to the ... effective performance of its statutorily mandated responsibilities.” Am. Library Ass’n v. FCC, 406 F.3d 689, 692 (D.C. Cir. 2005). The Commission has failed to make that showing. It relies principally on several Congressional statements of policy, but under Supreme Court and D.C. Circuit case law statements of policy, by themselves, do not create “statutorily mandated responsibilities.” Comcast Corp., 600 F.3d at 644.

\textsuperscript{50} The FCC has overstated the level of broadband competition in the United States.

Contrary to claims of those who feel the U.S. has “robust broadband competition,” it is clear that half of the states have a duopoly rather than true competitive markets. The only question for these states is how much of a market share the top two providers collectively command. In states such as Ohio and Nevada, where there is a 30+ percentage gap between the top two providers, some will argue this is a monopoly. The other contention, that consumers and businesses have a wealth of options for providers (one industry executive estimated “everyone has at least four wireless carries, plus cable, satellite” etc.), also has flaws. This is perhaps true when taking in the nation as a whole, but when analyzed at the state and county levels which is where in reality the selection of possible providers actually exists, there are far fewer choices. Even in the most competitive states, the bottom five competitors have 5% market share or less. These competitors are obviously not offering services throughout their states, so clearly any remaining providers are less than a competitive force. Furthermore, if others are adding dial-up service providers to their list of consumer choices, this is disingenuous distraction because consumers know dial-up is Internet access but it isn’t broadband.

rates and refrain from engaging in anticompetitive conduct.\textsuperscript{51} In every instance where a regulatory safeguard has appeared necessary, explicitly or implicitly, for an information service, the FCC has had to scramble to find a lawful basis to reassert jurisdiction. This process has forced the FCC to spend countless hours devising creative and not-always-successful ways to backtrack from its previously clear and unequivocal determination. One example is the FCC’s attempt to sanction Comcast for deliberately interfering with its subscribers’ peer-to-peer file transfers, which contained some identical content to the company’s pay-per-view cable television service.\textsuperscript{52} The Commission determined that Comcast did not have legitimate traffic-management reasons for meddling with subscriber traffic\textsuperscript{53} and that the company lacked candor in its representation of what tactics it had used.\textsuperscript{54} Notwithstanding the commonly shared view that Comcast’s conduct justified FCC investigation and a remedy to safeguard consumers, the D.C. Circuit Court of Appeals rejected the FCC’s attempt to invoke ancillary jurisdiction as the lawful basis for sanctioning Comcast. The court determined that the FCC lacked a direct statutory basis for intervening:

In this case the Commission cites...[no section in the Communications Act of 1934] to shed light on any express statutory delegation of authority found in Title II, III, VI, or, for that matter, anywhere else. That is, unlike the way it successfully employed policy statements in Southwestern Cable and Midwest Video I, the Commission does not rely on section 230(b) or section 1 to argue that its regulation of an activity over which it concededly has no express statutory authority (here Comcast’s Internet management practices) is necessary to further its regulation of activities over which it does have express statutory authority (here, for example, Comcast’s

\textsuperscript{51} Wireline Broadband Classification Order, supra note 3, at 14,884-85 (discussing sufficiency of intermodal competition and price decline as a result of that competition).

\textsuperscript{52} “[T]he evidence reviewed above shows that Comcast selectively targeted and terminated the upload connections of its customers’ peer-to-peer applications and that this conduct significantly impeded consumers’ ability to access the content and use the applications of their choice.” Comcast Sanction, supra note 11, at 13,054.

Peer-to-peer applications, including those relying on BitTorrent, have become a competitive threat to cable operators such as Comcast because Internet users have the opportunity to view high-quality video with BitTorrent that they might otherwise watch (and pay for) on cable television. Such video distribution poses a particular competitive threat to Comcast’s video-on-demand (“VOD”) service.

\textsuperscript{53} Id. at 13,030.

\textsuperscript{54} “Comcast’s statements in its comments and response to Free Press’s complaint raise troubling questions about Comcast’s candor during this proceeding.” Id. at 13,032 n.31.
management of its Title VI cable services). In this respect, this case is just like *NARUC II*. On the record before us, we see “no relationship whatever,” *NARUC II*, 533 F.2d [601,] 616, between the Order and services subject to Commission regulation.55

Faced with a clear rebuke, FCC Chairman Julius Genachowski attempted to fashion a rationale for subdividing broadband access so that the Commission could identify and apply limited regulation of now identifiable telecommunications service components.56 This newfound severability of telecommunications services ran completely counter to the FCC’s previous rationale used to apply the information-service classification unconditionally to broadband Internet access. The Commission previously recognized the need for a telecommunications link to provide bit-and-packet transmission across distances; however, the Commission determined that this component was not a standalone retail service because it was seamlessly integrated with a predominant information service.57 The Supreme Court affirmed the FCC’s statutory interpretation, which served as the basis to treat cable modem Internet access as an information service.58

55 Comcast Corp., 600 F.3d at 654.
56 See Julius Genachowski, Chairman, Fed. Commc'n's Comm'n, The Third Way: A Narrowly Tailored Broadband Framework (May 6, 2010), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-297944A1.doc (proposing to apply Title II regulation only to the bit transmission portion of ISP services and rejecting a renewed attempt to find a way to extend Title I ancillary jurisdiction or reclassifying all aspects of Internet access as a telecommunications service); see also Austin Schlick, General Counsel, Fed. Commc'n's Comm'n, A Third-Way Legal Framework for Addressing the Comcast Dilemma (May 6, 2010), available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-297945A1.doc (providing legal rationale for narrow application of selected sections of Title II regulatory authority over Internet access).
57 To support its finding that broadband Internet access constitutes an information service, the FCC subordinated the telecommunications transport function and emphasized the nature of what content subscribers receive:

Thus, whether a telecommunications service is being provided turns on what the entity is “offering . . . to the public,” and customers' understanding of that service. End users subscribing to wireline broadband Internet access service expect to receive (and pay for) a finished, functionally integrated service that provides access to the Internet. End users do not expect to receive (or pay for) two distinct services—both Internet access service and a distinct transmission service, for example. Thus, the transmission capability is part and parcel of, and integral to, the Internet access service capabilities. Accordingly, we conclude that wireline broadband Internet access service does not include the provision of a telecommunications service to the end user irrespective of how the service provider may decide to offer the transmission component to other service providers.

Wireline Broadband Classification Order, supra note 3, at 14,910-11.
This newfound ability to segregate and identify a new telecommunications component to Internet access service is a scramble and a stretch. Previously, the Commission conveniently and expediently argued no such segregation could occur. The FCC subsequently abandoned this strategy and now asserts that it can still intervene and respond to complaints about ISP conduct based on other creative and novel interpretations of the Communications Act of 1934, as amended.59

Ostensibly structured to offer an acceptable compromise, the FCC issued a Report and Order that imposes basic public-interest obligations on ISPs,60 including four principles established in a 2005 statement61 and requirements that ISPs operate with transparency, nondiscrimination, and a commitment not to block lawful traffic.62 The Commission identified exceptions for reasonable network management,63 specialized services,64 and

---

59 In re Preserving the Open Internet, 25 FCC Rcd. 17,905 (Dec. 23, 2010) (report and order) [hereinafter Open Internet Report and Order]; see also In re Preserving the Open Internet, 24 FCC Rcd. 13,064 (Oct. 22, 2009) (notice of proposed rulemaking) [hereinafter Open Internet NPRM].
60 Specifically, the FCC imposed rules on the providers of broadband Internet access service, defined as a mass-market retail service by wire or radio that provides the capability to transmit data to and receive data from all or substantially all Internet endpoints, including any capabilities that are incidental to and enable the operation of the communications service, but excluding dial-up Internet access service. This term also encompasses any service that the Commission finds to be providing a functional equivalent of the service described in the previous sentence, or that is used to evade the protections set forth in this Part.
62 The FCC attempts to couch its open access initiative as consistent with prior bipartisan actions:

[W]e adopt three basic rules that are grounded in broadly accepted Internet norms, as well as our own prior decisions: i. Transparency. Fixed and mobile broadband providers must disclose the network management practices, performance characteristics, and terms and conditions of their broadband services; ii. No blocking. Fixed broadband providers may not block lawful content, applications, services, or non-harmful devices; mobile broadband providers may not block lawful websites, or block applications that compete with their voice or video telephony services; and iii. No unreasonable discrimination. Fixed broadband providers may not unreasonably discriminate in transmitting lawful network traffic.

63 “A network management practice is reasonable if it is appropriate and tailored to achieving a legitimate network management purpose, taking into account the particular network architecture and technology of the broadband Internet access service.” Id. at 17,952 (differentiating between reasonable network management practices that could affect how subscribers access content and unreasonable discriminatory practices).
wireless access. Notwithstanding its prior decision to apply the information-service classification that requires the FCC to eschew regulatory oversight, the Commission emphasized that the public-interest duty to ensure an open Internet required it to establish clear and certain rules applicable to both fixed (i.e., wire-based) and mobile (i.e., wireless) ISPs.

Having faced instances where it saw the need to intervene and resolve complaints about unfair and anticompetitive practices of a major national ISP, the FCC presented compelling arguments to impose public-interest safeguards. But in concluding that retail ISPs operate as information-service providers, the Commission acted on the assumption that an ISP like Comcast would never engage in such practices because robust competition would punish such self-serving conduct with substantial customer migration to alternative carriers promising not to interfere with customers' broadband traffic.

The FCC's Open Internet Report and Order would obligate all ISPs to "disclose [their] network management practices, performance characteristics, and terms and

---

64 “[S]pecialized services,’ such as some broadband providers’ existing facilities-based VoIP and Internet Protocol-video offerings, differ from broadband Internet access service . . . .” Id. at 17,965.

65 We will closely monitor the robustness and affordability of broadband Internet access services, with a particular focus on any signs that specialized services are in any way retarding the growth of or constricting capacity available for broadband Internet access service. We fully expect that broadband providers will increase capacity offered for broadband Internet access service if they expand network capacity to accommodate specialized services. We would be concerned if capacity for broadband Internet access service did not keep pace. We also expect broadband providers to disclose information about specialized services’ impact, if any, on last-mile capacity available for, and the performance of, broadband Internet access service. We may consider additional disclosure requirements in this area in our related proceeding regarding consumer transparency and disclosure.

Id. at 17,966.

66 Despite the likelihood that wireless network access will grow and perhaps become the primary way people access the Internet, the FCC established relaxed anti-blocking rules based on spectrum and operational limitations not applicable to wire-based networks.

A person engaged in the provision of mobile broadband Internet access service, insofar as such person is so engaged, shall not block consumers from accessing lawful websites, subject to reasonable network management; nor shall such person block applications that compete with the provider’s voice or video telephony services, subject to reasonable network management.

Id. at 17,959.

64 Id. at 17,908.

67 See id.
conditions of their broadband services. The FCC adopted different requirements for fixed and broadband providers on the other two key requirements. Fixed providers may not unreasonably discriminate in transmitting lawful network traffic, nor can they block lawful content, applications, services, or nonharmful devices. Mobile broadband providers may not block access to lawful websites or applications that compete with their voice or video services.

The Report and Order rejects assertions that network neutrality requirements would stifle innovation, reduce

---

68 Id. at 17,906.

A person engaged in the provision of broadband Internet access service shall publicly disclose accurate information regarding the network management practices, performance, and commercial terms of its broadband Internet access services sufficient for consumers to make informed choices regarding use of such services and for content, application, service, and device providers to develop, market, and maintain Internet offerings.

Id. at 17,937.

69 “A person engaged in the provision of fixed broadband Internet access service, insofar as such person is so engaged, shall not block lawful content, applications, services, or non-harmful devices, subject to reasonable network management.” Id. at 17,942.

70 Id. at 17,959-60.

71 Network neutrality refers to the imposition of nondiscrimination, transparency, and other requirements on ISPs. The requirements are designed to foster a level, competitive playing field among content providers and to establish consumer safeguards so that Internet users have access limited only by legitimate concerns such as ISP network management and national security. See Rob Frieden, A Primer on Network Neutrality, 43 INTERECOMMICS: REV. EUR. ECON. POL’Y 4, 5 (2008).

incentives to invest in network infrastructure, and hamper employment in the Internet economy:

We believe these rules, applied with the complementary principle of reasonable network management, will empower and protect consumers and innovators while helping ensure that the Internet continues to flourish, with robust private investment and rapid innovation at both the core and the edge of the network. This is consistent with the National Broadband Plan goal of broadband access that is ubiquitous and fast, promoting the global competitiveness of the United States.\textsuperscript{72}

Despite the strident dissents from the two Republican Commissioners, the Report and Order appears to emphasize that the final rules logically follow from the nonpartisan consensus reached in documents created in 2005 and 2007.\textsuperscript{73} Further, the Report and Order claims that the requirements do

\textsuperscript{72} Open Internet Report and Order, supra note 59, at 17,906.

\textsuperscript{73} The FCC attempts to frame the Open Internet Report and Order as noncontroversial and a lawful exercise of statutory authority:

The rules we proposed in the \textit{Open Internet NPRM} and those we adopt today follow directly from the Commission’s bipartisan Internet Policy Statement, adopted unanimously in 2005 and made temporarily enforceable for certain broadband providers in 2005 and 2007; openness protections the Commission established in 2007 for users of certain wireless spectrum; and a notice of inquiry in 2007 that asked, among other things, whether the Commission should add a principle of nondiscrimination to the Internet Policy Statement. Our rules build upon these actions, first and foremost by requiring broadband providers to be transparent in their network management practices, so that end users can make informed choices and innovators can develop, market, and maintain Internet-based offerings. The rules also prevent certain forms of blocking and discrimination with respect to content, applications, services, and devices that depend on or connect to the Internet.

\textit{Id.} at 17,907-08 (citations omitted).
not violate the Constitution, particularly First Amendment expression rights of ISPs and the prohibition on government takings in the Fifth Amendment.

Additionally, the Report and Order extensively attempts to demonstrate that the FCC has lawful jurisdiction to promulgate network neutrality rules, primarily because Congress, in Section 706 of the Telecommunications Act, authorized the FCC to take all reasonable steps to promote widespread access to the Internet. In light of the Comcast case, the Commission must establish clear and direct statutory authority to impose new rules. The Commission heavily relied on Section 706 of the Telecommunications Act, which does not explicitly authorize regulation and rule making. The FCC inferred that the duty to encourage the deployment of “advanced telecommunications capability” authorizes the Commission to use whatever tools it considers necessary to achieve timely progress.

The FCC’s assumption of statutory authority requires two novel reinterpretations of the definition of telecommunications contained in the Communications Act. First, the FCC has to consider advanced telecommunications capability to include Internet access, despite having

---

74 See id. at 17,981-87.
76 Open Internet Report and Order, supra note 59, at 17,985.
78 See Open Internet Report and Order, supra note 59, at 17,966-81.
79 Comcast Corp. v. FCC, 600 F.3d 642, 661 (D.C. Cir. 2010).
80 Open Internet Report and Order, supra note 59, at 17,968.
82 The FCC inferred that Section 706 of the 1996 Act confers broad authority to revise the scope of regulatory oversight to promote Internet access:

As noted, Section 706 of the 1996 Act directs the Commission (along with state commissions) to take actions that encourage the deployment of “advanced telecommunications capability.” . . . Under Section 706(a), the Commission must encourage the deployment of such capability by “utilizing, in a manner consistent with the public interest, convenience, and necessity,” various tools including “measures that promote competition in the local telecommunications market, or other regulating methods that remove barriers to infrastructure investment.”

Open Internet Report and Order, supra note 59, at 17,968.
83 “[A]dvanced telecommunications capability,’ as defined in the statute, includes broadband Internet access.” Id. at 17,968 (citing 47 U.S.C. § 1302(d)(1) (defining “advanced telecommunications capability” as “high-speed, switched, broadband telecommunications capability that enables users to originate and receive high-quality voice, data, graphics, and video telecommunications using any
previously concluded that the technologies providing such access constitute information services that only integrate telecommunications (but do not constitute telecommunications services in and of themselves). The Commission previously determined that the telecommunications transmission of bits and packets in Internet access is not severable from the predominant information service offered, but instead provided as a subordinate part of an information service that an ISP offers to end users. Second, the FCC now has to elevate the significance of the telecommunications bit-transmission function in Internet access to trigger public-interest concerns about competition and anticompetitive practices, even though the Commission had previously qualified Internet-access technologies for an unregulated safe harbor status. Now the FCC wants to validate the telecommunications component as the driver for public-interest regulatory safeguards.

Despite having previously concluded that the broadband marketplace was robustly competitive and close to ubiquitous, the Commission cited to better-calibrated market penetration data to support its involvement:

Section 706(b) of the 1996 Act provides additional authority to take actions such as enforcing open Internet principles. It directs the
Commission to undertake annual inquiries concerning the availability of advanced telecommunications capability to all Americans and requires that, if the Commission finds that such capability is not being deployed in a reasonable and timely fashion, it "shall take immediate action to accelerate deployment of such capability by removing barriers to infrastructure investment and by promoting competition in the telecommunications market." In July 2010, the Commission "conclude[d] that broadband deployment to all Americans is not reasonable and timely" and noted that "[a]s a consequence of that conclusion," Section 706(b) was triggered. Section 706(b) therefore provides express authority for the pro-investment, pro-competition rules we adopt today.

Additionally, the FCC applied portions of Titles II, III, and VI of the Communications Act to ISPs despite the fact that Title II customarily applies to common carriers, Title III to broadcasters and wireless carriers, and Title VI to cable television operators. Instead of stating that ISPs operate as telecommunications service carriers when they provide essential first and last-mile access to the Internet—a scenario suggested by FCC Chairman Julius Genachowski and now apparently rejected—the Report and Order states that because some Internet-based services compete with traditional telephone, broadcast, and video services, the Commission has jurisdiction to impose rules and regulations to prevent anticompetitive practices and to promote competition.

The FCC justified the imposition of network neutrality rules on ISPs with the conclusion that ISPs have the incentive and ability to engage in anticompetitive practices that limit Internet openness in terms of content, applications, services, and devices accessed over, or connected to, broadband Internet access services. The Commission provided three examples suggesting that ISPs may have incentives to block or degrade content that competes with what the ISP or an affiliate offers, to impose surcharges on competing content providers in addition to end user subscription fees, and to degrade competitors’ traffic:

[1] Broadband providers may have economic incentives to block or otherwise disadvantage specific edge providers or classes of edge providers, for example by controlling the transmission of network traffic over a broadband connection, including the price and quality of access to end users. A broadband provider might use this power to

---

86 Open Internet Report and Order, supra note 59, at 17,972.
87 Id. at 17,972-80.
88 Id.
89 Id. at 17,907.
benefit its own or affiliated offerings at the expense of unaffiliated offerings.\textsuperscript{90}

[2] Broadband providers may have incentives to increase revenues by charging edge providers, who already pay for their own connections to the Internet, for access or prioritized access to end users. Although broadband providers have not historically imposed such fees, they have argued they should be permitted to do so. A broadband provider could force edge providers to pay inefficiently high fees because that broadband provider is typically an edge provider’s only option for reaching a particular end user. Thus broadband providers have the ability to act as gatekeepers.\textsuperscript{91}

[3] If broadband providers can profitably charge edge providers for prioritized access to end users, they will have an incentive to degrade or decline to increase the quality of the service they provide to non-prioritized traffic. This would increase the gap in quality (such as latency in transmission) between prioritized access and non-prioritized access, induce more edge providers to pay for prioritized access, and allow broadband providers to charge higher prices for prioritized access. Even more damaging, broadband providers might withhold or decline to expand capacity in order to “squeeze” non-prioritized traffic, a strategy that would increase the likelihood of network congestion and confront edge providers with a choice between accepting low-quality transmission or paying fees for prioritized access to end users.\textsuperscript{92}

The FCC considers the three examples of discrimination as more than theoretical in light of actual examples where ISPs, such as Comcast, have blocked or degraded traffic without legitimate network management concerns.\textsuperscript{93} Similarly,
the Commission has stated that the benefits in guarding against such anticompetitive practices outweigh the costs.\textsuperscript{94}

The FCC’s latest attempt to circumvent its information-service classification of broadband Internet access may not pass muster with a reviewing court.\textsuperscript{95} The Commission avoided repeating the Title I ancillary jurisdiction strategy as well as Chairman Genachowski’s proposed surgical removal of telecommunications-service elements from information services. But the Commission has come up with similarly triangulating strategies: Title III confers broad authority for the FCC to impose any necessary safeguard over spectrum-using services—arguably including wireless broadband\textsuperscript{96}—and Section 706 of the Telecommunications Act of 1996 both encourages and authorizes any well-articulated rationale for regulating information services, which promotes wider access to broadband services.\textsuperscript{97}

Had the FCC acknowledged years ago that public access to information services might trigger conflicts not readily resolved by the marketplace, the Commission would have been able to retain limited and nonintrusive jurisdiction to respond to complaints. Telecommunications and information markets and technologies have converged, and it is now more difficult for the FCC to determine the exact scope of its lawful jurisdiction and the line between regulated telecommunications services and unregulated information services. Rather than acknowledge the

\begin{itemize}
  \item The FCC attempted to downplay the significance of its order and the burdens it imposed:

  \begin{quote}
  By comparison to the benefits of these prophylactic measures, the costs associated with the open Internet rules adopted here are likely small. Broadband providers generally endorse openness norms—including the transparency and no blocking principles—as beneficial and in line with current and planned business practices (though they do not uniformly support rules making them enforceable) Even to the extent rules require some additional disclosure of broadband providers’ practices, the costs of compliance should be modest.
  \end{quote}

\end{itemize}

Open Internet Report and Order, supra note 59, at 17,928.


\textsuperscript{96} See, e.g., 47 U.S.C. § 332 (2006) (applying regulations to wireless commercial mobile radio service operators using Title III that generally address broadcast spectrum use).

\textsuperscript{97} See id. § 1302.
need to make ad hoc determinations and to resolve conflicts, the Commission blithely assumed that a competitive marketplace would provide solutions to consumers and remedies to any and all problems. Such reliance comes across as misguided, particularly in light of the conflicts the FCC has faced involving Internet access and how to justify its intervention.

Ironically, even as the FCC appears to have abandoned oversight of information services completely, it has devised a judicially approved model for asserting jurisdiction over a new hybrid service that combines telecommunications and information services: VoIP. The Commission has established an extensive body of decisions on what obligations VoIP service providers must undertake to serve the public interest. Bear in mind that many of these obligations impose significant costs on VoIP carriers, thereby reducing their competitiveness and ability to offer a cheaper alternative to existing wired and wireless services. Although VoIP arguably constitutes a type of information service, the FCC has managed to avoid having to make that determination even as the Commission requires VoIP operators to incur the same obligations as Title II–regulated common-carrier telephone companies. VoIP service providers that can receive or deliver calls to conventional wired and wireless networks must contribute to universal service funding programs designed to promote affordable dial-up telephone service; make arrangements to support subscriber

---

98 VoIP customers initiate and receive calls via their broadband links, e.g., DSL and cable modem services. The FCC considers broadband access an information service. In re Inquiry Concerning High-Speed Access to the Internet over Cable and Other Facilities, 17 FCC Rcd. 4798, 4802 (Mar. 15, 2002) (declaratory ruling and notice of proposed rulemaking), aff’d, Nat’l Cable & Telecomms. Ass’n v. Brand X Internet Servs., 545 U.S. 967 (2005); Wireline Broadband Classification Order, supra note 3. It follows that software and other applications carried via information-service links similarly qualify as information services.

99 The FCC has managed to avoid making a specific regulatory classification of VoIP, despite having imposed Title II regulatory requirements:

- To date, the Commission has not classified interconnected VoIP service as either an information service or a telecommunications service. The Commission has, however, extended certain obligations to providers of such service, including local number portability, 911 emergency calling capability, universal service contribution, CPNI protection, disability access and TRS contribution requirements, and section 214 discontinuance obligations.

In re Connect America Fund, 26 FCC Rcd. 4554, 4582 (Feb. 9, 2011) (notice of proposed rulemaking and further notice of proposed rulemaking) (citations omitted).

100 In re Universal Service Contribution Methodology, 21 FCC Rcd. 7518, 7538 (June 27, 2006) (report and order and notice of proposed rulemaking) (extending section 254(d) permissive authority to require interconnected VoIP providers to
access to emergency 911 service;\textsuperscript{101} cooperate with law enforcement authorities;\textsuperscript{102} incorporate the technical accommodations for persons with disabilities,\textsuperscript{103} such as deaf callers; support the ability of existing subscribers to keep their existing telephone numbers when switching services;\textsuperscript{104} and report service outages to the Commission.\textsuperscript{105}

The FCC can impose consumer-oriented safeguards on VoIP service providers based on a more persuasive and better-articulated assertion of ancillary jurisdiction. Because VoIP competes with conventional wired and wireless services subject to Title II regulation, the Commission can impose the very same requirements on VoIP carriers despite the lack of specific Title II authority.\textsuperscript{106} Reviewing courts have affirmed the Commission’s jurisdiction as well as its preemption of the states from imposing a different regulatory regime, or none at all.\textsuperscript{107} But success in selectively regulating VoIP service does not extend to other information services because a less-direct impact on a regulated service exists and also because of the FCC’s summary conclusion that all information services qualify for deregulation.

B. Eliminating Common Carrier Duties

The FCC has streamlined and even deregulated some telecommunications services based on criteria contained in the

\textsuperscript{101} In re IP-Enabled Servs., 22 FCC Rcd. 10,245 (June 3, 2005) (first report and order and notice of proposed rulemaking).
\textsuperscript{102} In re Communications Assistance for Law Enforcement Act & Broadband Access & Services, 20 FCC Rcd. 14,989 (Sept. 23, 2005) (first report and order and further notice of proposed rulemaking).
\textsuperscript{103} In re IP-Enabled Services, 22 FCC Rcd. 11,275 (June 15, 2007) (report and order); In re IP-Enabled Services, 22 FCC Rcd. 18,319 (Oct. 9, 2007) (order and public notice seeking comment) (granting in part and denying in part waivers of the FCC order); see also In re Contributions to the Telecommunications Relay Services Fund, 26 FCC Rcd. 3285 (Mar. 3, 2011).
\textsuperscript{104} In re Telephone Number Requirements for IP Enabled Services Providers, 22 FCC Rcd. 19,531 (Nov. 8, 2007) (report and order, declaratory ruling, order on remand, and notice of proposed rulemaking); In re Matters of Local Number Portability Porting Interval and Validation, 25 FCC Rcd. 6953 (May 20, 2010) (report and order) (establishing short deadlines for conversions).
\textsuperscript{106} In re Connect America Fund, 26 FCC Rcd. 4554, 4582 (Feb. 9, 2011) (notice of proposed rulemaking and further notice of proposed rulemaking) (citations omitted).
Telecommunications Act\textsuperscript{108} and, more broadly, in light of expanded competition. In many instances the Commission wisely has forborne from applying conventional “command and control,” “heavy-handed” regulation in light of carriers’ ability to self-regulate and consumers’ ability to pursue service options.\textsuperscript{109} However, the Commission has accelerated the deregulatory glide path in some market segments based on wishful thinking and flawed assessments of the robustness and sustainability of competition.\textsuperscript{110} The markets for equipment,\textsuperscript{111}

\textsuperscript{108} Section 10 of the Telecommunications Act of 1996, codified at 47 U.S.C. § 160 (2006), requires the FCC to forbear from any statutory provision or regulation if the Commission determines that:

\begin{itemize}
  \item[(1)] enforcement of such regulation or provision is not necessary to ensure that the charges, practices, classifications, or regulations by, for, or in connection with that telecommunications carrier or telecommunications service are just and reasonable and are not unjustly or unreasonably discriminatory;
  \item[(2)] enforcement of such regulation or provision is not necessary for the protection of consumers;
  \item[(3)] forbearance from applying such provision or regulation is consistent with the public interest.
\end{itemize}

47 U.S.C. § 160(a) (2006). In making such determinations, the Commission must also consider “whether forbearance from enforcing the provision or regulation will promote competitive market conditions.” \textit{Id.} § 160(b). Section 160(d) specifies, however, that “[e]xcept as provided in section 251(f) . . . the Commission may not forbear from applying the requirements of section 251(c) or 271 . . . until it determines that those requirements have been fully implemented.” \textit{Id.} § 160(d). Section 332(c) of the Communications Act, 47 U.S.C. § 332(c), authorizes the Commission to refrain or forbear from enforcing any provision other than the core requirements of sections 201, 202, and 208, which respectively require just and reasonable charges, practices, classifications, and regulations, prohibit unreasonable discrimination and carrier practices, and require the FCC to investigate complaints.

\textsuperscript{109} See Earthlink, Inc. v. FCC, 462 F.3d 1 (D.C. Cir. 2006) (affirming the FCC’s decision to forbear from imposing most local loop unbundling requirements on incumbent carriers); U.S. Telecom Ass’n v. FCC, 359 F.3d 554, 588 (D.C. Cir. 2004) (upholding the FCC’s nationwide decision to refrain from requiring § 251 unbundling fiber broadband elements and reversing the Commission’s decision not to eliminate other unbundling requirements in light if the adverse impact on carrier investment incentives).

\textsuperscript{110} The FCC previously did not even require applicants for regulatory forbearance to demonstrate how marketplace conditions specifically supported less government oversight:

We acknowledge that we have not previously required petitioners to specify in the petition how the requested relief meets each of the three forbearance criteria, and that a requirement to do so will burden applicants to the extent that they must develop their supporting arguments in advance of filing. We do not, however, consider this an unreasonable expectation, and we find that the benefit to both the Commission and clarity and precision outweighs the burden on the petitioner of explaining how forbearance from each regulation or statutory provision meets each prong.

\textit{In re Petition to Establish Procedural Requirements to Govern Proceedings for Forbearance Under Section 10 of the Communications Act of 1934, as Amended}, 24 FCC Rcd. 9543, 9551 (June 29, 2009) (report and order) [hereinafter Forbearance Criteria Order].

\textsuperscript{111} In re Use of the Carterfone Device in Message Toll Telephone Service, 13 F.C.C.2d 420 (June 26, 1968) (decision); \textit{In re Telerent Leasing Corp.}, 45 F.C.C.2d 204
wiring located on customers’ premises,\textsuperscript{112} and long-distance telephone services\textsuperscript{113} provide clear examples of prudent regulatory streamlining. But similar initiatives for the first-, last-, and middle-mile services\textsuperscript{114} that link end users with major

(Feb. 5, 1974), aff’d sub nom. N.C. Util. Comm’n v. FCC, 537 F.2d 787 (4th Cir. 1976); In re Mebane Home Telephone Co., 53 F.C.C.2d 473, 474 (June 4, 1975), aff’d sub nom. Mebane Home Tel. Co. v. FCC, 535 F.2d 1324 (D.C. Cir. 1976); see also Pub. Util. Comm’n of Tex. v. FCC, 886 F.2d 1325 (D.C. Cir. 1989) (noting long established FCC policy that carriers and non-carriers alike have a federal right to interconnect to the public telephone network in ways that are privately beneficial if they are not publicly detrimental). Previous FCC opposition to this principle failed to pass muster with a reviewing court that interpreted the Communications Act as mandating the right of consumers to attach equipment to the network in ways that were privately beneficial but not publicly harmful. Hush-A-Phone Corp. v. United States, 238 F.2d 266 (D.C. Cir. 1956). “The interveners’ tariffs [prohibiting the use of a plastic device to enhance privacy and low volume conversations], under the Commission’s decision, are in [sic] unwarranted interference with the telephone subscriber’s right reasonably to use his telephone in ways which are privately beneficial without being publicly detrimental.” \textit{Id.} at 269.


\textsuperscript{114} The FCC categorizes Internet access into three types based on geographical location and function:

Today, the Internet has evolved from its early stages and is comprised of three types of interconnected networks. The first category, Backbone Providers, supply long-distance high-speed “connections between a small number of interconnection points.” Second, there are Middle-Mile Providers who supply regional distributive functions; for example, a connection from a Backbone Provider to a distant city’s central office maintained by an ISP. Finally, there are Last-Mile Providers who connect Middle-Mile Providers to end users (consumers). Although ISPs were historically considered Last-Mile Providers, it is often the case for broadband capable networks that the ISP is both the Last-Mile Provider and the Middle-Mile Provider. This system of connected networks is most analogous to a road system: Backbones represent interstate highways; Middle-Mile networks are the intrastate highways; and Last-Mile networks are the local roads that ultimately reach consumers.


Middle-mile facilities are shared assets for all types of last-mile access. As such, the cost analysis is very similar regardless of last-mile infrastructure. The local aggregation point can vary based on technology (e.g., a cable headend, LEC central office or a wireless mobile switching center (MSC)) while the Internet gateway is a common asset. Middle-mile facilities are widely deployed but can be
broadband long-haul networks exemplify premature abandonment of regulatory safeguards in light of the onset of little competition, particularly in rural areas.\(^{115}\)

expensive in rural areas because of the difficulties of achieving local scale, thereby increasing the investment gap. On a per-unit basis, middle-mile costs are high in rural areas due to long distances and low aggregate demand when compared to middle-mile cost economics in urban areas. While there may be a significant affordability problem with regard to middle-mile access, it is not clear that there is a middle-mile fiber deployment gap.


\(^{115}\) The FCC’s conclusions about broadband competitiveness has generated substantial opposition:

The course the Commission has followed over the past eight years has turned out to be spectacularly wrong in all of those aspects. There is little to no competition for broadband services in the residential and “middle mile” markets. As a result, U.S. consumers pay higher rates for services with slower speeds than do consumers in other industrialized nations. Our record of online innovation has slowed to a crawl. The U.S.’s standing in the world ranking of broadband adoption falls continually. (One can look at various rankings and dispute any given position, but the trend in all of them is clear: America is clearly falling behind.)


The reason the U.S. is falling behind can be traced directly to the decisions the Commission made over the past 10 years to reclassify broadband service, taking it out of the environment of Title II while moving it into the more legally murky area of Title I by classifying broadband as an “information service” instead of as a “telecommunications service.” Now is the time to recognize that this deliberate decision to deregulate by redefinition failed to produce the promised land of “intermodal competition” and reverse that decision.

Id. at 177.

Rural broadband networks are fundamentally similar to broadband networks in other areas in that, in order to have broadband access to the Internet, they must include local access, or last-mile, broadband access to the end user and backhaul, or middle-mile, capabilities to an available Internet peering point. The last-mile network connects residential and business end users to a local ISP. In this configuration, the middle-mile or backhaul component connects the local ISP to an Internet peering point or node. In rural settings, either or both of these components may not support robust broadband connectivity.

In three instances of streamlined regulatory oversight discussed below, the FCC eliminated statutory duties to deal, which, in turn, short-circuited both the prospect for true facilities-based competition and effective judicial review.\textsuperscript{116} In its zeal to eliminate common-carrier regulations, based on a questionable finding of robust and sustainable competition, the FCC has abandoned requirements that local exchange carriers: (1) provide market entrants interconnection with their switching and routing facilities on congressionally mandated favorable terms and conditions;\textsuperscript{117} (2) separate their basic transmission facilities from services that provide enhancements to these basic transmission links;\textsuperscript{118} and (3) refrain from offering end-user retail services at rates below the wholesale rate offered other carriers.\textsuperscript{119}

In all three instances the FCC eliminated regulatory requirements based on the view that they were not needed to ensure that consumers could acquire diverse services at competitive rates. After failing to convince the FCC that such streamlining did not serve the public interest, consumer advocates and recent market entrants were similarly unsuccessful at convincing appellate courts that the Commission erred in its fact finding.\textsuperscript{120} On two separate occasions the Supreme Court has stated clearly that if the FCC determines that no regulatory safeguards are necessary, then reviewing courts

\textsuperscript{116} See infra Part II.B.3.


\textsuperscript{120} For example, a reviewing court did not question the FCC's conclusion that a sufficiently competitive market existed for telecommunications services linking end users with ISPs and other service providers. Ad Hoc Telecomm. Users Comm. v. FCC, 572 F.3d 903, 904 (D.C. Cir. 2009).
should not second guess the Commission and therefore should not apply a more rigorous antitrust standard or duty to deal.\footnote{121} Thus, if the FCC overstates the competitiveness and regulatory capability of telecommunications-service markets, recent case precedent states that appellate courts will not correct the Commission’s mistakes but instead summarily validate the Commission’s determination that such carriers have no duty to deal with other carriers.

1. Abandonment of Local Loop Unbundling

The Telecommunications Act of 1996\footnote{122} sought to stimulate local exchange service competition by creating a combination of specific common-carrier responsibilities on telecommunications carriers\footnote{123} with additional requirements on the Bell Telephone companies that were spun off from AT&T in 1984.\footnote{124} In exchange for satisfying a fourteen-point competitive checklist,\footnote{125} the spun-off Bell Telephone companies could seek FCC authorization to provide long-distance telephone services, a line of business prohibited since AT&T’s divestiture.\footnote{126} Included in that list was a requirement that they provide network access on an à la carte or combined basis at rates well below what the incumbent carriers would seek to charge even at wholesale.\footnote{127} Congress hoped that the Bell companies’ entry

\footnotesize{\begin{itemize}
\item \footnote{121} Verizon Comm'ns Inc. v. Law Offices of Curtis V. Trinko, LLP, 540 U.S. 398, 410 (2004); Linkline, 555 U.S. at 450.
\item \footnote{123} 47 U.S.C. § 251 (2006) (duties applicable to all telecommunications carriers). 47 U.S.C. § 252(c)(3) requires all telecommunications to provide, to any requesting telecommunications carrier for the provision of a telecommunications service, nondiscriminatory access to network elements on an unbundled basis at any technically feasible point on rates, terms, and conditions that are just, reasonable, and nondiscriminatory in accordance with the terms and conditions of the agreement and the requirements of this section and section 252 of this title. An incumbent local exchange carrier shall provide such unbundled network elements in a manner that allows requesting carriers to combine such elements in order to provide such telecommunications service.
\item \footnote{124} 47 U.S.C. § 271 (duties that the Bell telephone companies must satisfy to qualify for the opportunity to pursue prohibited lines of business such as most long distance telephone services).
\item \footnote{126} Bellsouth Corp. v. FCC, 162 F.3d 678, 680-81 (D.C. Cir. 1998).
\item \footnote{127} The Supreme Court did not dispute the right of Congress to require the FCC to create new rate-setting methods with an eye toward expediting market entry in the local exchange marketplace:
\end{itemize}
into long-distance services would further stimulate competition in that market. Congress also believed that the interconnection requirements imposed on these carriers would jump-start local service competition.\[128\] But over time, the Bell companies faced a robustly competitive long-distance telephone service market with low margins and less-than-desired upside business opportunities.\[129\] The mandated promotional pricing of local exchange facilities stimulated market entry by new competitive local exchange carriers (CLECs), but sustainable, long-term competition by facilities-based carriers did not result.\[130\]

The Act thus appears to be an explicit disavowal of the familiar public-utility model of rate regulation (whether in its fair-value or cost-of-service incarnations) presumably still being applied by many States for retail sales, see In re Implementation of Local Competition in Telecommunications Act of 1996, 11 F.C.C.R. 15,499, 15,857, ¶ 704 (1996) (First Report and Order), in favor of novel ratsetting designed to give aspiring competitors every possible incentive to enter local retail telephone markets, short of confiscating the incumbents’ property.


\[128\] The Telecommunications Act of 1996 required incumbent local exchange carriers to cooperate with market entrants.

Until the 1990’s, local phone service was thought to be a natural monopoly. States typically granted an exclusive franchise in each local service area to a local exchange carrier (LEC), which owned, among other things, the local loops (wires connecting telephones to switches), the switches (equipment directing calls to their destinations), and the transport trunks (wires carrying calls between switches) that constitute a local exchange network. Technological advances, however, have made competition among multiple providers of local service seem possible, and Congress recently ended the longstanding regime of state-sanctioned monopolies. The Telecommunications Act of 1996 (1996 Act or Act), Pub. L. 104-104, 110 Stat. 56, fundamentally restructures local telephone markets. States may no longer enforce laws that impede competition, and incumbent LECs are subject to a host of duties intended to facilitate market entry.


\[129\] The long distance toll service marketplace has become robustly competitive with low profit margins.

Until the 1970s, AT&T had a virtual monopoly on long distance service in the United States. In the 1970s, competitors such as MCI and Sprint began also to offer long distance service. With the gradual emergence of competition, basic rates dropped, calling surged, and AT&T’s dominance declined. More than 1,900 toll companies now offer long distance service of which more than 1,400 are wireline carriers. These carriers remain subject to the Commission’s jurisdiction. The Commission, however, has chosen to rely on competition, rather than regulation, as much as possible. Thus, the Commission forbears from regulating most aspects of long distance service.


\[130\] Legislative and FCC attempts to promote local exchange competition failed:

It was both the intent of Congress and the target of intense and sustained FCC efforts to open up the incumbent local exchange carriers’ (ILECs) local access lines
Frustrated by the combination of low long-distance margins and the ongoing duty to bolster the market share of newcomers, incumbent carriers sought judicial relief. Initially, even the Supreme Court favored the FCC’s interpretation of the ‘96 Act’s requirements. The Court determined that the FCC could lawfully require promotional pricing that used a costing model which justified access prices well below existing wholesale rates instead of actual, current, and already-incurred costs. Similarly, the Court held that such mandatorily low interconnection rates did not constitute an unconstitutional taking of incumbent-carrier property because the carriers never proved that any undertaking resulted in a financial loss, only less-than-desired financial gains. However, the Court and other lower appellate tribunals later agreed that the FCC’s interconnection pricing mandate lacked sufficient calibration to ensure that the promotional pricing only occurred where absolutely necessary to jump-start competition. As time

to competitive local exchange carriers (CLECs) who could then compete against the ILECs for “last mile” services without having to build their own access lines. Seldom have the forces of public policy in telecommunications been as powerfully aligned as they were on the issue of local-loop unbundling. And yet, the effort was a failure—the evidence for which is the demise of the CLECs. The reasons for this failure are clear: (i) the interface between the regulated monopoly owning the local-access line and the CLECs who wished to use it was highly complex; and (ii) the ILECs not only owned the local loops, they also competed in the retail market for access services with the very CLECs who had to use their facilities. The result was that ILECs had every incentive to make life miserable for the CLECs in any way they could, and the complexity of the interface gave them plenty of opportunity.


132 In Verizon Communications, Inc. v. FCC, the Supreme Court rejected ILECs’ arguments that using a theoretical, most-efficient-cost model, instead of actual historical costs, constituted a taking that violated the Fifth Amendment. Id. The Court noted that no party had disputed any specific rate established by the FCC’s forward-looking, long-run incremental cost-pricing methodology, and concluded that “[r]egulatory bodies required to set [just and reasonable] rates . . . have ample discretion to choose methodology.” Id. at 499. Additionally the Court stated that the Telecommunications Act of 1996 did not specifically require historical costs, particularly in light of its explicit prohibition on the use of conventional “rate-of-return or other rate-based proceeding” . . . which has been identified with historical cost ever since Hope Natural Gas was decided.” Id. at 499-500; see also AT&T Corp. v. Iowa Utils. Bd., 525 U.S. 366 (1999) (largely upholding the FCC’s implementation of the Congressional mandate contained in Section 251 of the Telecommunications Act of 1996 as a reasonable exercise of its rulemaking authority, including its requirement that ILECs unbundle network elements and offer CLECs the opportunity to pick and choose from an à la carte menu or platform of elements).

133 See, e.g., United States Telecom Association v. FCC, 290 F.3d 415 (D.C. Cir. 2002) (rejecting the FCC’s local exchange network unbundling requirements as insufficiently calibrated); U.S. Telecom Ass’n v. FCC, 359 F.3d 554, 578-85 (D.C. Cir. 2004) (again reversing the FCC for failing to create local requirements based on the specific level of local competition).
passed, and as many market entrants did not fully migrate from reselling incumbent carrier services to operating their own networks, reviewing courts became less deferential to the FCC’s precompetitive initiatives. Several appellate courts eventually rejected the FCC’s national pricing mandates based on the conclusions that Congress only required incumbent carriers to offer such rates in localities where the absence of such a financial catalyst would impair the onset and sustainability of competition.134

Reviewing courts grew weary with the ongoing role of the FCC, not only in the matter of whether and how a carrier must interconnect with a competitor but also the terms, conditions, and rates of such interconnection. The courts were persuaded that the FCC’s pricing methodology might bolster artificial competition, sustainable only because the FCC was all but guaranteeing a margin between the low rates incumbent carriers had to charge and the higher retail rates CLECs could charge customers.135 The courts also became persuaded that the FCC’s pricing methodology removed incentives for CLECs to

134 Appellate courts required the FCC to limit precompetitive initiatives to that perceived as minimally necessary to achieve success:

The purpose of the [1996 Telecommunications] Act is not to provide the widest possible unbundling, or to guarantee competitors access to ILEC network elements at the lowest price that government may lawfully mandate. Rather, its purpose is to stimulate competition-preferably genuine, facilities-based competition. Where competitors have access to necessary inputs at rates that allow competition not only to survive but to flourish, it is hard to see any need for the Commission to impose the costs of mandatory unbundling.

U.S. Telecom Ass’n, 359 F.3d at 576 (ordering elimination of all unbundling requirements for access to long distance and CMRS carriers).

135 Reviewing courts determined that the Telecommunications Act of 1996 sought to promote competition by allowing market entrants the temporary option of profitably reselling incumbent carrier services.

We also made clear that the Commission’s broad and analytically insubstantial concept of impairment failed to pursue the “balance” between the advantages of unbundling (in terms of fostering competition by different firms, even if they use the very same facilities) and its costs (in terms both of “spreading the disincentive to invest in innovation and creating complex issues of managing shared facilities”).

migrate from the resale of incumbent carrier facilities to making their own investments in new infrastructure. In response, the FCC exempted new technologies from any unbundling requirement and established dates for the elimination of interconnection and preferential access pricing for CLECs.

2. Elimination of Structural Safeguards

The FCC also eliminated structural separation rules. These rules required incumbent carriers with market power to create one or more separate subsidiaries to pursue markets that add value to and enhance basic leased lines. These requirements, articulated in the FCC's First and Second Computer Inquires, sought to establish a bright line between basic telecommunications services and the array of enhancements that evolved into what are now called information services. The Commission sought to create a level, competitive playing field between ventures unaffiliated with a carrier providing basic network access and an information-service affiliate of the basic network-providing carrier.

Reviewing courts determined that the FCC correctly refused to mandate sharing of competitively used facilities:

We therefore uphold the Commission's rules concerning hybrid loops, FTTH, and line sharing on the grounds that the decision not to unbundle these elements was reasonable, even in the face of some CLEC impairment, in light of evidence that unbundling would skew investment incentives in undesirable ways and that intermodal competition from cable ensures the persistence of substantial competition in broadband.

U.S. Telecom Ass'n, 359 F.3d at 585.


For background on the FCC's Computer Inquiries, see Robert Cannon, The Legacy of the Federal Communications Commission's Computer Inquiries, 55 Fed.
Structural separation prevented facilities-based incumbent carriers from offering preferential interconnection terms and conditions to corporate affiliates.\(^{141}\)

Carriers subject to the separate-subsidiary requirement and other safeguards that mandated functional separation between basic and enhanced services bristled at these requirements. They believed that the requirements were both unnecessary and costly.\(^{142}\) Over time, these carriers succeeded in persuading the FCC to abandon these safeguards despite never proving how such requirements resulted in lost efficiency and synergy.\(^{143}\) Bear in mind that the complaining carriers

\begin{footnotesize}
\footnotesize
\begin{enumerate}
\item Incumbent carriers framed the separate subsidiary requirement as unnecessary and inefficient.
\end{enumerate}
\end{footnotesize}

Parties supporting the removal of the structural separation requirements for the provision of enhanced services by AT & T and BOCs argue that, in the current telecommunications environment, the costs of those requirements outweigh their benefits. On the cost side of the equation, they contend that structural separation has imposed substantially greater burdens on the affected carriers, and ultimately on the public, than anticipated when we established those requirements in Computer II. In particular, a large number of parties assert that structural separation has deprived the public of innovative services that could be provided efficiently through AT & T’s and the BOCs’ extensive communications networks and thus made available to a large number of potential customers.


The FCC abandoned structural separation requirements based on carrier assertions of lost operational synergy and efficiency:

The following factors guide us toward replacing the Computer Inquiry obligations for wireline broadband Internet access service providers with a less regulatory framework: the increasing integration of innovative broadband technology into the existing wireline platform; the growth and development of entirely new broadband platforms; the flexibility to respond more rapidly and effectively to new consumer demands; and our expectation of the availability of alternative competitive
willingly created separate subsidiaries to provide “yellow page” directory advertising and wireless services,\textsuperscript{144} perhaps because such separateness accrued tax benefits and some degree of insulation from having to compensate the parent carrier for access to existing billing and database-management systems.

Even as the FCC eliminated local loop unbundling (LLU) and structural safeguards, national regulatory authorities (NRAs) in other nations have embraced them.\textsuperscript{145} Carriers facing such obligations have not experienced financial distress and the competitive environment has shown measureable broadband transmission to the currently required wireline broadband common carrier offerings. We believe our actions today will enhance each of these factors.

\textit{Wireline Broadband Classification Order, supra note 3, at 14,895.}

Deployment to consumers of these technologies then, at best, is delayed and, in many cases, may be avoided altogether. Broadband Internet access services are also not developing in ways that neatly fall within existing regulatory classifications or the current Computer Inquiry requirements (i.e., they cannot be easily separated into discrete information service and telecommunications service components). As a result, unlike cable modem providers or other broadband Internet access service competitors, wireline carriers must make either of two less-than-optimal choices when they seek to deploy advanced network equipment: either they must decide not to use all the equipment's capabilities, thereby reducing their operational efficiency; or they must defer deployment while the manufacturer re-engineers it to facilitate compliance with the Computer Inquiry rules, thereby creating unnecessary costs and service delays.

\textit{Id. at 14,887-88.}

\textsuperscript{144} For example AT&T divides itself into four subsidiaries, two of which provide wireless and directory publishing service:

\begin{itemize}
  \item AT&T has four main operating segments: wireless, wireline, advertising solutions, and other. The wireless segment consists of AT&T's subsidiary, AT&T Mobility, which provides wireless services to both business and consumer customers. This segment represents approximately 43 percent of 2009 total segment operating revenues.
  \item The advertising solutions segment includes AT&T's directory operations, which publish Yellow and White Pages directories and sell directory advertising and Internet-based advertising and search.
\end{itemize}

\textit{In re Applications of AT&T Inc. and Cellco Partnership D/B/A Verizon Wireless for Consent to Assign or Transfer Control of Licenses and Authorizations and Modify a Spectrum Leasing Arrangement, WT Docket No. 09-104, Memorandum Opinion and Order, 25 FCC Rcd. 8704, 8706 (2010).}

\textsuperscript{145} “[E]xperience both in the United Kingdom and elsewhere has indicated that, where access to the incumbents' networks has been allowed, it has provided a sound platform for the successful deployment of new services. Many of these new services—VoIP is an example—provide a significant source of competition.” Michael H Ryan, \textit{Promoting Network-Based Competition in UK Fixed-Line Markets: A Failed Policy,} 5 \textit{CONVERGENCE} 63, 72 (2009); Bob Bell, \textit{Broadband Deregulation—Similar Legislation, Different Results: A Comparative Look at the United States and the European Union,} 10 \textit{TUL. J. TECH. & INTELL. PROP.} 77, 94-98 (2007); \textit{Organization for Economic Co-Operation and Development, Developments in Local Loop Unbundling 5} (2000), available at http://www.oecd.org/dataoecd/25/24/6869228.pdf.
improvement. For example, Britain’s dominant carrier, British Telecom, split itself into two firms in 2006, one providing first- and last-kilometer access to telecommunications infrastructure and the other offering competitive services. The United Kingdom marketplace has become robustly competitive without harming incumbent British Telecom’s financial viability and stock attractiveness.

The nations of the European Union continue to embrace structural separation and LLU. Other nations with LLU requirements include Japan, Korea, New Zealand, Switzerland, South Africa, Australia, and Hong Kong.

Many national regulatory authorities endorse local loop unbundling as a vehicle for stimulating competition and expediting development of next generation broadband networks:

Korea has acquired world-class broadband internet services through a successful combination of industrial and competition policy. From the start, the Ministry of Information and Communication aggressively pursued industrial policy in the sector, but without stifling competition. It fostered competition in the market by lowering entry barriers and intervening to prevent KT from gaining too much of a competitive edge. It also adopted a local loop unbundling strategy to address concerns about unfair competition. The success of the government’s broadband internet strategy is apparent in the penetration ratio . . . . What can we learn from this Korean example? At an early stage of development, the government recognized the need for fundamental infrastructure. As an industrial policy measure, it required market entrants to install their own facilities while helping to create the market conditions that would make this affordable. Later, after sufficient facilities had been set up throughout Korea, the government changed tack and began to enforce an “essential facilities” doctrine that rested on local loop unbundling. This enabled new entrants to secure a foothold in an established market on a competitive basis. This demonstrates that under certain circumstances industrial policy can function alongside competition policy to achieve an ultimate economic policy goal, without producing undesirable side effects from a competition policy perspective.


See Network Unbundling, INFODEV, ICT REGULATION TOOLKIT 4.5.5 (Dec. 28, 2011), http://www.ictregulationtoolkit.org/en/Section.3426.html; Robert W. Crandall,
3. Courts Infer the Absence of a Common Carrier Duty to Deal

Appellate courts have determined that there is no antitrust remedy if the FCC has relaxed its oversight of carrier interconnection terms and conditions based on its expert assessment of marketplace competition. Put another way, if the FCC determines that the scope of competition is sufficient to trigger abandonment of regulatory safeguards, reviewing courts have no basis to second guess the Commission. In application, this means that reviewing courts have great reluctance to impose more burdensome safeguards than what the FCC, in its expert judgment, has deemed unnecessary.

Verizon v. Law Offices of Curtis V. Trinko resolved uncertainty about whether antitrust claims can exist based on the obligations imposed on ILECs by the Telecommunications Act of 1996 and, if so, whether individual customers have standing to assert such claims. The Supreme Court granted certiorari, limited to the question of whether the court of appeals erred in reversing the district court’s dismissal of the respondent’s antitrust claims.

The Court held that the “savings clause” contained in the ‘96 Act does not foreclose application of antitrust laws to ILEC behavior. However, the Court noted that such inclusion in the text of the Communications Act does not provide significantly greater scrutiny of or safeguards against anticompetitive practices. The relaxation of existing regulatory oversight performed by the FCC and state regulatory agencies does not create a mandate for new antitrust safeguards for courts to enforce:

But just as the 1996 Act preserves claims that satisfy existing antitrust standards, it does not create new claims that go beyond existing antitrust standards; that would be equally inconsistent with

---


“Section 601(b)(1) of the 1996 Act is an antitrust-specific saving clause providing that ‘nothing in this Act or the amendments made by this Act shall be construed to modify, impair, or supersede the applicability of any of the antitrust laws.’ Trinko, 540 U.S. at 406 (citing 110 Stat. 143, 47 U.S.C. § 152).
the saving clause’s mandate that nothing in the Act “modify, impair, or supersede the applicability” of the antitrust laws.\textsuperscript{153}

Having concluded that the ’96 Act does not foreclose antitrust cases, the Court easily rejected the applicability of the Sherman Act to a claim that Verizon discriminated against competitors when they sought access to individual, unbundled network services provided by Verizon:

We conclude that Verizon’s alleged insufficient assistance in the provision of service to rivals is not a recognized antitrust claim under this Court’s existing refusal-to-deal precedents. This conclusion would be unchanged even if we considered to be established law the “essential facilities” doctrine crafted by some lower courts, under which the Court of Appeals concluded respondent’s allegations might state a claim.\textsuperscript{154}

The Court concluded that both the FCC and state regulatory agencies can investigate claims that an ILEC had failed to comply with ’96 Act requirements and, in turn, can impose financial penalties, remediation measures, and additional reporting requirements for noncompliance:

Finally, we do not believe that traditional antitrust principles justify adding the present case to the few existing exceptions from the proposition that there is no duty to aid competitors. Antitrust analysis must always be attuned to the particular structure and circumstances of the industry at issue. Part of that attention to economic context is an awareness of the significance of regulation.\textsuperscript{155}

The Supreme Court’s deference to the FCC’s deregulatory campaign has gone so far as to allow an incumbent carrier to engage in predatory price squeezing, or to offer end users lower rates than what it charges competitors.\textsuperscript{156} In 2003, several ISPs filed suit against Pacific Bell Telephone Co. contending that the company attempted to monopolize the market for DSL broadband Internet access by creating a price squeeze where ISP competitors were obligated to pay a higher wholesale price than what Pacific Bell offered on a retail basis.\textsuperscript{157} Both the district court and the Ninth Circuit Court of Appeals agreed that the ISPs could present their price squeeze claim, despite the Supreme Court’s ruling in \textit{Trinko}.

\textsuperscript{153} Id. at 407.
\textsuperscript{154} Id. at 419.
\textsuperscript{155} Id. at 411.
\textsuperscript{157} Id. at 443-44.
The Supreme Court assumed that Pacific Bell had no antitrust duty to deal with any ISPs based on the FCC’s premise that ample facilities-based competition existed. Curiously, the Court did not mention that Pacific Bell could have avoided a unilateral duty to deal with ISPs based on the FCC’s conclusion that DSL, and presumably its component parts, constituted information services and not common-carrier-provided telecommunications services. But for a voluntary concession to secure the FCC’s approval of AT&T’s acquisition of another ILEC, the Court noted that Pacific Bell would not have a duty even to provide ISPs with wholesale service. The Court granted certiorari to resolve the narrow question of whether ISP plaintiffs can bring a price-squeeze claim under Section 2 of the Sherman Act when the defendant carrier has no antitrust-mandated duty to deal with the plaintiffs. The lower courts concluded that the Trinko precedent did not bar such a claim, but the Supreme Court reversed this holding.

On procedural grounds, the Court’s decision chided the ISP plaintiffs for changing the nature of their claim from a price squeeze to one characterizing Pacific Bell’s tactics as predatory pricing. On substantive grounds, the Court noted

158 “DSL now faces robust competition from cable companies and wireless and satellite services.” Id. at 443.
159 “As a condition for a recent merger, however, AT T remains bound by the mandatory interconnection requirements, and is obligated to provide wholesale DSL transport service to independent firms at a price no greater than the retail price of AT & T’s DSL service.” Id.
160 “We granted certiorari, 554 U.S. 916, . . . to resolve a conflict over whether a plaintiff can bring price-squeeze claims under § 2 of the Sherman Act when the defendant has no antitrust duty to deal with the plaintiff.” Id. at 435-46.
161 The Court supported the theoretical possibility that an antitrust claim could survive in a deregulated environment.

Our grant of certiorari was limited to the question whether price-squeeze claims are cognizable in the absence of an antitrust duty to deal. The Court of Appeals addressed only AT & T’s motion for judgment on the pleadings on the plaintiffs’ original complaint. For the reasons stated we hold that the price-squeeze claims set forth in that complaint are not cognizable under the Sherman Act.

Id. at 455-56.
162 The Court noted that the plaintiffs appeared to have shifted their claim from Pacific Bell engaging in a price squeeze to one alleging predatory pricing.

This case has assumed an unusual posture. The plaintiffs now assert that they agree with Judge Gould’s dissenting position that price-squeeze claims must meet the Brooke Group requirements for predatory pricing. They ask us to vacate the decision below in their favor and remand with instructions that they be given leave to amend their complaint to allege a Brooke Group claim. In other words, plaintiffs are no longer pleased with their initial theory of the case, and ask for a mulligan to try again under a different theory.
that a new emphasis on predatory pricing would have required
determination of whether the retail price was set below cost, a claim the ISPs did not make. The Court determined that the case did not become moot because of the change in economic and antitrust arguments. But the decision evidences great skepticism as to whether the ISPs had any basis for a claim. In the Court’s reasoning, the ISPs failed to make a claim that Pacific Bell’s retail DSL prices were predatory, and the ISPs also failed to refute the Court’s conclusion that Pacific Bell had no duty to deal with the ISPs (i.e., to provide wholesale service). The Court could apparently ignore the voluntary concession AT&T made that created a duty to deal because that concession may have triggered FCC oversight, but the concession could not change whether an antitrust duty to deal arose. The Court read the Trinko case as foreclosing any antitrust claim if no antitrust duty to deal exists.

The Court remanded the case to the district court to determine whether the ISP plaintiffs had a viable predatory pricing claim. The Court expressed the need for clear

Id. at 446.

The Court referenced Brooke Group Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209 (1993), which supports the inference that a predatory pricing claim can be established only with proof of below-cost pricing coupled with evidence that the defendant can subsequently recoup any lost profits. Linkline, 555 U.S. at 446-47.

The Court determined that a shift in framing what anticompetitive practice occurred did not by itself render the claim moot.

We do not think this case is moot. First, the parties continue to seek different relief. AT & T asks us to reverse the judgment of the Court of Appeals and remand with instructions to dismiss the complaint at issue. The plaintiffs ask that we vacate the judgment and remand with instructions that they be given leave to amend their complaint. The parties thus continue to be adverse not only in the litigation as a whole, but in the specific proceedings before this Court.

Linkline, 555 U.S. at 446.

“The challenge here focuses on retail prices—where there is no predatory pricing—and the terms of dealing—where there is no duty to deal.” Id. at 449. “If there is no duty to deal at the wholesale level and no predatory pricing at the retail level, then a firm is certainly not required to price both of these services in a manner that preserves its rivals’ margins.” Id. at 452.

“In this case, as in Trinko, the defendant has no antitrust duty to deal with its rivals at wholesale; any such duty arises only from FCC regulations, not from the Sherman Act.” Id. at 450.

The Court remanded the case for lower court determination whether a viable antitrust claim existed.

It is for the District Court on remand to consider whether the amended complaint states a claim upon which relief may be granted in light of the new pleading standard we articulated in Twombly, whether plaintiffs should be given leave to amend their complaint to bring a claim under Brooke Group, and such other matters properly before it.
antitrust rules and apparently viewed consumer access to low retail prices—predatory or not—as sufficient reason for courts to refrain from intervening. Remarkably, the Court did not seem troubled by the threat of all ISPs’ competitors exiting the market, an event that surely would enable the surviving incumbent carrier to raise rates: “For if AT&T can bankrupt the plaintiffs by refusing to deal altogether, the plaintiffs must demonstrate why the law prevents AT&T from putting them out of business by pricing them out of the market.”

This case evidences a strong reluctance on the part of the Supreme Court to support any review over the pricing strategies of carriers. Presumably the plaintiffs could have petitioned the FCC to review the broadband wholesale prices, but the Commission could have claimed that it had no jurisdiction to investigate because the DSL service at issue constituted an information service not subject to Title II pricing and nondiscrimination requirements. In light of the regulatory objectives contained in the ‘96 Act, which the Court deemed “much more ambitious than the antitrust laws,” more powerful safeguards against anticompetitive practices already exist. The Court opted not to second guess why the FCC refrained from using its lawful authority to remedy an obvious price squeeze.

C. Eliminating Cellular Radio Spectrum Caps

In 2003, the FCC eliminated a cap on the amount of spectrum a single wireless telecommunications carrier can acquire based on a determination of ample competition.

Id. at 456.

Id. at 456-57.

The holding in Nat’l Cable & Telecommunications Ass’n v. Brand X Internet Services, 545 U.S. 967 (2005), and the reversal of the FCC’s attempt to sanction Comcast for meddling with subscribers’ use of cable modem broadband links, Comcast Corp. v. FCC, 600 F.3d 642 (D.C. Cir. 2010), confirm that the FCC has no direct statutory mandate to regulate the terms and conditions by which a carrier offers information services including DSL.


Despite evidence to the contrary, the FCC concluded that a robustly competitive wireless telecommunication market existed:

Measures of market concentration in the record show a substantial continuing decline in concentration in most local CMRS [commercial mobile radio service] markets. We find that considerable entry has occurred and that meaningful competition is present, particularly given the presence of such earmarks of competition as falling prices, increasing output, and improving service quality and options. Specifically, concentration in CMRS markets, as measured by subscriber share, is falling.
Coupled with the Commission’s approval of each and every merger application it had received,\textsuperscript{172} the Commission all but guaranteed a concentrated marketplace for wireless services.\textsuperscript{173} In light of increasing reliance on wireless services to serve all consumers’ information, communications, and entertainment requirements, the FCC should have concluded that such consolidation would adversely affect the level of competition and the public interest. Advocates for merger approval have heralded efficiency gains from scale, the possibility of increased employment, spectrum scarcity, and extraordinary growth in demand for services.\textsuperscript{174} To these advocates, a spectrum cap would prevent a single carrier from satisfying demand and a proliferation of carriers presumably would not be able collectively to achieve such goals.

When it removed the spectrum cap, the FCC made summary assertions without using any serious or rigorous analysis about the consequences. The Commission never considered that removing a spectrum cap would eliminate an


\textit{Id.} at 22,690.


\textit{Id.} at 22,690.


Nonetheless, there are factors that moderate concern regarding the spectrum access barrier to entry. In particular, the need for direct access to spectrum is not absolute because carriers can compete in the provision of CMRS without direct access to spectrum through resale, or a mobile virtual network operator ("MVNO") arrangement.
ex ante safeguard that helps prevent anticompetitive consequences before harm has occurred. Arguably, ex ante safeguards are more essential in light of the Commission’s elimination of common-carrier duties to deal and case law that all but eliminates antitrust remedies. As a basis for comparison, other nations, including the United Kingdom, support spectrum caps in the mobile wireless marketplace. The UK’s telecommunications regulator has acknowledged that high barriers to entry and the potential for excessive concentration[176] justify spectrum caps:

We also propose to put in place safeguard caps to guard against longer term[177] risks to competition from very asymmetric holdings of spectrum. While we do not think that spectrum needs to be held equally for there to be effective competition or equality of opportunity to compete, we do think that there could be a risk if some national wholesalers held a very large share of mobile spectrum. While it is difficult to speculate about future possible developments, we consider it is possible that in the longer term there could be technological (e.g. beyond LTE) or market developments that meant that very asymmetric holdings of spectrum represented a risk to competition, especially for sub-1 GHz spectrum.

175 In light of the substantial deregulation that has occurred, the remaining regulatory oversight provides essential safeguards.

[A] sector regulator can introduce ex ante means, of which spectrum caps are one example, to help ensure that markets remain truly competitive. To the extent that policy makers believe they should have a portfolio of ex post and ex ante measures at their disposal to facilitate and ensure effective competition in markets for the sake of users, consumers, and overall welfare, then both a sector regulator in telecommunications and a Competition Authority have valuable roles to play.


176 The United Kingdom telecommunications regulatory authority considered it essential to impose spectrum caps on wireless carriers:

We consider that if we put in place no measures in the combined award to promote competition, there is a material risk of an outcome that would lead to lower competitive intensity in the provision of higher quality data services compared to competition in the wholesale market today, and compared to what might be possible. This is because we consider there is a material risk of only two or three national wholesalers emerging from the auction capable of providing higher quality data services in a profitable way. This is especially the case given that there are high barriers to entry to the national wholesale market, including the difficulty of obtaining access to suitable spectrum.


177 Id. at 49, ¶ 5.83.
Only recently, with 91.2 percent of the wireless market controlled by four national carriers, has the Commission begun to express doubts about whether concentration in the wireless marketplace generates sufficient competition.\footnote{With an eye toward providing better fact-based assessments of industry competitiveness, the FCC’s recent reports on the wireless marketplace use a more sophisticated and granular assessment:

[R]ather than reaching an overarching, industry-wide determination with respect to whether there is “effective competition,” the Report complies with the statutory requirement by providing a detailed analysis of the state of competition that seeks to identify areas where market conditions appear to be producing substantial consumer benefits and provides data that can form the basis for inquiries into whether policy levers could produce superior outcomes.} Previously the Commission expressed no concern that incumbent carriers would acquire the lion’s share of any newly available spectrum. For example, in the auctions for choice 700 MHz spectrum, which were made available when television broadcasters converted to digital transmissions, the two largest incumbent carriers, AT&T and Verizon, spent $16 billion of the $19.6 billion collected by the U.S. government.\footnote{AT&T and Verizon acquired the most spectrum and bid the most money in the FCC’s auction of 700 MHz wireless spectrum: According to an analysis by The Associated Press, the two telecom companies bid more than $16 billion, constituting the vast majority of the overall $19.6 billion that was bid in the FCC auction. With Verizon Wireless and AT&T dominating the auction so completely, hopes that the auction would allow for the creation of a new nationwide wireless service provider were dashed.}

In re Implementation of Section 6002(b) of the Omnibus Budget Reconciliation Act of 1993, 25 FCC Rcd. 11,407, 11,411, 11,407 (May 20, 2010) (fourteenth report) [hereinafter 14th Wireless Competition Report]. The Commission largely disputes its previous determinations of robust competition. For example, in 2006 the FCC reported that despite having approved a major merger, “[e]ven with one less nationwide mobile telephone carrier to choose from, U.S. consumers continue to benefit from robust competition in the CMRS marketplace.” In re Implementation of Section 6002(B) of the Omnibus Budget Reconciliation Act of 1993 Annual Report and Analysis of Competitive Market Conditions With Respect to Commercial Mobile Services, 21 FCC Rcd. 10,947, 11,029 (Sep. 29, 2006) (eleventh report). More recent Commission reports are less confident about the sufficiency of competition: “Over the past five years, concentration has increased in the provision of mobile wireless services. The two largest providers, AT&T and Verizon Wireless, have 60 percent of both subscribers and revenue, and continue to gain share (accounting for 12.3 million net additions in 2008 and 14.1 million during 2009).” Id. at 11,412. The Commission uses the Herfindahl-Hirschman Index to measure wireless industry concentration and reports that the current figure of 2848 exceeds the 1800 figure used by the Department of Justice to identify “highly concentrated” industries. See id. at 11,451-55; see also Annual Report and Analysis of Competitive Market Conditions With Respect to Mobile Wireless, Including Commercial Mobile Services, 26 FCC Rcd. 9664, 9697, Table 4, Service Provider Share of Subscribers and Revenues (Year-End 2009) (June 27, 2011) (fifteenth report).}
In light of the Commission’s favorable treatment of merger requests, AT&T Wireless applied to acquire T-Mobile. AT&T claimed the merger would help it abate a severe spectrum shortage and promote the company’s ability to provide wireless broadband services to rural locales on an accelerated basis. The company had sought to shift attention from the market-concentrating impact of the merger because acquiring T-Mobile’s 14 percent market share would boost AT&T’s share to over 40 percent, which, combined with Verizon’s share, would result in two companies controlling almost 80 percent of the market. AT&T sought to frame the merger as a means for the company to improve customer service and to compensate for delays in FCC regulatory reform, especially the Commission’s inability to make more spectrum available for wireless services.

AT&T’s now failed merger with T-Mobile constitutes an exception to a long list of approved mergers made possible by the FCC’s removal of a spectrum cap. Had the Commission retained the cap, the wireless marketplace may today have had more competition, innovation, and consumer choice. The four major carriers do not deviate significantly from a business model that offers subscribers a subsidized handset in exchange for a two-year service commitment and a hefty financial penalty for early termination of service. Wireless carriers charge rates that contribute to the recoupment of the handset subsidy and subscribers have few options for cheaper service if they activate...

---


180 See Acquisition of T-Mobile USA, Inc. by AT&T Inc. Description of Transaction, Public Interest Showing and Related Demonstrations, supra note 174.

181 “As we have shown, AT&T is facing severe capacity constraints in markets throughout the United States, and this merger is the surest and most efficient solution to those constraints.” Id. at 5.


184 See Applications of AT&T Inc. and Deutsche Telekom Ag for Consent to Assign or Transfer Control of Licenses and Authorizations, DA 11-1955 (order) (Nov. 29, 2011).

a used and unsubsidized handset. Had the spectrum cap remained in force, perhaps one or more carriers would have pursued a different business plan, maybe concentrating on data services and offering an open interface to content and software instead of the tightly controlled access erected by the four major carriers and handset manufacturers such as Apple.

U.S. wireless carriers claim they must aggressively compete by offering consumers world-class service in terms of monthly minutes of use, price, and innovation. On the positive side, the carriers correctly report that their rate plans offer large baskets of voice minutes and—at least until recently—unlimited data access plans. Additionally, carriers typically offer services that do not debit the monthly usage allotment when a subscriber calls another subscriber of the same carrier. On the other hand, U.S. wireless carriers offer services with nearly identical price points. Service terms do not stimulate competition and innovation even as these carriers generate some of the world’s highest margins and average revenue per user (ARPU). Provided subscribers do not deviate from relatively narrow, carrier-defined usage parameters, both carriers and customers can benefit. However, one can only speculate how much more robust, innovative, and dynamic the industry could have become had the FCC retained the spectrum cap.

Instead, the FCC overstates the positive benefits accruing from an increasingly concentrated industry. By using carrier-provided estimates of ARPU, average minutes of use, and other metrics, the FCC states that U.S. wireless carriers operate in a robustly competitive and innovative marketplace:

American consumers are the world’s wireless winners because today’s wireless ecosystem has evolved into a virtuous cycle of innovation and fierce competition. The U.S. regulatory approach has enabled American consumers to benefit from better value and more cutting-edge wireless products and services than consumers in other countries. Due to flexible, market-driven policies, the U.S. wireless industry is the most innovative and competitive. We are the example that other countries try to emulate.

---

186 U.S. wireless carriers claim they operate in a robustly competitive and innovative marketplace:

American consumers are the world’s wireless winners because today’s wireless ecosystem has evolved into a virtuous cycle of innovation and fierce competition. The U.S. regulatory approach has enabled American consumers to benefit from better value and more cutting-edge wireless products and services than consumers in other countries. Due to flexible, market-driven policies, the U.S. wireless industry is the most innovative and competitive. We are the example that other countries try to emulate.


187 See, e.g., Wireless Industry Innovation: We’re #1, CTIA—THE WIRELESS ASS’N BLOG (June 14, 2011), http://blog.ctia.org/2011/06/14/wireless-industry-innovation-were-1/ (compiling a list of industry leading accomplishments by U.S. wireless carriers).

188 “One of the main benefits of choosing an AT&T Mobile Phone Plan is unlimited calls to other AT&T wireless mobile users.” AT&T Wireless Phone Service, AT&T, http://www.att-services.net/att-wireless.html (last visited Jan. 8, 2012).

189 “The average monthly subscriber bill (ARPU) in the United States, at $51.54, is much higher than the Western European average of $33.45.” 14th Wireless Competition Report, supra note 178, at 11,619.
and cost-per-minute of service, the FCC has reported a mostly happy story about the U.S. wireless marketplace. Only recently has the Commission started to acknowledge the highly concentrated nature of the wireless marketplace. The Commission has generally dismissed any problems drawn from credible and frequently used measures of severe industry concentration. Factoring in Verizon’s $28 billion acquisition of Alltel, a company with a 5.2 percent market share, the Herfindahl-Hirschman Index (HHI) generated a concentration score of 2848, well above the 1800/2500 figure that triggers a Justice Department and Federal Trade Commission “highly concentrated” market finding. Apparently for wireless markets, other factors support a decision not to worry about the HHI score, including the availability of many subsidized handsets, non-price rivalry, and the $3.4 billion the four major wireless carriers spent on advertising in 2009. Additionally, the FCC has reported to Congress that CMRS carriers have at least 586 MHz of spectrum available. However, a close examination of the frequency bands identified by the Commission generates questions whether carriers can offer a

15th CMRS Competition Record, supra note 182, at 9679.

The Herfindahl-Hirschman Index (HHI), which is calculated by summing the squared market shares of all firms in any given market, is a commonly used measure of industry concentration. Antitrust authorities in the United States generally classify markets into three types: Unconcentrated (HHI < 1500), Moderately Concentrated (1500 < HHI < 2500), and Highly Concentrated (HHI > 2500).


In re Implementation of Section 6002(b) of the Omnibus Budget Reconciliation Act of 1993, 26 FCC Rcd. 9664, 9748 (June 27, 2011) (fifteenth report); see also 14th Wireless Competition Report, supra note 178, at 11,491-92.

14th Wireless Competition Report, supra note 178, at 11,566.
functionally equivalent service option based on propagational characteristics of the available spectrum and company business plans. For example, Clearwire, a company identified as providing a competitive alternative to CMRS, concentrates on data services to users and only offers VoIP service to users with wireless-modem-equipped, portable computers. The company does not provide a functional and competitive alternative to mobile services accessible via small handsets like those used by CMRS subscribers.194

III. CONCLUSION

National regulatory authorities such as the FCC typically have a statutory duty to serve the public interest and to recalibrate the nature and scope of their oversight when circumstances change. Technological innovations surely promote the possibility of more competition, but the countervailing trends of convergence create incentives for incumbents to diversify and serve new markets while expanding in size and scale. The cross-currents of potentially greater competition, but also consolidation of control by incumbents, should motivate NRAs to streamline regulations with caution and on an incremental basis. The FCC did not embrace this course of action and opted instead to make expansive deregulatory pronouncements based largely on nonempirical, overly optimistic assessments about the future sustainability of existing and future competition.

In the four case studies examined in this article, the FCC has identified problems necessitating its intervention or reassessment, but the Commission’s prior acts now prevent it from crafting quick and lawful solutions.195 When it opted to


195 The Commission rarely has the inclination or authority to undo a streamlined regulation in light of changed circumstances. A rare exception occurred when the Commission approved the merger of Sirius and XM satellite digital audio radio services (SDARS).

At that time, the Commission agreed that market forces produced by the robust competition between two SDARS competitors would ensure that listeners would receive noncommercial educational and public interest programming on the SDARS service. In the absence of such competitive forces post-merger, we find the potential harm to programming diversity greater than was the case in 1997.
apply unconditionally the information-services classification to all types of broadband Internet access, the FCC abdicated its authority even to resolve legitimate complaints of discriminatory and anticompetitive conduct. When it freed Title II–regulated common carriers of many core responsibilities—such as the duty to cooperate with competitors on fair terms, conditions, and prices—the Commission made it possible for reviewing courts to conclude that these carriers no longer had a duty to deal with each other subject to FCC oversight. Even a blatantly anticompetitive practice, such as offering retail rates below the wholesale rate offered to a competitor, does not trigger a judicial remedy because reviewing courts can defer to the FCC’s expert conclusion that marketplace competition would discipline carriers and offer readily available and cheaper alternatives to carriers’ engaging in price squeezes. When the FCC eliminated spectrum caps, it allowed incumbent carriers to achieve necessary scale, but also to benefit from extraordinarily high barriers to market entry all but guaranteeing a concentrated market, which is compounded by lax merger review.

The FCC has executed a strategy that favors incumbents best equipped to exploit streamlined or eliminated regulation for private gain. The competition identified or predicted by the Commission has failed to reach effective and sustainable levels. Rather than imposing so-called heavy-handed regulations, the FCC has removed regulatory safeguards that would require scrutiny of incumbents’ efforts to achieve market dominance, including tactics that might constitute unfair trade practices and violations of competition policies.

Only recently has the FCC changed its approach and recognized anticompetitive conduct and market concentration. The FCC has determined that it should resolve complaints regarding the allegedly anticompetitive practices of certain ISPs. The Commission no longer reports to Congress that the mobile wireless marketplace unconditionally operates with effective competition, or that Americans enjoy ubiquitous access to

---

\textit{In re Applications for Consent to the Transfer of Control of Licenses, 23 FCC Rcd. 12,348, 12,413 (Aug. 5, 2008) (memorandum opinion and order and report and order).}

\textsuperscript{196} The FCC no longer unconditionally concludes that the U.S. wireless marketplace evidences effective competition:

\text[We find that the mobile wireless ecosystem is sufficiently complex and multifaceted that it would not be meaningful to try to make a single, all-inclusive finding regarding effective competition that adequately encompasses the level of]
competitive broadband services. Additionally, the Commission has launched a reassessment of whether middle-mile telecommunications links between end users and carriers are priced at competitive levels. The Commission apparently now sees the need to impose duties to deal fairly and on reasonable terms and conditions even for carriers who claim regulatory streamlining exempts them from government oversight.

It remains to be seen whether and how the FCC can maneuver around all the consumer-protection tools it has abandoned. Already courts have rejected the Commission’s creative and novel invocations of ancillary jurisdiction in lieu of direct statutory authority. Had the Commission acted cautiously it would have lost the ability to make a big deregulatory pronouncement, but years later it would be in a position to act when needed.

Sadly, remedies for the FCC and the nation cannot arrive anytime soon, because Congress appears unable to reach consensus on necessary amendments to the Communications Act of 1934. Whether and how the FCC should regulate has become a contentious issue based largely on economic and political philosophy and not empirical evidence. The FCC needs a clear statutory basis to provide public-interest safeguards for consumers of information services and to make sensible and limited retreats from several deregulatory initiatives. Such reassessments would not signal a resumption of intrusive and potentially harmful regulation. Instead the FCC would have clear legislative authority to assess the current state of telecommunications and information-service markets and to make midcourse corrections in the scope of deregulation.

Absent new statutory authority the FCC will continue to struggle with no certainty whether an assertion of ancillary competition in the various interrelated segments, types of services, and vast geographic areas of the mobile wireless industry.

197 The FCC states that “that broadband is not being deployed in a reasonable and timely fashion to all Americans.” Inquiry Concerning the Deployment of Advanced Telecommunications Capability to All Americans in a Reasonable and Timely Fashion, and Possible Steps to Accelerate Such Deployment Pursuant to Section 706 of the Communications Act of 1996, as Amended by the Broadband Data Improvement Act, GN Docket No. 10-159, FCC 11-78, ¶ 1 (May 20, 2011) (seventh broadband progress report and order on reconsideration).

198 In re Connect America Fund, 26 FCC Rcd. 4554, 4676 (Feb. 9, 2011) (notice of proposed rulemaking and further notice of proposed rulemaking) (seeking comment on reasons for high middle mile costs and whether to use universal funding support to expand capacity and reduce price).

199 See generally Frieden, supra note 26, at 277-312.
jurisdiction will pass muster with a reviewing court. The Commission has achieved success in applying what it considers necessary consumer safeguards for VoIP, but similar efforts to curb ISP anticompetitive practices have failed. The Commission lacks clear guidance on the reach of its jurisdiction at the very time it needs to provide guidance to stakeholders, particularly ones that use the Internet to serve as a medium for a combination of voice, data, and video services.
NRSRO Nullification

WHY RATINGS REFORM MAY BE IN PERIL

Jason W. Parsont

One strategic move might be made by the Big Three [Moody's, S&P and Fitch] that would destabilize the status quo: they could decide to surrender their NRSRO status, and thereby avoid the more demanding provisions of the Dodd-Frank Act, which only apply to NRSROs. . . . When the burdens outweigh the benefits, it makes sense for them to abandon NRSRO status—if they can.

INTRODUCTION

In the words of Senator Christopher Dodd, Congress has “spent an inordinate amount of time on the rating agency question.” Scholars too have been deeply engaged in answering this question and so have market participants, the Securities and Exchange Commission (SEC), and others.

The rating-agency question asks whether there is a regulatory mechanism that can be adopted to encourage the credit rating agencies (CRAs)—including those that are nationally recognized statistical rating organizations (NRSROs)—to produce more accurate credit ratings on debt securities, or whether another assessment of credit risk exists

---

1 Careers-in-Law-Teaching Fellow, Columbia Law School, J.D. For helpful comments on earlier drafts, I am grateful to Frank Partnoy, Claire Hill, Paul Mahoney, Victor Goldberg, Dan Dunson, Ari Blaut, Jordan Abramson, Jonathan Schalit, Yuliya Guseva, Irene Ten Cate, Elizabeth Sepper, Aarthi Anand, Jennifer Sheridan and all other participants in the Associates’ and Fellows’ Workshop at Columbia Law School. I am also grateful to John Coffee and Charles Whitehead for their important advice and for the support of friends and family.


that could replace the need for credit ratings. Recently, Congress found that inaccurate credit ratings on a type of debt security—structured-finance products, such as subprime residential mortgage-backed securities and collateralized-debt obligations (CDOs)—significantly contributed to the mismanagement of risk by investors during the financial crisis of 2008. This, in turn, was a root cause of the crisis.

Given the close scrutiny that has been devoted to this question, including Congress’s recent legislative solution (collectively, Ratings Reform), this article’s purpose is not to propose a new answer, but to help refine the proposed answers. It tackles a fundamental problem that has the potential to cause any regulatory solution to the rating-agency question to crumble: “NRSRO Nullification,” which is the exercise of the NRSROs’ right to voluntarily withdraw from the regulatory regime. Such action would undermine, if not completely nullify, Ratings Reform and any future regulatory solutions. The concept of NRSRO Nullification also encompasses the exercise of the corollary right of the approximately seventy-six unregulated CRAs (non-NRSROs) to refrain from registering and thereby avoid the regulatory framework governing NRSROs.

Ratings Reform was a compromise between the views of two competing camps—the Free Market Camp and the Reform

---

5 See infra note 66.
6 See infra note 69.
10 See 15 U.S.C.A. § 78o-7(a) (setting forth the right of CRAs to voluntarily register, or refrain from registering, as NRSROs).
Camp. The Free Market Camp generally advocates deregulation and replacing credit ratings with market measures or a professional-judgment analysis as the solution to the rating-agency question. Its goal is to decrease reliance on credit ratings. The Reform Camp, on the other hand, generally advocates closer regulation—including purging destructive conflicts of interest—to improve the accuracy and reliability of credit ratings. While Ratings Reform may have been a compromise, the thrust of the combined legislation favors the Reform Camp’s solution due to its emphasis on using regulation and oversight to improve the quality of credit ratings, even as it seeks to deemphasize their importance.

The prospect of NRSRO Nullification is thus most problematic for those in the Reform Camp (including this author), because it would permit the NRSROs to foil Congress’s intent. In its most drastic form, all ten NRSROs could exercise their withdrawal right, which would precipitate de facto deregulation. It would be more likely, however, that only the seven smallest NRSROs would exercise this right, which would return the regulated club to only the Big Three—Moody’s, S&P and Fitch. Either result would deprive society of a regulatory mechanism to effectively promote accurate and reliable ratings. In the absence of regulation, there would be no way to collect comparative performance data, manage or prohibit conflicts of interest, or realign incentives so that accurate ratings—instead of issuer-friendly ratings—would serve the NRSROs’ best business interests. If such regulation only applied to the Big Three, then the comparative data set and controls would be confined to this group, and there would be significantly less incentive to compete over accuracy. While some in the Free Market Camp might welcome NRSRO Nullification, they too should be wary since market

---

11 See Coffee, supra note 1, at 246 (identifying the two camps of reformers); see also infra note 63.

12 See infra Part I.C.

13 See infra Part I.C.

14 The seven smallest NRSROs are A.M. Best Co. (A.M. Best), Dominion Bond Rating Service Limited (DBRS), Kroll Bond Rating Agency, Inc. (Kroll), Japan Credit Rating Agency (JCR), Rating and Investment Information (R&I), Egan-Jones Ratings Company (Egan-Jones), and Morningstar Credit Ratings (Morningstar). See infra notes 64, 102, 142.

15 The Big Three are Moody’s Investors Services, Inc. (Moody’s), Standard & Poor’s Financial Services LLC, a subsidiary of The McGraw-Hill Companies, Inc. (S&P), and Fitch, Inc., Fitch Ratings Ltd., and its subsidiaries (Fitch). See infra Part III.A.3.

participants will continue to rely on NRSRO ratings in the near
term irrespective of congressional action seeking to decrease
reliance on them.\textsuperscript{17}

NRSRO Nullification, moreover, is realistic for at least
some subset of the regulated club, because Ratings Reform has
substantially increased the burdens of regulation \textit{vis-à-vis} the
benefits.\textsuperscript{18} Indeed, this is precisely the reason some non-
NRSROs have refrained from registering and some NRSROs
have stopped rating structured-finance products or curtailed
plans to expand.\textsuperscript{19} Since the NRSROs are only subjected to
Ratings Reform if they consent, there is a fragile equilibrium
that must be maintained between the benefits and burdens of
NRSRO status.\textsuperscript{20} The resolution of the two most critical
unresolved aspects of Ratings Reform—the Franken Proposal\textsuperscript{21}
and the new standards of creditworthiness\textsuperscript{22}—will impact this
equilibrium. To the extent the final form of these items
continues to increase NRSRO-specific burdens without
providing NRSRO-specific benefits, such measures could tip
the balance toward NRSRO Nullification, especially for the
seven smallest NRSROs.\textsuperscript{23}

The final form of these items could also prevent NRSRO
Nullification if such items provide sufficient benefits to avoid
the tipping point. The Franken Proposal’s CRA Board,\textsuperscript{24}
which would act as both a “rater” of the NRSROs and an allocator of

(characterizing abandonment of the entire NRSRO regulatory regime as “an unrealistic
pipedream” that would nonetheless be a preferable solution).

\textsuperscript{17} See Claire A. Hill, Limits of Dodd-Frank’s Rating Agency Reform, 15 CHAP.
L. REV. 133, 144 (2011) (“[P]eople will continue to be influenced by the agencies . . . no
matter what the government does . . . [S]o it behooves government to make them
better if at all possible.”); see also 156 CONG. REC. S3955, 3956 (daily ed. May 19, 2010)
(statement of Sen. Al Franken) (“Here is the problem. Eliminating federally mandated
reliance on NRSRO credit ratings doesn’t change the fact that State laws, pension fund
policies, and other private market actors will still explicitly rely on NRSRO ratings.”).

\textsuperscript{18} For a description of how the benefits and burdens of NRSRO status have
changed over three different periods, see infra Part II.

\textsuperscript{19} See infra notes 229-30 and accompanying text.

\textsuperscript{20} Professor Coffee recently alluded to this problem by suggesting that the
Big Three might strategically abandon their NRSRO status to avoid the more
demanding provisions of Ratings Reform if the burdens were to outweigh the benefits.
See supra note 1 and accompanying text. I distinguish my claim by suggesting the
more likely problem is that the other seven NRSROs would surrender their NRSRO
status for precisely this reason while the Big Three would remain.

\textsuperscript{21} See infra Part III.A.1.

\textsuperscript{22} See infra Part III.A.2.

\textsuperscript{23} Based on today’s existing distinctions between NRSRO and non-NRSRO
status, it appears that the smallest seven would be more likely to opt out of the
regulatory regime than the Big Three. See infra Part III.A.3.

\textsuperscript{24} See infra notes 235-40 and accompanying text (describing the mechanics of
the CRA Board).
initial (but not secondary) rating assignments, is a potential solution for keeping the NRSROs voluntarily regulated while also maximizing accurate and reliable ratings. Through its rating function, the CRA Board could provide a reliable signal to the market about the best-performing NRSROs over time. If the new standards of creditworthiness permit investment fiduciaries to optionally rely on credit ratings, such a signal could significantly influence the preferences that such fiduciaries have regarding CRAs. Through its allocating function, the CRA Board could also reward good performance with increased market share. While both functions have the potential to create competition over accuracy, the allocating function may be premature because there is currently no definition of accuracy that the market accepts. My central recommendation, which would only put into place the Franken Proposal’s rating function, seeks to address this problem as part of a broader goal of finding an optimal mechanism to address both the rating-agency question and NRSRO Nullification together.

While this article also considers closing the voluntary registration loophole by having Congress adopt a mandatory registration requirement as an alternate way to prevent or reverse NRSRO Nullification, the article concludes that doing so is not necessary if my central recommendation is adopted.

There is also recent precedent demonstrating that the NRSROs would be willing to foil Congress’s intent. As part of Dodd-Frank, Congress sought to impose negligence exposure under Section 11 of the Securities Act of 1933 on the NRSROs for misleading ratings disclosed in a registration statement. The NRSROs had historically not been subject to negligence—only recklessness—because of a safe harbor, known as Rule

“Investment fiduciaries,” as used herein, means those persons (such as a fund’s board of directors and investment advisor) responsible for determining creditworthiness, whether or not under the new standards of creditworthiness, at their respective institutions (e.g., broker-dealers, funds, banks, insurance companies, etc.).

See infra Part IV.

See infra Part III.B.

See infra Part IV.C.

See Dodd-Frank § 939G (“Rule 436(g) . . . shall have no force or effect.”); see House-Senate Conference Committee Holds a Meeting on the Wall Street Reform and Consumer Protection Act, FIN. MARKETS REG. WIRE, June 15, 2010 (statement of Rep. Mary Jo Kilroy) (“Included in the House offer is a simple commonsense proposal that will help change this dynamic, a proposal that would nullify SEC Rule 436(g) and hold all CRAs accountable under Section 11 liability, a standard which already covers many experts in the financial world—accountants, auditors, lawyers, investment bankers and the directors, officers and executives of the issuer.”).
436(g), that shielded them from such exposure.\textsuperscript{30} Congress did not, however, repeal the NRSROs’ existing right to withhold consent to negligence exposure.\textsuperscript{31} Thus, the NRSROs collectively withheld consent shortly after Dodd-Frank passed and thereby nullified the intent of Congress.\textsuperscript{32} While it is troubling that the NRSROs escaped negligence exposure in this manner, NRSRO Nullification would pose a significantly larger problem because it would allow an escape not just from negligence exposure but from the entire regulatory regime. While it would be headline news if the Big Three left the regulated club, few words would likely be uttered if the other seven surrendered their NRSRO status.

This article proceeds in four parts. Part I provides background to the rating-agency question by discussing what credit ratings are and what it means for them to be inaccurate when made. It then describes the debate over the rating-agency question and the ultimate shape of Ratings Reform. Part II examines the question of why a CRA would want to be an NRSRO. It describes the benefits and burdens of being an NRSRO prior to 2006 and in the aftermath of Congress’s two recent attempts at Ratings Reform, the Credit Rating Agency Reform Act of 2006 (CRARA)\textsuperscript{33} and Dodd-Frank.\textsuperscript{34} It shows that being an NRSRO has become significantly less attractive over time. Part III identifies and discusses the financial and reputational implications that today’s most critical unresolved items—the Franken Proposal and the new standards of creditworthiness—will have on the NRSROs’ decision to withdraw from Ratings Reform. It then assesses the impact that today’s existing distinctions in concert with such unresolved items will have on this decision and the legal implications of NRSRO Nullification. This part also assesses the extent to which such items and their alternatives will promote accurate

\textsuperscript{30} See 17 C.F.R. § 230.436(g) (2009).
\textsuperscript{31} Securities Act of 1933 § 7(a), 15 U.S.C.A. § 77g(a) (West 2010) ("If . . . any person [e.g., an NRSRO] whose profession gives authority to a statement made by him [e.g., a credit rating], is named as having prepared or certified any part of the registration statement . . . . the written consent of such person shall be filed with the registration statement.").
\textsuperscript{32} See Ford Motor Credit Company LLC, SEC No-Action Letter, Item No. 1120 (Regulation AB) (Nov. 23, 2010) [hereinafter Ford No-Action Letter], available at http://www.sec.gov/divisions/corpfin/cf-noaction/2010/ford072210-1120.htm ("[T]he rating agencies indicated that they were not willing to provide their consent . . . .").
\textsuperscript{34} Dodd-Frank, Title IX, Subtitle C.
and reliable ratings. It concludes that certain proposals for resolving these items would be more likely to tip the balance toward NRSRO Nullification, while other proposals would be more likely to prevent this result, and that the smaller seven are more likely to opt out of Ratings Reform than the Big Three.

In Part IV, I present my central recommendation for resolving the Franken Proposal and new standards of creditworthiness in consonance with the dual goals of promoting accurate and reliable ratings and preventing NRSRO Nullification—adopt a refined version of the rating function suggested by the Franken Proposal, but not the allocating function. In addition, adopt the SEC’s current proposal to permit partial reliance by investment fiduciaries on any credit ratings under the new standards of creditworthiness subject to one additional requirement: investment fiduciaries seeking to rely on NRSROs that the rater deems good performers must certify agreement with the rater’s methodology for defining accuracy, while investment fiduciaries seeking to rely on non-NRSROs or NRSROs that the rater deems poor performers must publicly explain their disagreement with the rater’s methodology or show why certain non-NRSROs, when compared with NRSROs, produce ratings of equal or better quality. Under this proposal, it will not be necessary to close the voluntary registration loophole through a mandatory registration requirement.

I. THE RATING-AGENCY QUESTION

This part provides background to the rating-agency question by discussing what credit ratings are and what it means for them to be inaccurate when made. Then it describes the debate over the rating-agency question and Congress’s Ratings Reform solution.

A. What Are Credit Ratings?

Credit ratings are letter- and number-based assessments of risk that are “designed to measure and predict the probability of default, or the expected loss...for an individual debt obligation or for an obligor.” The ratings scales

typically range from triple-A (denoted by S&P and Fitch as “AAA” and by Moody’s as “Aaa”)—which represents the least risk of default—to C or D, which represents default or high vulnerability to default. Ratings that fall within the four highest categories assigned (typically “Baa3,” “BBB-” or higher) are known as “investment grade,” while lower ratings are known as non-investment grade (sometimes referred to as speculative, high-yield, or junk). The chart on the following page provides a comparison of the Big Three’s long-term credit ratings for individual debt obligations.

The CRAs characterize these symbols as forward-looking opinions and not as guarantees of future performance. Moody’s states that the CRAs “do not predict which specific bonds within a category are expected to default. Rather, credit ratings communicate that the higher the rating category, the lower the expected frequency of default.”

Since ratings are predictions about future risk and are not hard-and-fast facts, a credit rating should not be considered inaccurate when made simply because the rated security performed worse than expected. Good-faith predictions of future risk frequently prove wrong.

---


37 See, e.g., Letter from Michel Madelain, Chief Operating Officer, Moody’s Investor’s Serv., to Elizabeth M. Murphy, Sec’y, Sec. & Exch. Comm’n 5 (Dec. 14, 2009), available at http://www.sec.gov/comments/s7-25-09/s72509-5.pdf (“[R]atings are inherently and completely forward-looking, rather than backward-looking, in nature.”).

38 Id. at 3. Moody’s compares its ratings to actuarial predictions made by life insurance companies:

[An actuary would predict that, in the next five years, a 25 year-old non-smoker will be less likely to die than an 80 year-old smoker; nonetheless, in the next five years, some 25 year-old non-smokers will die, while some 80 year-old smokers will survive. Similarly, a rating analyst is predicting that a Ba1 bond will be more likely to default than a Aaa bond; nonetheless, some Aaa bonds will default, while most Ba1 bonds will not default.]

Id. at 4.
<table>
<thead>
<tr>
<th>Moody's</th>
<th>S&amp;P</th>
<th>Fitch</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Investment Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aaa</td>
<td>AAA</td>
<td>AAA</td>
</tr>
<tr>
<td>Aa</td>
<td>AA</td>
<td>AA</td>
</tr>
<tr>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>Baa</td>
<td>BBB</td>
<td>BBB</td>
</tr>
<tr>
<td>Baa3</td>
<td>BBB-</td>
<td>BBB-</td>
</tr>
<tr>
<td><strong>Non-Investment Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ba1</td>
<td>BB+</td>
<td>BB+</td>
</tr>
<tr>
<td>Ba</td>
<td>BB</td>
<td>BB</td>
</tr>
<tr>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Caa</td>
<td>CCC</td>
<td>CCC</td>
</tr>
<tr>
<td>Ca</td>
<td>CC</td>
<td>CC</td>
</tr>
<tr>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td></td>
</tr>
</tbody>
</table>

39 See Moody’s Investors Serv., supra note 36, at 4 (“Moody’s appends numerical modifiers 1, 2, and 3 to each generic rating classification from Aa through Caa. The modifier 1 indicates that the obligation ranks in the higher end of its generic rating category; the modifier 2 indicates a mid-range ranking; and the modifier 3 indicates a ranking in the lower end of that generic rating category.”).

40 See Standard & Poor’s, supra note 36 (“Ratings from ‘AA’ to ‘CCC’ may be modified by the addition of a plus (+) or minus (-) sign to show relative standing within the major rating categories.”).

41 See Fitch Ratings, supra note 36 (“The modifiers ‘+’ or ‘-’ may be appended to a rating to denote relative status within major rating categories.”).
B. How Can a Credit Rating Be Inaccurate When Made?

Since any retrospective analysis comparing actual defaults against before-the-fact predictions will show some level of inaccuracy, one must distinguish between those ratings that were inaccurate when made as a result of neglect or bad faith and those ratings that were made in good faith at the outset but only proved inaccurate as a result of naturally occurring market forces. The available evidence identifies at least five different scenarios when neglect or bad faith in the initial production of ratings appears to have contributed to inaccurate ratings.

The first scenario involves the failure of rating analysts to comply with or adhere to available procedures and methodologies for producing credit ratings. For example, two financial economists reported that the CRAs were regularly making subjective adjustments in certain cases rather than following consistent policies. They reported that “only 1.3% of AAA CDOs closed between January 1997 and March 2007 met the rating agency’s reported AAA default standard,” with the rest falling short. They concluded that “the AAA tranches should have been rated ‘as approximately BBB’ and that if the AAA tranches in their sample of 916 CDOs were so downgraded to BBB, the total overvaluation ‘cumulates to $86.2 billion in cost to investors.’” This example suggests that there was a systematic failure to comply with objective procedures and methodologies for producing ratings. Many other examples have also been alleged where the CRAs failed to comply with their own models or failed to implement existing models that would have produced more accurate, but less issuer-friendly, ratings. As recently as September 2011,

---

---
the SEC found that “[o]ne of the larger NRSROs reported that it had failed to follow its methodology for rating certain asset-backed securities” and “[a]ll of the NRSROs failed to follow their ratings procedures in some instances.”

The second scenario involves the use of identical rating symbols to express grossly different risks. In spite of years of statistical evidence showing that different securities such as corporate bonds and CDOs had grossly different default rates across the same risk symbols, the CRAs did not seek to alert investors to these differences in risk by using different symbols. For example, two financial economists reported “that the five year cumulative default rate on corporate bonds receiving a ‘Baa’ rating from Moody’s between 1983 and 2005 was only 2.2%, but the same five year cumulative rate between 1994 and 2005 on CDOs with a Baa rating was 24%—a more than ten to one disparity!” Similarly, during the financial crisis of 2008, “[r]ating agencies gave triple-A ratings to 75% of the $3.2 trillion subprime mortgages that lost sizable value only months after the ratings were made.” Such triple-A ratings signaled to investors a degree of safety commensurate with U.S. treasury bonds. It is troubling that the CRAs used the symbol associated with the benchmark of safe securities to represent securities that proved to be, and perhaps should have always been known to have been, significantly more risky.

The third and most often cited scenario involves the existence of two inherent conflicts of interests: the issuer-pays conflict and the ratings-shopping problem. The issuer-pays
conflict describes the business model where the issuer, rather than the investor, pays the rating agency for a rating. The ratings-shopping problem occurs where issuers shop privately (as opposed to final) ratings and award business to the rating agency willing to give it the highest, rather than the most accurate, final rating.\textsuperscript{52} Ample evidence supports the notion that such inherent conflicts of interest put unmanaged pressure on the CRAs to inflate ratings in order to maintain their market share. For example, e-mails uncovered by a Senate committee reveal that one Moody’s managing director admitted that its behavior in terms of handing out triple-A ratings for mortgage-backed securities made it “either incompetent at credit analysis, or like we sold our soul to the devil for revenue.”\textsuperscript{53} An S&P official said that its mortgage team had “become so beholden to their top issuers for revenue [that] they ha[d] developed a kind of Stockholm syndrome which they mistakenly tag as customer value creation.”\textsuperscript{54} One UBS banker warned S&P as follows: “Heard your ratings could be 5 notches back of mo[o]dy’s equivalent. This is going to kill your [residential business]. It may force us [UBS] to do moodyfitch only cds [sic].”\textsuperscript{55}

The fourth scenario involves the failure of the CRAs or their proxies to do due diligence.\textsuperscript{56} Most cogently, Professor


\textsuperscript{52} There is also a related conflict of interest associated with client concentration. Issuances of asset-backed securities equaled and then subsequently exceeded those of corporate bonds beginning in 2002 and “the top six underwriters [of asset-backed securities] controlled over 50 percent of the mortgage-backed securities underwriting market in 2007, and the top eleven underwriters each had more than 5 percent of the market and in total controlled roughly 80 percent of this very lucrative market”; thus it becomes clear that a rating agency’s market share could be significantly diminished if a small concentration of clients became unhappy and sought to take business elsewhere. \textit{See Enhancing Investor Protection and the Regulation of Securities Markets: Hearing Before the S. Comm. on Banking, Housing, and Urban Affairs, 111th Cong. 55-56 (2009) [hereinafter Enhancing Investor Protection]} (written statement of John C. Coffee, Jr.). By contrast, for corporate bonds, whose ratings have proven far more accurate, “no one client accounted for more than 1% of their business.” \textit{See Coffee, supra} note 1, at 237.


\textsuperscript{54} \textit{Id.}


\textsuperscript{56} \textit{See, e.g.,} John Patrick Hunt, \textit{Credit Rating Agencies and the “Worldwide Credit Crisis”: The Limits of Reputation, the Insufficiency of Reform, and a Proposal for Improvement}, 2009\textit{ Colum. Bus. L. Rev.} 109, 170-71 (“Fitch explains that it ‘does not audit or verify . . . or . . . perform any other kind of investigative diligence into the accuracy . . . or completeness’ of the information it receives.” (citation omitted)); Darcy, \textit{supra} note 51, at 617 (“CRAs do not perform their own due diligence.”); Paul
Coffee pointed out that CRAs were alone among financial gatekeepers in that they did not conduct “factual verification with respect to the information on which their valuation models rely.” He explains that the problem is “that no valuation model, however well designed, can outperform its informational inputs; hence, use of unverified data results in the well-known ‘GIGO Effect’—Garbage In, Garbage Out.”

While the CRAs used to rely on due diligence firms to test the creditworthiness of securitized mortgages, this practice mostly ended after the year 2000 with the CRAs’ tacit approval.

The fifth scenario involves “less-than-thorough business practices.” According to the SEC, “[w]hen the firms didn’t have enough staff to do the job right, they often cut corners.” A senior analytical manager at one of the Big Three stated in an e-mail that “[w]e do not have the resources to support what we are doing now.” As the CRAs began expanding their coverage of issuances and began taking on more complex instruments, resources to produce each rating with integrity declined. At present, there is a large disparity in staffing resources when comparing the Big Three against the smaller seven NRSROs. While the Big Three each employ over one thousand credit analysts and supervisors, the smaller seven each employ considerably fewer.

---


57 See Coffee, supra note 1, at 244.
58 Id.
59 Id. at 241 (“Factual verification of the creditworthiness of securitized mortgages largely disappeared after 2000, as investment banks and deal arrangers ceased to pay for such activities, and the CRAs did not insist on their continuation.”).
61 Id. at 595 (citation omitted).
63 See id. (“Rating agencies also began rating substantially greater numbers of issuers and increasingly complex instruments. But the resources expended per rating declined. As they expanded ratings to cover large numbers of structured finance products, including tranches of various collateralized debt obligations, some NRSROs neglected to divert resources to update rating models and methodologies or recruit additional staff needed to ensure quality.”).
64 A.M. Best (120), DBRS (95), R&I (78), JCR (57), Morningstar (24), Kroll (13), and Egan-Jones (5). See SEC SEPT. 2011 SUMMARY REPORT, supra note 47, at 8.
C. The Debate over the Rating-Agency Question

Billions, and sometimes trillions, of dollars have been alleged to have been lost during the financial crisis of 2008.\(^65\) Congress’s account holds that inaccurate credit ratings misled investors with respect to the level of risk they were assuming.\(^66\) Other accounts hold that sophisticated investors over-relied on credit ratings and share equal blame for making the same mistakes as the CRAs.\(^67\) Substantially all agree, however, that inaccurate and unreliable credit ratings on structured finance products\(^68\) were a root cause of the financial crisis of 2008.\(^69\)

\(^65\) See 156 Cong. Rec. S3664, 3673 (daily ed. May 13, 2010) (statement of Sen. Al Franken) (“This conflict of interest has cost American investors and pensioners billions of dollars because supposedly risk-free investments have failed or been downgraded to junk status.”); see id. at S3675 (statement of Sen. George Lemieux) (“We have a chance to address the issue of the rating agencies, because, but for their failure to do their job, we may not have had this debacle that destroyed, as some estimate, $600 trillion worth of wealth.”).

\(^66\) Congress found that,

[i]n the recent financial crisis, the ratings on structured financial products have proven to be inaccurate. This inaccuracy contributed significantly to the mismanagement of risks by financial institutions and investors, which in turn adversely impacted the health of the economy in the United States and around the world. Such inaccuracy necessitates increased accountability on the part of the credit rating agencies.


\(^67\) See Hill, supra note 46, at 598 (explaining that both market participants and rating agencies “drank the Kool-Aid”); see Frank Partnoy, Overdependence on Credit Ratings Was a Primary Cause of the Crisis 9 (San Diego Legal Studies Paper No. 09-015, 2009), available at http://ssrn.com/abstract=1430653 (“Without overreliance on ratings, investors would more likely have looked through the complexity of CDO and SIV transactions to the underlying mortgage-backed securities, and prices would have more accurately reflected market estimates of default probability, recovery, and correlation.”).

\(^68\) This phenomenon was limited to asset-backed securities (i.e., structured finance products) and not other types of debt securities. See, e.g., Hunt, supra note 56, at 170-71 (“Official reports on the crisis did not indicate that agencies did a poor job in the corporate segment. Indeed, regulatory authorities’ studies drew a fundamental distinction between agencies’ traditional and structured-finance ratings and criticized only the latter.”); Coffee, supra note 1, at 236 (“The failure of the CRAs was largely limited to structured financial products. Similar problems have not characterized the ratings of corporate bonds.”).

\(^69\) See Partnoy, supra note 62, at 13 (“[I]naccurate and unreasonable credit ratings from NRSROs were a primary cause of the recent crisis. . . .”); see also Coffee, supra note 1, at 232 (“Few disinterested observers doubt that inflated credit ratings and conflict-ridden rating processes played a significant role in exacerbating the 2008 financial crisis.”); 156 Cong. Rec. S3965, 3977 (daily ed. May 19, 2010) (statement of Sen. Christopher Dodd) (“I agree with my colleagues that erroneous credit ratings on asset backed securities played a central role in the financial crisis and that we need to improve the regulation of credit ratings.”).
The critics generally divide into two camps: the Free Market Camp and the Reform Camp.\footnote{See Coffee, supra note 1, at 231 ("[R]eformers divide into two basic camps: (1) those who see the ‘issuer pays’ model of the major credit ratings firms as the fundamental cause of inflated ratings, and (2) those who view the licensing power given to credit ratings agencies by regulatory rules requiring an investment grade rating from an NRSRO rating agency as creating a de facto monopoly that precludes competition.").} A fundamental difference between the two camps concerns their views over the informational value of ratings, which influences how the camps might ask and answer the rating-agency question.

The Free Market Camp believes that ratings have little informational value and that better indicators of risk are generally provided by market measures such as credit spreads and credit-default-swap spreads.\footnote{A credit spread for a bond is the difference between a bond’s yield and the yield of a comparable risk-free bond. Higher yields (and therefore wider credit spreads) reflect the market’s view of the relative riskiness of such bond. Credit default swaps are effectively insurance policies that investors can buy to protect themselves against an entity’s default. Credit-default-swap spreads are equal to the premium on such protection. For higher credit risks, the premium would be higher (and therefore the spread would be wider) and vice versa. See Frank Partnoy, The Siskel and Ebert of Financial Markets?: Two Thumbs Down for the Credit Rating Agencies, 77 WASH. U. L.Q. 619, 624 (1999) [hereinafter Siskel & Ebert] (credit spreads); see also Mark J. Flannery, Joel F. Houston & Frank Partnoy, Credit Default Swap Spreads as Viable Substitutes for Credit Ratings, 158 U. PA. L. REV. 2085, 2088 (2009) [hereinafter Credit Default Swap Spreads] (credit-default-swap spreads).} One study found that credit-default-swap spreads incorporate new risk information more quickly than credit ratings.\footnote{See generally Credit Default Swap Spreads, supra note 71.} While they acknowledge that investors frequently rely on ratings to assess the riskiness of their investments, they argue that this type of reliance is misplaced. For these reasons, the Free Market Camp would ask, do assessments of credit risk exist that are more accurate and reliable than credit ratings?

The Reform Camp, on the other hand, believes that ratings still provide valuable information about risk and that such market measures do not provide a superior substitute, especially in the case of complicated and obscure structured finance products. They believe that CRAs are in a better position than any other market actors to assess risk and that, while blind reliance on ratings should be discouraged, measured reliance is justified from an efficiency standpoint as long as regulatory measures ensure that the incentives to produce accurate and reliable ratings are properly aligned with the CRAs’ interests. For these reasons, the Reform Camp would ask, how can CRAs, including those who are NRSROs,
be encouraged through regulation to produce more accurate and reliable credit ratings?

Professor Partnoy, the leading advocate for the Free Market Camp, historically favored replacing references to NRSRO credit ratings in statutes and regulations with market-based measures such as credit spreads or credit-default-swap spreads. To ease concerns about volatility, he suggested using thirty-to-ninety day rolling averages of such spreads. He has recently advocated, in light of SEC proposals, that a professional judgment analysis taking into account multiple factors including such market measures and credit ratings could also be a viable substitute for sole reliance on NRSRO credit ratings. He has long argued that references to NRSRO credit ratings in statutes and regulations transformed the NRSROs from providers of risk information to sellers of “regulatory licenses.” Regulatory licenses are “the valuable property rights associated with the ability of a private entity [such as an NRSRO], rather than the regulator, to determine the substantive effect of legal rules.” Since NRSROs were given this quasi-governmental power in a wide variety of contexts, they were effectively guaranteed continued business even if they performed poorly. This partially explains the phenomenon of widespread inaccurate credit ratings during the recent financial crisis and previous crises. If statutes and regulations were to require a different proxy for risk than NRSRO credit ratings, this would remove the power of the NRSROs to grant regulatory

---

73 See Siskel & Ebert, supra note 71, at 704 (recommending credit spreads in place of credit ratings).
74 See generally Credit Default Swap Spreads, supra note 71.
75 See Partnoy, supra note 62, at 17 (“Investors concerned about the volatility of market prices could use 30-day or 90-day rolling averages.”). In addition, to ease concerns that market-based credit spreads would only be available after a market for the bond has arisen, he has suggested “pre-issuance estimates of credit spreads (i.e., ‘price talk’), in much the same way investors now rely on pre-issuance estimates of credit ratings, which are not issued until the bonds are issued (credit spreads are available at the same time).” See Siskel & Ebert, supra note 71, at 706 n.391.
76 See, e.g., Partnoy, supra note 67, at 16 (advocating professional judgment analysis to replace sole reliance on credit ratings that would include market measures, such as credit spreads and swap spreads, as one factor of the analysis).
77 See Partnoy, supra note 62, at 2. See generally Siskel & Ebert, supra note 71. For further discussion of the NRSRO’s “power to license” over time see infra Parts II.A.2, II.B.2, and II.C.2.
78 Siskel & Ebert, supra note 71, at 623.
79 See Partnoy, supra note 62, at 2 (“A regulatory license is a key that unlocks the financial markets. Credit rating agencies profit from providing ratings that unlock access to the markets, regardless of the accuracy of their ratings.”). Professor Partnoy also criticizes, and makes suggestions to remedy, the lack of accountability of the credit rating agencies and the lack of competition. See id. at 14-18.
licenses and might cause the NRSROs to once again depend only on their reputations to maintain business."

Professor Coffee, the leading advocate for the Reform Camp, disputes this view. He argues that credit ratings cannot simply be replaced by credit spreads or other market measures, and it is unrealistic to expect even sophisticated institutional investors to do their own credit analysis in the context of complex and opaque debt instruments such as CDOs. He compares such "do-it-yourself financial analysis" to "do-it-yourself brain surgery." Under this view, NRSRO-dependent regulatory licenses may actually improve the quality of ratings to the extent they prevent investors from relying on unestablished ("fly-by-night") CRAs instead of those with an established track record. Regulatory licenses are viewed as much less problematic since they do not alone explain the dominance of the Big Three NRSROs. Such NRSROs were dominant before NRSRO status was introduced in the 1970s and have remained dominant since the expansion of the NRSRO club from three to ten rating agencies. Professor Coffee argues that the issuer-pays conflict of interest, rather than the relaxation of high standards from the exploitation of regulatory licenses, is the main impediment to accurate ratings. He also touches upon the independent ratings-shopping problem that some accounts treat as equally problematic. Professor Coffee argues that the solution to the issuer-pays conflict must "either (1) divorce issuer payment of the CRA from issuer selection of the CRA, or (2) encourage (and implicitly subsidize) an alternative 'subscriber pays' market for ratings." Any solution adopted must realign the incentives of the CRAs so that they are rewarded for accuracy instead of issuer-friendly ratings: "if the incentives

---

80 See Siskel & Ebert, supra note 71, at 682 ("The regulatory license view is quite simple. Absent regulation incorporating ratings, the regulatory license view agrees with the reputational capital view: rating agencies sell information and survive based on their ability to accumulate and retain reputational capital.").

81 See Coffee, supra note 1, at 233.

82 Id.

83 See id. at 248. ("If licensing power alone could explain the dominance of the Big Three, then the newer members of the SEC’s NRSRO club should be sharing in a collective oligopoly.").

84 Id. at 232-34.

85 Id. at 255; see also, e.g., 156 CONG. REC. S3955, 3956 (daily ed. May 19, 2010) (statement of Sen. Al Franken) ("Right now, credit rating agencies have incentives to hand out top AAA ratings to every product because they need to maintain their business. If they hand out low ratings, issuers of financial products can go shop around for a higher rating from a different rating agency.").

86 Coffee, supra note 1, at 234.
remain poorly aligned, regulatory oversight alone is unlikely to ensure ratings accuracy."

D. Congress’s Solution

While the debate over the rating-agency question has been ongoing for over a decade, reform has often “seemed stuck in an ever-repeating cycle of futility.” This came to an end in 2006 when Congress adopted the first leg of Ratings Reform, known as CRARA. While CRARA contained certain free market aspects such the NRSROs’ voluntary withdrawal right, it did not heed the advice of the Free Market Camp to rollback regulatory licenses. The clear purpose of CRARA was to “improve ratings quality.” It sought to do so through a narrowly tailored regulatory regime.

Unfortunately, much of the damage that CRARA sought to prevent had already been done by the time it became effective on June 26, 2007, and new problems were observed in the financial crisis of 2008. As a result, Congress passed the second leg of Ratings Reform in July 2010 as part of the Dodd-Frank Act. While Dodd-Frank builds on CRARA’s Reform Camp architecture, it is a clearer compromise than CRARA between the views of the two camps. It seeks to achieve the Free Market Camp’s goal of decreasing reliance on ratings, while at the same time pursuing measures to encourage ratings to be more accurate and reliable. It seeks to do the former by eliminating substantially all NRSRO regulatory licenses on the federal level and replacing them with new standards of creditworthiness by July 2012. It seeks to achieve the latter through a multifaceted approach of heightened regulation, accountability, and transparency. The crown jewel of the Reform Camp approach, which has not yet been (and

---

87 Id. at 278.
88 Hill, supra note 17, at 133.
89 See supra notes 8, 33.
91 See supra note 66.
92 See Hill, supra note 17, at 143 (“The reform has two important goals. One is to decrease reliance on ratings . . . . The other goal is to improve the quality of ratings.”).
94 See id. §§ 932-938.
may not be) adopted, is known as the Franken Proposal.\textsuperscript{95} It is intended to address the issuer-pays conflict and the ratings-shopping problem by creating a centralized board (the CRA Board) controlled by investors to assign NRSROs to issuers with respect to initial, but not secondary, ratings in order to divorce issuer selection from issuer payment. The CRA Board also would serve as a “rater” of at least some subset of the NRSROs and would increase or decrease rating assignments based on past performance. The SEC must adopt the Franken Proposal or one of its alternatives after delivering its findings and recommendations to Congress no later than July 2012.\textsuperscript{96}

II. WHY BE AN NRSRO?

This part examines the question of why a CRA would want to be an NRSRO. It describes the benefits and burdens of being an NRSRO prior to 2006 and in the aftermath of CRARA and Dodd-Frank. It shows that being an NRSRO has become significantly less attractive over time.

A. Pre-2006

Before 2006, the CRAs placed a high value on NRSRO status primarily because it afforded financial and reputational benefits without any significant burdens. It conferred—among other advantages—minimal competition because the club had few members, captive business because of regulatory licenses, and a shield against negligence exposure. These, in turn, brought increased visibility and sent a signal to market that NRSRO ratings had the government’s seal of approval as proxies for safety and quality. Investment adviser regulation under the Investment Advisers Act of 1940 (the Advisers Act),\textsuperscript{97} which also conferred some reputational benefits, was the only burden of NRSRO status, and it was minimal.

\textsuperscript{95} See Solicitation of Comment to Assist in Study on Assigned Credit Ratings, Exchange Act Release No. 64,456, at 49 (May 10, 2011) [hereinafter Solicitation of Comment], available at http://www.sec.gov/rules/other/2011/34-64456.pdf (includes a copy of the Franken Proposal’s proposed provisions in Section 15E(w) of the Securities Exchange Act of 1934 as an appendix [hereinafter the Franken Proposal], as those provisions would have been added by Section 939D of H.R. 4173 (111th Congress), as passed by the Senate on May 20, 2010).

\textsuperscript{96} See Dodd-Frank § 939F(d).

\textsuperscript{97} 15 U.S.C.A. § 80b-6 (West, Westlaw through P.L. 112-71 (excluding P.L. 112-55 and 112-56)).
1. The Exclusive Club

From the early 1900s, when the credit rating business first took shape, through 1975, when NRSRO status was first introduced by the SEC, the Big Three (or their predecessors) were the primary issuers of credit ratings on debt securities. In 1975, the SEC recognized the Big Three as the original NRSROs. While four additional CRAs were later admitted, each of them was subsequently merged with or acquired by the Big Three such that the club only comprised its original three members as of January 2003.

While two additional CRAs joined the club by March 2005, the recognition of new NRSROs was tightly controlled by the SEC. “The single most important factor in the SEC staff’s assessment . . . [was] whether the rating agency [was] ‘nationally recognized’ in the United States as an issuer of credible and reliable ratings by the predominant users of securities ratings.” As a result, new entrants faced a chicken-and-egg dilemma in becoming NRSROs: “it [was] nearly impossible to obtain clients [without] a track record for reliable ratings, yet such a track record [was] difficult to generate unless one first ha[d] clients.”

While the SEC’s approach created a high barrier to entry for those outside of the club, it made membership in the club more valuable through exclusivity since fewer members meant less competition. This was especially valuable in light of a market norm by which issuers regularly sought two and sometimes three ratings on an issuance of debt. By one estimate, Moody’s and

---

98 See SEC REPORT OF 2003, supra note 9, at 5. For a history of the NRSROs during the twentieth century, see generally Siskel & Ebert, supra note 71.

99 See SEC REPORT OF 2003, supra note 9, at 8-9.

100 These firms were (1) Duff and Phelps, Inc.; (2) McCarthy Crisanti & Maffei, Inc.; (3) IBCA Limited and its subsidiary, IBCA, Inc.; and (4) Thomson BankWatch, Inc. Id. at 9.

101 Id. at 5.


103 See SEC REPORT OF 2003, supra note 9, at 9-10.

104 Id. at 9.

105 Coffee, supra note 1, at 234.

106 See Hill, supra note 46, at 590.
S&P held “nearly 80 percent of the market”\textsuperscript{107} in 2006. Fitch presumably held most of the balance.\textsuperscript{108} Many criticized this lack of competition as lowering the quality of ratings.\textsuperscript{109}

2. The Power to License

To distinguish safe securities from risky securities, the SEC needed a measure of safety and quality. The NRSRO concept was developed specifically for this purpose in 1975.\textsuperscript{110} In the context of broker-dealer net capital requirements,\textsuperscript{111} the SEC sought to encourage broker-dealers to hold “investment grade” securities rather than “non-investment grade” securities by permitting broker-dealers to subtract from their net worth a smaller percentage (known as a haircut) for investment grade securities.\textsuperscript{112} This allowed such broker-dealers the benefit of disclosing a higher net capital figure. Since the NRSROs were given the regulatory authority to determine the meaning of the term “investment grade,” they were empowered to determine the substantive effect of these rules—in the parlance of Professor Partnoy, they were given the power to grant “regulatory licenses.”\textsuperscript{113} As a result, market actors, such as broker-dealers, became dependent upon NRSRO ratings in making investment decisions.

After 1975, NRSRO credit ratings became widely incorporated as proxies for safety and quality “in federal and state legislation, rules issued by financial and other regulators, foreign regulatory schemes, and private financial contracts.”\textsuperscript{114} For example, they came to influence the holdings of money market...

\textsuperscript{107} 152 CONG. REC. E1957, 1957 (daily ed. Sept. 29, 2006) (statement of Brian Carroll submitted by Rep. Fitzpatrick) (“[O]nly three NRSROs have staff No Action letters: Moody’s, S&P and Fitch Inc., with the first two capturing nearly 80 percent of the market.”).
\textsuperscript{108} Id.
\textsuperscript{109} See, e.g., 151 CONG. REC. H5255, 5255 (daily ed. June 28, 2005) (statement of Rep. Fitzpatrick) (“Two firms dominate the ratings market with SEC approval and this, Mr. Speaker, creates an uncompetitive marketplace, stifles competition from other rating agencies, lowers the quality of ratings and allows conflicts of interest to go unchecked. It is bad for the market and it is hurtful to individual investors.”).
\textsuperscript{110} See SEC REPORT OF 2003, supra note 9, at 5 (“Since 1975, the Commission has relied on ratings by market-recognized credible rating agencies for distinguishing among grades of creditworthiness in various regulations under the federal securities laws.”).
\textsuperscript{111} See 17 C.F.R. § 240.15c3-1 (2011).
\textsuperscript{113} See Siskel & Ebert, supra note 71, at 623.
\textsuperscript{114} See SEC REPORT OF 2003, supra note 9, at 6.
funds,\textsuperscript{115} banks\textsuperscript{116} and insurance companies.\textsuperscript{117} They also allowed issuers of registered securities not widely followed by the market to use short-form registration statements instead of long forms.\textsuperscript{118}

NRSRO-dependent laws and regulations created an artificial demand in the market to obtain NRSRO credit ratings, which meant guaranteed business for the NRSROs. Commentators criticized this “regulatory licensing” power since such guaranteed business did not reward accuracy, but only membership in the NRSRO club: they were “insulated from the standard market penalty for being wrong—the loss of business.”\textsuperscript{119} As an illustrative example of the financial benefits of such membership, as of December 31, 2006, Moody’s reported net income of approximately $754 million\textsuperscript{120} and, during that December, its market capitalization nearly peaked at $19.32 billion.\textsuperscript{121}

3. The Negligence Shield

In 1981, the SEC reversed a long-standing policy prohibiting the disclosure of credit ratings in registration statements because it believed such ratings would help investors make informed investment decisions about the level of risk they would assume.\textsuperscript{122} Such disclosure, however, would have made the NRSROs “experts” under the Securities Act of 1933, which would subject their ratings to negligence exposure under Section 11. Since this would have imposed a higher liability standard on the NRSROs in the context of the sale of

\textsuperscript{115} See Investment Company Act Rule 2a-7, 17 C.F.R. § 270.2a-7 (2010).
\textsuperscript{117} See, e.g., CAL. INS. CODE § 1192.10 (2010).
\textsuperscript{118} See, e.g., Form S-3, 17 C.F.R. § 239.13 (2010).
\textsuperscript{119} See Calomiris & Mason, supra 48; see also Frank Partnoy, How and Why Credit Rating Agencies Are Not Like Other Gatekeepers, in FINANCIAL GATEKEEPERS: CAN THEY PROTECT INVESTORS? 59, 60-61 (Yasuyuki Fuchita & Robert E. Litan eds., 2006), available at http://ssrn.com/abstract=900257 (In a minimally competitive environment, such regulatory licenses may therefore be responsible for "generating" economic rents for NRSROs that persist even when they perform poorly and otherwise would lose reputational capital.").
registered debt securities than it would have in any other context, the NRSROs informed the SEC that they would not consent to such disclosure.

Since the SEC was comfortable that the antifraud rules would be sufficient to hold the NRSROs accountable for reckless behavior, it created a shield to protect NRSRO credit ratings disclosed in registration statements from negligence exposure in order to encourage disclosure of such ratings. Non-NRSROs, by contrast, would not benefit from such a shield and would therefore need to consent to negligence exposure for ratings disclosed in a registration statement. This gave the NRSROs a considerable competitive advantage over all other CRAs. Commentators cited this benefit as evidence that the NRSROs were held to a lower standard than other comparable gatekeepers such as lawyers, auditors, and even other CRAs.

4. The Seal of Approval

Because the SEC only allowed a small group of “nationally recognized” CRAs into the club and only gave such CRAs the power to license and a negligence shield, it increased the visibility of these CRAs and sent a signal to the market

---

123 CRAs generally have First Amendment protection from common law negligence claims because credit ratings are thought to constitute a form of speech that is of public concern and therefore entitled to the highest form of protection. See, e.g., Approaches to Improving Credit Rating Agency Regulation: Hearing Before the H. Subcomm. on Capital Markets, Insurance, and Gov’t, 111th Cong. 1-2 (2009) (statement of Eugene Volokh, Gary T. Schwartz Professor of Law, UCLA School of Law), available at http://www.house.gov/apps/list/hearing/financialsvcs_dem/volokh.pdf. Under this standard, rating agency speech is protected unless a rating is made with actual malice, meaning “with knowledge that it was false or with reckless disregard of whether it was false or not.” See N.Y. Times Co. v. Sullivan, 376 U.S. 254, 279-80 (1964). Some have indicated that the actual malice standard equates to the scienter element of a Rule 10b-5 claim, but that a state of mind of negligence is not enough. See, e.g., Letter from Laurence H. Tribe & Thomas C. Goldstein, Legal Consultants to Moody’s Investors Serv., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n (Dec. 14, 2009), available at www.sec.gov/comments/s7-24-09/s72409-13.pdf.

124 See 1981 Release, supra note 122, at 42,027 n.27.

125 See id. at 42,025 & n.1. (“The rating organization would continue to be subject to liability under the antifraud provisions of the Federal securities laws . . . . See, e.g., Section 17(a) of the Securities Act (15 U.S.C. 77q(a)); Section 10(b) of the Securities Exchange Act of 1934 (15 U.S.C. 78j(b)) and Rule 10b-5 thereunder (17 CFR 240.10b-5); and Section 206 of the Investment Advisers Act of 1940 (15 U.S.C. 80b-6).”)

126 See PARTNOY, supra note 62, at 14 (“Rating agencies have been sued relatively infrequently, and rarely have been held liable.”).

127 See, e.g., Partnoy, supra note 119, at 83-84 (“The unique problems associated with [NRSROs] as gatekeepers stem from . . . their lack of exposure to civil and criminal liability. Unlike other gatekeepers, [NRSROs] are explicitly immune to prosecution for certain violations of securities law, including section 11 of the Securities Act of 1933 . . . .”).
that their ratings could be relied upon as proxies for safety and quality in consonance with their regulatory functions. Some interpreted this signal as conferring governmental approval on the quality of debt securities, much like a USDA Grade A seal would signify the quality of meat products to consumers.\footnote{See Siskel & Ebert, supra note 71, at 684-86.} Non-NRSROs, in particular, argued that the “lack of NRSRO status substantially hindered their businesses’ rate of growth [since] . . . the marketplace views the NRSRO designation as the equivalent of the ‘Good Housekeeping Seal Of Approval.’”\footnote{SEC REPORT OF 2003, supra note 9, at 38.}

By this account, the government would be responsible for the brand name status of the Big Three. By other accounts, however, NRSRO status itself would not be responsible for such reputational benefits because the Big Three dominated the industry long before NRSRO status came into being. These accounts would attribute the stature of the Big Three to the “reputational capital”\footnote{Reputational capital is a metaphor for the value that a company’s reputation has in generating future business. Since a CRA’s reputation and ability to maintain future business are on the line each time it issues a rating, the CRA is said to be pledging its reputational capital to generate the confidence of investors. See generally John C. Coffee, Jr., Gatekeepers: The Professions and Corporate Governance (2006) (stating the reputational capital view and recognizing its limits).} they acquired over time after being first movers in the industry.\footnote{See Coffee, supra note 1, at 263 (“Their supremacy thus seems more based on ‘first mover’ advantages and the difficulty of entering the field without a proven track record. If, as widely assumed, economies of scale characterize the production of financial information, the first entrant can operate more efficiently and exclude later entrants.”).}

In the absence of informative disclosure, moreover, the SEC had no meaningful mechanism during this period to actually test the accuracy and reliability of the NRSROs’ output. So the reputations of the NRSROs were not based on objective performance measures but merely on industry perceptions. The SEC noted in 2003 that “[w]ithout such an assurance as to the quality of the ratings issued by a rating agency, it would be foolhardy to rely upon ratings as a proxy for credit quality in regulation.”\footnote{SEC REPORT OF 2003, supra note 9, at 38.} Yet, national recognition—and not accurate performance—was the primary qualifying attribute for attaining NRSRO status, and investors relied accordingly on NRSRO ratings as a shorthand for risk. Those outside of the club were excluded from this benefit.
5. Investment Adviser Regulation

Investment adviser regulation under the Advisers Act imposed the sole burden on NRSRO status during the pre-2006 period. Its duties, not narrowly tailored to NRSROs, included filing Form ADV, recordkeeping requirements, and periodic examinations. While the NRSROs were purportedly subject to the same duties as all other investment advisers, they viewed their registrations as voluntary and thought that certain requirements such as the recordkeeping rules did not apply to them. The SEC generally acquiesced to this view. This in turn hampered the SEC’s ability to perform meaningful examinations and led the agency to characterize the NRSROs as not subject to much “formal regulation or oversight.” Advisers Act regulation, however, did confer one benefit: it sent a message to the market that NRSROs were regulated entities whether or not the SEC was really watching them.

---

133 Form ADV required NRSROs to disclose, among other things, information about its advisory business and other business activities, financial industry affiliations, participations or interests in client transactions, disciplinary history, owners and executive officers, and the locations of books and records. It did not require any disclosures specifically tailored to NRSROs such as the disclosure of procedures and methodologies, performance data, rating action histories, or a description of the data relied on when forming ratings opinions. See, e.g., Moody’s Inv. Servs., Form ADV (Feb. 2005), available at http://www.adviserinfo.sec.gov/%28S%282ztk5owxx43l2wyi11rpdaf%29%29iapd/content/viewform/adv022005/sections/iapd_AdvIdentifyingInfoSection.aspx?ORG_PK=111146&RGLTR_PK=&STATE_CD=&FLNG_PK=0588E7BC000801480305C5F0624CAAAD056C8CC0.


135 See Lowe v. SEC, 472 U.S. 181, 208, 211 (1985). The Lowe majority held that a publisher of nonpersonalized investment advice that was circulated broadly to the public fell into an exclusion from the defined term “investment adviser.” The reason was that the “publications [did] not fit within the central purpose of the Act because they [did] not offer individualized advice attuned to any specific portfolio or to any client’s particular needs.” Id.

136 See SEC REPORT OF 2003, supra note 9, at 20 (“T]he effectiveness of the Commission’s examination being hampered by, among other things, the lack of recordkeeping requirements tailored to NRSRO activities, the NRSROs’ assertions that the document retention and production requirements of the Investment Advisers Act of 1940 are inapplicable to the credit rating business, and their claims that the First Amendment shields the NRSROs from producing certain documents to the Commission.”).


138 See SEC REPORT OF 2003, supra note 9, at 4.

139 See HAZEN, supra note 134 (describing phenomenon in 1985 that entities such as the NRSROs could have deregistered as investment advisers after the Supreme Court’s holding in Lowe: “Presumably, these publishers believed it would add to their credibility to be able to state that they were registered with the SEC.”).

After the passage of CRARA in 2006, the ratio of benefits to burdens with respect to the NRSROs began to shift. Between 2006 and 2010, the NRSRO club doubled in size, the NRSROs’ power to license began to marginally diminish, and the SEC considered repealing—but ultimately did not repeal—the NRSROs’ negligence shield. Congress and the SEC also took measures to negate the perception that such credit ratings had the government’s seal of approval. While these measures did little to alter the financial and reputational benefits of NRSRO status (especially for the Big Three), the introduction of NRSRO regulation, which conferred some reputational benefits, added significant new legal burdens to all of the NRSROs.

1. The Expanding Club

In light of critiques regarding the exclusive nature of the NRSRO club, Congress sought to make entry easier through objective standards that relied more on the judgment of investors than on the judgment of the SEC.140 As a proxy for such investors, Congress required certifications from at least ten nonaffiliated qualified institutional buyers that used such ratings for at least three years.141 Nonetheless, no mechanism was put into place to actually test the accuracy and reliability of the NRSROs’ output. By December 2007, the NRSRO club doubled to ten.142

In addition, Congress adopted other measures to foster competition. For example, it adopted new performance disclosure requirements to permit the market to compare the performance of different NRSROs143 and an Equal Access Rule to give lesser established NRSROs an equal opportunity to rate structured finance products by requiring issuers to provide the

---

142 See Commission Orders Granting NRSRO Registration, U.S. SEC. EXCHANGE COMMISSION, http://www.sec.gov/divisions/marketreg/ratingagency.htm (last visited June 20, 2011). The new members were R&I, JCR, Kroll (the successor to LACE Financial Corp.), Morningstar (the successor to Realpoint LLC) and Egan-Jones. As of August 10, 2011, no additional credit rating agencies have been admitted to the club. Jeannette Neumann, Call to Downsize Giants of Ratings, WALL ST. J. (Aug. 10, 2011), http://online.wsj.com/article/SB100014240531190448904564984938844849956.html?mod=ITP_moneyandinvesting_0#articleTabs%3Darticle.
same information to hired and non-hired NRSROs. Since the Equal Access Rule is NRSRO-specific, it represents a new benefit of NRSRO status.

Nonetheless, as of year-end 2010, there had not yet been any tangible impact on competition: the Big Three still issued approximately 97 percent of all outstanding ratings across all categories reported, and in the structured finance product realm, they issued approximately 94 percent of all outstanding ratings. Some studies, moreover, suggested that greater competition was inadvertently producing a race to the bottom, possibly as a result of the issuer-pays conflict and ratings-shopping problem.

2. The First Licensing Rollbacks

In response to criticism over the NRSROs’ power to license, the SEC put out three proposals in July 2008 to amend and replace NRSRO-dependent regulatory licenses. These proposals included dismantling regulatory licenses with respect to broker-dealers, funds, and issuers of registered securities. While six commentators supported the proposals, the majority of commentators opposed them.

144 17 C.F.R. § 240.17g-5(a)(3) (2011); see also SEC 2011 ANNUAL REPORT, supra note 35, at 18 (“[The Equal Access Rule] allow[s] non-hired NRSROs to access information relating to the issuance of structured products that was previously only readily available to hired NRSROs. An NRSRO may then be able to break into the structured finance sector by providing unsolicited ratings on these securities. This would also allow market participants to see differences between credit ratings issued by a non-hired NRSRO and those issued by hired NRSRO and to observe, over time, the differences in the quality of the ratings.”).

145 At present, the Equal Access Rule is being considered as a potential alternative to the Franken Proposal. See infra Part III.A. In fact, the majority of commentators on the Franken Proposal prefer enhancing the Equal Access Rule instead of adopting the Franken Proposal. See infra text accompanying notes 184-87.

146 See SEC SEPT. 2011 SUMMARY REPORT, supra note 47, at 6-7.


As a result, the SEC left all NRSRO regulatory licenses in place except for one: a rule known as the municipal securities exception to the affiliated underwriter prohibition.\textsuperscript{151} This first licensing rollback foreshadowed the SEC’s proposed approach to the new standards of creditworthiness in the post-Dodd-Frank period by replacing a fund’s reliance on an NRSRO’s determination of “investment grade” with the professional judgment of its board.\textsuperscript{152} Since the NRSROs’ regulatory licensing benefit remained substantially intact during this period, substantive damage to the NRSROs cannot be attributed to this rollback. The financial crisis, however, took its toll. As of December 31, 2010, Moody’s reported net income of $508 million (down approximately 33 percent from December 31, 2006),\textsuperscript{153} and its market capitalization reached a low of $4.72 billion (down approximately 77 percent from its peak) in June 2010.\textsuperscript{154}

3. The Negligence Proposals

In light of criticism that the NRSROs were held to a lower standard than similarly situated gatekeepers such as lawyers, auditors, and other CRAs,\textsuperscript{155} the SEC issued two releases that together would have repealed the negligence shield for NRSROs and required the disclosure of their credit ratings.\textsuperscript{156} This would have forced the NRSROs to consent to

\textsuperscript{151} 17 C.F.R. 270.10f-3 (2008). The old NRSRO-dependent rule prohibited registered funds from knowingly purchasing municipal securities from an affiliated underwriter unless they were determined by an NRSRO to be investment grade. See SEC October 2009 Release, supra note 150, at 23-24.

\textsuperscript{152} See infra Part II.C.2.

\textsuperscript{153} See Moody’s 2010 Annual Report, supra note 120, at 27; see also Moody’s 2006 Annual Report, supra note 120, at 20.

\textsuperscript{154} See Moody’s Corporation Historical Market Cap Data, supra note 121.

\textsuperscript{155} See, e.g., Concept Release on the Possible Rescission of Rule 436(g) Under the Securities Act, Securities Act Release No. 9071, Exchange Act Release No. 60,798, Investment Company Act Release No. 28,943 (Oct. 7, 2009) (“NRSROs describe the credit ratings that they provide as opinions with respect to the registrant or security of the registrant, and the Commission notes that other professionals provide opinions upon which investors rely, such as legal opinions, valuation opinions, fairness opinions and audit reports, and we treat these opinions as subject to the Securities Act’s provisions for experts, including our requirements that registrants include the consents of such professionals if their reports are referenced in registration statements. It appears to us that NRSROs and other credit rating agencies are experts similar to other parties subject to liability under Section 11 and that it may no longer be consistent with investor protection to exempt NRSROs from the provisions of the Securities Act applicable to experts.”).

\textsuperscript{156} Id.; see also Credit Ratings Disclosure, Securities Act Release No. 9070, Exchange Act Release No. 60,797, Investment Company Act Release No. 28,942 (Oct. 7,
negligence exposure with respect to their ratings, much as auditors must do with respect to audited financial statements. To the extent that no NRSRO would be willing to consent, however, issuers would not be able to issue registered bonds. Many commentators warned that this would shut down the registered bond market and would likely have a number of other collateral consequences. Fitch pointed out that forced exposure to a negligence standard would defeat the purpose of being an NRSRO. While the majority of commentators expressed an increased willingness to hold the CRAs accountable for their mistakes, most were divided on the utility of this particular set of proposals. To date, the SEC has not acted on it.

4. Attempts to Negate the Seal of Approval

Other criticism suggested that investors placed too much reliance on NRSRO credit ratings. As a result, Congress and the SEC took measures to change the market perception of NRSRO credit ratings. Congress prohibited any representation or implication that any NRSRO “has been designated, sponsored, recommended, or approved, or that the abilities or qualifications thereof have in any respect been passed upon, by the United States or any agency, officer, or employee thereof.” The SEC likewise implemented a number of rules to cosmetically remove nonsubstantive references to NRSRO credit ratings.

2009) (discussing proposal that would have replaced the current permissive disclosure standard for ratings with a required disclosure standard).

157 If credit ratings were required to be disclosed and the CRAs refused consent, issuers would not have been able to issue registered bonds because their registration statements would have contained an omission of a material fact required to be stated therein. This would have subjected such issuers to Section 11 liability.

158 Commentators warned of the following collateral consequences, among others: a migration from registered deals to unregistered deals (e.g., 144A deals), a contraction in ratings coverage, and less accurate defensive ratings due to incentives favoring caution rather than candor. See, e.g., Letter from Robert E. Buckholz, Jr., Chair, & Trevor Ogle, Sec’y, Comm. on Sec. Regulation, N.Y.C. Bar to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n (Mar. 1, 2010); Letter from Deven Sharma, President, Standard & Poor’s to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n (Dec. 14, 2009); Letter from Michel Madelain, Chief Operating Officer, Moody’s Investors Servs. to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n (Dec. 14, 2009), all available at http://www.sec.gov/comments/s7-24-09/s72409.shtml.

159 See Letter from Charles D. Brown, General Counsel, Fitch Ratings, to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n (Dec. 14, 2009) (“The proposal seems to defeat the entire purpose of becoming an NRSRO if one of the perceived benefits of recognition (use of the credit ratings by registrants) creates a significant new liability.”).

160 See, e.g., SEC October 2009 Release, supra note 150, at 1 (“NRSRO ratings in Commission rules may have contributed to an undue reliance on those ratings by market participants.”); Partnoy, supra note 67, at 1.

ratings in statutes and regulations. Its initiative explicitly sought to negate the perception of a seal of approval.

The market, however, did not change any of its previous habits due to these signals to stop over-relying on NRSRO ratings, nor did it punish the Big Three in light of the widespread negative media attention focused on them during the financial crisis of 2008. The Wall Street Journal pointed out that “[the Big Three’s] dominance of the business didn’t change after they lost some credibility for being overly optimistic about the performance of thousands of mortgage-related bonds before and during the financial crisis.” It also remarked that the market continued to ignore the judgments of lesser established NRSROs such as Egan-Jones. Some attribute this phenomenon to “sticky” market practices, which refers to the strong incentives that market participants have to honor existing market norms, such as relying on the Big Three’s credit ratings, even if they appear to be performing poorly. While some may also attribute the phenomenon to the continued dependence on NRSRO regulatory licenses during this period, this does not explain why the other seven NRSROs had so little success in increasing their market share in relation to the Big Three. The most convincing reason for the phenomenon is that “no other ratings firm[s] can come close to matching the number of analysts, broad coverage and decades of experience.” Thus, just as the failure of the Big Three during the financial crisis of 2008 did not impact their market share, attempts to negate the government’s seal of approval

162 These changes were in addition to the aborted proposals to remove regulatory licenses from SEC regulations and the first licensing rollback discussed above. See SEC October 2009 Release, supra note 150, at 11-14. An example of such a cosmetic change was the test to determine whether an Alternative Trading System (ATS) must be registered as an exchange. This test previously depended upon the number of investment grade and non-investment grade corporate debt securities trading on the ATS. The SEC consolidated these two categories into “corporate debt securities” to remove the superfluous “investment grade” distinction. See id. at T-9.

163 Id. at 36 (The SEC stated that the “initiative [was] designed to address the concern that the inclusion in the Commission’s rules and forms of requirements relating to security ratings could create the appearance that the Commission had, in effect, given its ‘official seal of approval’ on ratings, which could adversely affect the quality of due diligence and investment analysis . . . .”).

164 Neumann, supra note 142.

165 Id.

166 Hill, supra note 17, at 144 (“Market practices are sticky, and market actors have strong incentives to abide by them. Even now, after Moody’s, Standard & Poor’s, and Fitch have done so badly, and when other rating agencies are NRSROs, the Big Three are still highly influential.”).

167 Neumann, supra note 142.
had no substantive impact on the Big Three’s market share and no positive effect on the growth of the smaller seven.

5. NRSRO Regulation

In light of the SEC’s inability to regulate and oversee the NRSROs under the Advisers Act, Congress adopted a new, narrowly tailored regulatory regime for NRSROs. NRSRO regulation imposes substantial new legal burdens on the NRSROs without providing any substantive NRSRO-specific benefits. The new legal burdens include new disclosure obligations, recordkeeping rules, prohibitions on unfair business practices, management of certain conflicts of interest, prohibition of other conflicts of interest, and new and enhanced SEC penalty and examination powers.

Although the new regulatory regime also provides some new legal benefits, such benefits do not bear on the NRSRO Nullification analysis. Two of these benefits were designed to apply equally to all CRAs: protection from SEC or State regulation of the substance of credit ratings and the procedures and methodologies used to determine them, and protection from SEC or State regulation of the procedures and methodologies used to determine them.

---


169 See 17 C.F.R. § 240.17g-2.


172 See 15 U.S.C.A. § 78o-7(h); see also 17 C.F.R. § 240.17g-5(c).

173 The SEC has the power to penalize the NRSROs in that they could lose their NRSRO status or be censured, limited, or suspended if they commit one or more enumerated bad acts, including securities laws violations or felonies, or if they fail to maintain adequate financial and managerial resources to consistently produce credit ratings with integrity. See Securities Exchange Act of 1934 § 15E(d), 15 U.S.C.A. § 78o-7(d). These powers are in addition to their general powers under the Exchange Act. See, e.g., Securities Exchange Act of 1934 §§ 21, 21A, 21B, 21C, 32, 15 U.S.C.A. §§ 78u, 78u-1, 78u-2, 78u-3, 78ff. The SEC also has examination authority over all records of NRSROs. See Securities Exchange Act of 1934 § 17(b), 15 U.S.C. § 78q.

174 See Securities Exchange Act of 1934 § 15E(c), 15 U.S.C. § 78o-7(c) (“Notwithstanding any other provision of this section, or any other provision of law, neither the Commission nor any State (or political subdivision thereof) may regulate...”)
against private rights of action (now repealed).\textsuperscript{175} The other benefits include an exemption for the NRSROs from registration under the Advisers Act\textsuperscript{176} and the NRSROs’ voluntary withdrawal right together with the CRAs’ voluntary registration right.\textsuperscript{177} The former does not bear on NRSRO Nullification because the SEC has long treated NRSRO registrations as voluntary under the Advisers Act and continues to treat non-NRSRO registrations in the same manner. While the latter permits NRSRO Nullification, it does not, by itself, affect the decision to opt in or out of the regulatory regime. Since this regime is NRSRO-specific, however, regulation under the new regime continues to confer the same, if not an enhanced, reputational benefit as regulation under the Advisers Act did. The new NRSRO-specific legal burdens, on the other hand, impose new costs that weigh in favor of NRSRO Nullification.

C. Post-Dodd-Frank (2010–Present)

After the passage of Dodd-Frank in July 2010, most of the distinct financial benefits of NRSRO status from the previous eras were called into question or removed: the club was no longer exclusive, measures were proposed to increase competition, the power to license was set to be substantially eliminated on the federal level by July 2012, and the negligence shield was repealed (though, as described herein, negligence exposure was not imposed). In addition, the legal burdens of NRSRO status were again increased substantially through heightened regulation. In spite of reputational damage from the financial crisis of 2008, the Big Three’s credibility nonetheless appeared to remain mostly intact while the

\textsuperscript{175} See Securities Exchange Act of 1934 § 15E(m), 15 U.S.C.A. § 78o-7(m) (West 2008), amended by Dodd-Frank, Pub. L. No. 111-203, § 933(a), 124 Stat. 1376, 1883 (2010). Dodd-Frank’s amendment to section 15E(m) repealed the NRSROs’ protection against private rights of action granted under CRARA, which previously read as follows: “Nothing in this section may be construed as creating any private right of action, and no report furnished by a nationally recognized statistical rating organization determines credit ratings.”. Id.


opinions of the other seven NRSROs continued to be mostly ignored. It remains to be seen how the two most critical unresolved items—the Franken Proposal and the new standards of creditworthiness—will impact this calculus once they are finalized.

1. The Prospect of a New Elite Club

Since expanding the NRSRO club and other initial measures to foster competition among the NRSROs did not produce tangible results, Congress instead began focusing on a different strategy to increase competition: encouraging the existing members of the club to compete over accuracy. The leading proposal, which has not yet been (and may not be) adopted, is the Franken Proposal. This proposal seeks to increase competition among the existing members of the club by creating an environment where all NRSROs compete over accuracy on a level playing field without having to inflate their ratings to win business. It purports to do this by creating a series of benefits that only accrue to a new elite club for which only NRSROs would be eligible. The members of the new elite club would be known as qualified NRSROs (QNRSROs). To join this new elite club and get these benefits, NRSROs would have to be selected by the investor-controlled CRA Board. The CRA Board would rate the QNRSROs based on their track records for accuracy and reward the best performers, over time, with the most business. This would purportedly encourage a race to the top to issue accurate ratings instead of today’s race to the bottom to assign issuer-friendly ratings.

---

178 See supra note 95.
179 See 156 CONG. REC. S3664, 3676 (daily ed. May 13, 2010) (statement of Sen. Charles Schumer) (“[T]he provision Senator Franken is offering and I am proud to cosponsor goes to the heart of the conflict of interest. . . . [T]he smaller rating agencies and investor-paid agencies will have a level playing field to compete against the big three.”).
180 See Franken Proposal, supra note 95, § 15E(w)(1)(B).
181 While this new membership requirement creates a new barrier to entry, it appears that such barrier is meant to be exceedingly low to weed out only those NRSROs that are simply unqualified to rate structured finance products. Senator Franken has been clear that one of the express purposes of his proposal is to give the smaller NRSROs a level playing field to compete with the Big Three. See 156 CONG. REC. S3664, 3674 (daily ed. May 13, 2010) (statement of Sen. Al Franken) (“Standard & Poor’s and Moody’s and Fitch do about, what, 94 percent of the business. The other agencies will get a chance because what will be rewarded is accuracy.”).
182 See Franken Proposal, supra note 95, § 15E(w)(5).
183 For an analysis of the Franken Proposal and recommendations for improving it, see infra Parts III.A.1.b.i and IV.A.
The SEC is also considering five alternative proposals to the Franken Proposal to create incentives for NRSROs to compete over accuracy. These proposals include relying only on the existing Equal Access Rule; creating an investor-owned NRSRO to compete with the existing issuer-paid NRSROs; requiring the price-tag for ratings to be split between issuers, secondary market sellers, and investors; legally mandating a user-pays system; and preserving the issuer-pays model but putting the payment decision into the hands of security holders. These proposals vary with respect to the allocation of new NRSRO-specific benefits and burdens. Of these, the large majority of commentators support relying only on the existing Equal Access Rule with certain enhancements, even though no NRSRO has yet used it, while a smaller segment favors implementing a form of the Franken Proposal. None of the other alternatives appear to have gained much traction.

Congress ordered the SEC to study the Franken Proposal and alternative systems and deliver a report by July 2012 with their recommendations. After delivering the report, to the extent necessary or appropriate in the public interest or for the protection of investors, the SEC must establish an assignment system for selecting NRSROs to determine initial credit ratings for structured finance products that prevents the issuer and other issuer-aligned parties from selecting the NRSRO that will determine and monitor initial credit ratings. It must implement the Franken Proposal unless it determines

---

184 See Solicitation of Comment, supra note 95, at 38-47.
185 For an analysis of the alternatives to the Franken Proposal, see infra Part III.A.1.b.
186 See Letter from Robert Doblas, President, Morningstar Credit Ratings LLC, to Elizabeth M. Murphy, Sec'y, Sec. Exch. Comm'n 6 (Sept. 13, 2011) [hereinafter Morningstar Letter], available at http://www.sec.gov/comments/4-629/4629-24.pdf (“Since the [Equal Access Rule] became effective, we know of no NRSRO that has issued an unsolicited initial rating as a result of the information available under this rule.”).
189 See id. § 939F(d). Notably, DBRS points out that “[a]lthough Section 939F is awkwardly constructed . . . the correct reading of the provision requires the Commission to make the threshold public interest/protection of investors determination before engaging in any new rulemaking on assigned credit ratings.” Letter from Daniel Curry, President, & Mary Keogh, Managing Dir., Regulatory Affairs, DBRS Ltd., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 2 (Sept. 13, 2011) [hereinafter DBRS Letter], available at http://www.sec.gov/comments/4-629/4629-21.pdf. This means that the SEC need not adopt a system separating issuer payment and selection to the extent it finds that it is not necessary or appropriate in the public interest or for the protection of investors.
that an alternative system would be better. The resolution of this item has the potential to bestow important new benefits upon the NRSROs, but it also has the potential to impose significant new burdens.

2. The Elimination of Most Regulatory Licenses

As part of Dodd-Frank, Congress eliminated substantially all NRSRO-dependent regulatory licenses on the federal level effective July 2012 by striking statutory references to NRSRO credit ratings and causing each federal agency to do the same with respect to their regulations. In each case, the relevant federal agencies, such as the SEC, must then replace such references or any “requirement of reliance on credit ratings and . . . substitute in such regulations such standard of credit-worthiness as each respective agency shall determine as appropriate for such regulations.” A question remains, however, whether such NRSRO-dependent regulatory licenses will reemerge unscathed as part of the new standards meant to replace them. Some market participants contend that the requirement to replace such references or any requirement of reliance on credit ratings plainly forbids the federal agencies from incorporating any new form of reliance on NRSRO credit ratings into their new standards of creditworthiness, much like the SEC did with the first licensing rollback. Other market participants, however, interpret the “determine as appropriate” clause to mean that the federal agencies have discretion to incorporate NRSRO ratings as part of the new standards of creditworthiness so long as such incorporation is consistent with Congress’s intent to reduce, as opposed to eliminate, reliance on ratings.

104 Dodd-Frank § 939.
105 See id.
106 See id. § 939A.
107 Id. § 939A(b) (emphasis added).
108 See, e.g., Letter from Dennis M. Kelleher, President & CEO, & Stephen W. Hall, Sec. Specialist, Better Mkts., Inc., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 7 (July 5, 2011) [hereinafter Better Markets Letter] (“The Dodd-Frank Act requires the Commission and other agencies to remove references to credit ratings from their regulations, and to substitute alternative standards of credit-worthiness as each agency deems appropriate. Allowing broker-dealers to continue using credit ratings when assessing credit risk would violate this statutory mandate.” (emphasis omitted)).
109 See, e.g., Letter from Suzanne Rothwell, Managing Member, Rothwell Consulting LLC, to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 7 (July 5, 2011) (“It is unclear whether the ‘as appropriate’ language reflects the intention of Congress that each agency retains some flexibility to continue to rely on credit rating standards.
In its proposed new standards of creditworthiness, the SEC has taken the latter view by including partial reliance on NRSRO credit ratings as part of a broader professional judgment analysis.¹⁹⁶ It appears, however, that such partial reliance on NRSRO credit ratings would be permissive—not mandatory—so investment fiduciaries would ultimately decide whether such partial reliance would be appropriate in satisfying their legal duties.¹⁹⁷

Moreover, it appears that the SEC would not limit such permissive reliance to NRSRO credit ratings.¹⁹⁸ Thus, fiduciaries could equally rely on non-NRSRO credit ratings to the extent they deem appropriate and consistent with their legal duties. Since reliance on credit ratings for these purposes would no longer be limited to NRSROs, this would succeed in eliminating substantially all NRSRO-dependent regulatory licenses on the federal level. Many of the comment letters received to date, especially with respect to the rules governing money market funds, oppose this result and request that the SEC find a way to preserve the NRSRO-dependent rules or lobby Congress to amend Dodd-Frank to this end.¹⁹⁹

Although NRSRO-specific regulatory licenses may be substantially eliminated on the federal level in the near future, they will not be defunct. A handful of NRSRO-dependent

¹⁹⁶ See, e.g., SEC April 2011 Release, supra note 112, at 11 (“Under the proposed amendments, a broker-dealer, when assessing credit risk, could consider the following factors, to the extent appropriate, with respect to each security: . . . Internal or external credit risk assessments (i.e., whether credit assessments developed internally by the broker-dealer or externally by a credit rating agency, irrespective of its status as an NRSRO, express a view as to the credit risk associated with a particular security . . . .)”; see also References to Credit Ratings in Certain Investment Company Act Rules and Forms, Securities Act Release No. 9193, Investment Company Act Release No. 29,592, at 9 (Mar. 3, 2011) (“Fund boards of directors (which typically rely on the fund’s adviser) would still be able to consider quality determinations prepared by outside sources, including NRSRO ratings, that fund advisers conclude are credible and reliable, in making credit risk determinations.”).

¹⁹⁷ See sources cited supra note 195.

¹⁹⁸ See sources cited supra note 195.

¹⁹⁹ See, e.g., Letter from Timothy W. Cameron, Managing Dir., SIFMA’s Asset Mgmt. Grp., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 2-3 (Apr. 18, 2011) (arguing that ratings provisions should be retained in Rule 2a-7 and supporting efforts to urge Congress to amend Section 939A of Dodd-Frank to retain the references to ratings); see also Letter from C. David Messman, Sec’y & Chief Legal Officer, Wells Fargo Funds Mgmt., LLC, to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 1 (Apr. 25, 2011) (same).
regulatory licenses will still remain on the federal level in addition to state NRSRO-dependent regulatory licenses and those in the contracts and other documents of private entities. NRSROs will also continue to benefit from international regulatory licenses though none would bear on the NRSRO Nullification analysis. Thus, a fraction of the previous power to license will continue to be a financial benefit for the NRSROs. Given that the full scope of the regulatory licensing benefit remains unresolved, it is impossible to speculate how the NRSROs’ bottom lines may be affected in the future by this impending change.

3. The Repeal of the Negligence Shield

As part of Dodd-Frank, Congress finally repealed the NRSROs’ long-standing negligence shield under Section 11 for ratings disclosed in a registration statement, but perplexingly, Congress did not thereby expose the NRSROs to negligence. This is because Congress did not repeal the NRSROs’ existing right to refuse consent to negligence exposure, nor did it require that any ratings had to be disclosed (as the SEC considered at one point). As they had threatened to do for some time, the NRSROs withheld their

---

200 The most significant regulatory licenses that survive Dodd-Frank on the state level are those in insurance regulation. See John Patrick Hunt, Credit Ratings in Insurance Regulation: The Missing Piece of Financial Reform, 68 WASH. & LEE L. REV. 1667, 1686 (2011) (“If the insurance regulators maintain rating-dependent regulation, then Dodd-Frank’s purpose in eliminating credit ratings from federal financial regulation will be substantially frustrated.”).

201 Since NRSRO status is a U.S. legal concept, foreign regulatory schemes and treaties generally do not rely on NRSROs, but rather their own parallel construct. For example, under the Basel III framework, the parallel to NRSROs are external credit assessment institutions or “ECALs.” See BASEL COMMITTEE ON BANKING SUPERVISION, BANK FOR INTERNATIONAL SETTLEMENTS, BASEL III: A GLOBAL REGULATORY FRAMEWORK FOR MORE RESILIENT BANKS AND BANKING SYSTEMS 52 (2011), available at http://www.bis.org/publ/bcbs189.pdf. But see SEC REPORT OF 2003, supra note 9, at 8 n.22 (“In El Salvador . . . a rating agency can register as a ‘classifier of risk’ under the country’s securities laws if the rating agency is an NRSRO as recognized by the SEC.” (citations omitted)).

202 See supra note 29.

203 See supra note 31.

204 See supra note 156.
consent to allow issuers to disclose their ratings in future registration statements in order to avoid such exposure. In spite of some convincing rhetoric that Congress intended negligence exposure through the repeal of the negligence shield, the SEC issued temporary guidance assuring the market that the NRSROs’ interpretation of their consent right was correct and that previously required ratings disclosure could henceforth be omitted.

By removing this NRSRO-specific negligence shield, Congress put the NRSROs and non-NRSROs on a level playing field in this respect. As a result, registration statements have generally not included ratings disclosure since the end of July 2010, although issuers have continued to include ratings disclosure in other types of prospectuses. While the repeal of the negligence shield opens up the possibility that CRAs may seek to distinguish themselves from the pack by voluntarily consenting to negligence exposure in the future, there is currently no indication that any would be willing to do so. Since the negligence shield no longer provides a benefit to the NRSROs, however, it removes a distinction that previously weighed against NRSRO Nullification.

---

206 See supra note 32.
207 See supra note 29.
208 See Ford No-Action Letter, supra note 32 (“[D]isclosure of a rating in a registration statement requires inclusion of the consent by the rating agency to be named as an expert. . . . Pending further notice, the Division will not recommend enforcement action to the Commission if an asset-backed issuer as defined in Item 1101 of Regulation AB omits the ratings disclosure required by Item 1103(a)(9) and 1120 of Regulation AB from a prospectus that is part of a registration statement relating to an offering of asset-backed securities.”).
209 Issuers can disclose credit ratings in free writing prospectuses because such prospectuses do not become a part of the final registration statement and therefore are not subject to Section 11 liability. Such free writing prospectuses can include information, such as credit ratings, even though they are not included in the registration statement. See Securities Act of 1933 § 10(b), 15 U.S.C.A. § 77j(b) (West 2011); see also 17 C.F.R. §§ 230.408(b), 433 (2011). Typically, such free writing prospectuses are term-sheets filed with the SEC at the time the issuers begin selling their securities. See, e.g., Ford Credit Auto Lease Trust 2011-A, Free Writing Prospectus (June 27, 2011), available at http://www.sec.gov/Archives/edgar/data/1524342/000095012511062759/u50533f2fwp.htm (showing credit ratings by Moody’s and Fitch). In addition to free writing prospectuses, permissive disclosure of NRSRO credit ratings was also historically allowed in Rule 134 tombstone advertisements without Section 11 liability. See Regulation S-K § 10(c), 17 C.F.R. § 229.10(c). The SEC rescinded this rule in July 2011 effective September 2, 2011. See Security Ratings, Securities Act Release No. 9245, at 6 (July 27, 2011) (“[W]e are removing Rule 134(a)(17) under the Securities Act.”).
Reform introduced two other forms of heightened liability—enhanced SEC penalty powers and exposure to private rights of action under Section 18 of the Exchange Act—that are also NRSRO-specific.

4. The Demise of the Seal of Approval

Dodd-Frank, unlike previous efforts, substantially eliminated references to NRSRO credit ratings on the federal level and, by repealing the negligence shield, eliminated such references in registration statements. The elimination of federal government reliance on NRSRO ratings, which has not yet become effective, would appear to officially sever any perception that NRSRO credit ratings are government-sponsored (at least at the federal level). The elimination of credit ratings from registration statements, moreover, has signaled to investors that the federal government is no longer making it easy for investors to find and rely on such ratings even though they are still available on NRSRO websites and in nonregistration statement prospectuses. Moreover, the fact that the NRSROs refused to accept negligence exposure should raise a red flag for investors since the NRSROs themselves do not have sufficient confidence in their ratings to bear this exposure voluntarily.

While the demise of the government’s purported seal of approval will significantly reduce the visibility of NRSRO credit ratings in statutes, regulations, and registration statements, there has been no shortage of visibility for the Big Three from the media, especially in the realm of debt sold by

---

212 See supra note 175; see also Dodd-Frank, Pub. L. No. 111-203, § 932, 124 Stat. 1376 (2010) (replacing various references to “furnishing” with references to “filing.”). Filing, as opposed to furnishing, brings reports and other documents filed with the SEC (e.g., Form NRSRO) within the purview of liability under Sections 18 of the Exchange Act for false or misleading statements made in such reports or other documents. See Securities Exchange Act of 1934 § 18, 15 U.S.C.A. § 78r (West 2012).
213 Dodd-Frank also lowered the pleading standard in 10b-5 litigation against all CRAs and added a due diligence defense, which will heighten the settlement value and ability to win judgments against the CRAs in future actions and will promote increased due diligence. See Dodd-Frank § 933.
214 See supra note 209 and accompanying text.
sovereign nations. Most significantly, S&P sought to make it abundantly clear that it is at odds with the U.S. government by taking the unprecedented step of downgrading the United States’ triple-A credit rating to double-A in August 2011. Tremendous volatility in the stock markets followed. And while Egan-Jones, one of the seven smaller NRSROs, beat S&P to the punch weeks earlier, “[a]lmost no one paid attention.” This shows yet again that, in spite of any perceived seal of approval attributed to NRSRO status, the smaller NRSROs do not share this benefit equally with the Big Three. The seal of approval benefit may have always been illusory.

Congress’s message, moreover, was not solely that investors and regulators should reduce reliance on ratings. Through heightened regulation and the reintroduction of due diligence, Congress also signaled that the watchdog is now back on guard. This competing message may prove especially salient if the Franken Proposal is adopted. Since QNRSROs would have a new means of regulatory visibility and a new investor-controlled CRA Board to signal the value of NRSRO ratings to investors, this could provide an unprecedented form of visibility and positive press for lesser established NRSROs. It could thereby reverse the long-time reliance on national recognition as a proxy for safety and quality and replace it with an objective assessment of good performance. This might introduce a new investor-sanctioned seal of approval that could provide significant reputational benefits to the best performers.


See Neumann, supra note 142.

To make clear that there would be no government seal of approval, the Franken Proposal proposes the following disclaimer: “This initial rating has not been evaluated, approved, or certified by the Government of the United States or by a Federal agency.” See Franken Proposal, supra note 95, § 15E(w)(6).
5. Heightened NRSRO Regulation

In light of the new problems observed during the financial crisis of 2008, Congress adopted a number of heightened regulatory measures. These measures once again increase NRSRO-specific burdens without adding any offsetting benefits. Such heightened regulation includes new and enhanced disclosure obligations, new and enhanced recordkeeping rules, additional measures to combat conflicts of interest, the establishment of an Office of Credit Ratings, corporate governance reforms, new qualification standards for rating analysts, new whistle-blower duties, a duty to consider information from sources other than the issuer when making rating decisions, procedures to assess the probability that an issuer will default, clear definitions of symbols used to denote credit ratings, and consistent use of any such symbols.

As a result of the expected high costs of compliance, some CRAs that have been considering registration as NRSROs, such as

---

219 See generally Securities Exchange Act of 1934 §§ 15E(q)-(s), 15 U.S.C.A. §§ 78o-7(q)-(s) (West 2011). In addition, the SEC has proposed (1) an enhanced Form NRSRO including a new standard format Transition/Default Matrix that all NRSROs would have to use to show performance data; (2) enhanced rating history disclosures for each rating outstanding on or after June 26, 2007; and (3) a new Information Disclosure Form to accompany each rating action, which will include the disclosure of thirteen items that include the main assumptions used in constructing procedures and methodologies, potential limitations of credit ratings, a description of the data relied on for the purpose of determining such ratings, and the use of third party due diligence services. See SEC May 2011 Release, supra note 211, at 56-88, 99-119, 459-63.

220 See SEC May 2011 Release, supra note 211, at 315-19 (discussing proposed paragraphs (a)(9) and (b)(12)-(15) to be amended to Rule 17g-2, 17 C.F.R. 240.17g-2).

221 See, e.g., Dodd-Frank, Pub. L. No. 111-203, § 932(a)(4), 124 Stat. 1376, 1874-76 (2010); SEC May 2011 Release, supra note 211, at 25-26, 37-45 (discussing proposed paragraph (c)(8) to Rule 17g-5, 17 C.F.R. § 240.17g-5 and proposed new Rule 17g-8(c), which would be codified at 17 C.F.R. § 240.17g-8).

222 See Securities Exchange Act of 1934 §§ 15E(p)(1), (3), 15 U.S.C.A. §§ 78o-7(p)(1), (3). The Office of Credit Ratings will monitor ratings determinations, promote ratings accuracy, "ensure that such ratings are not unduly influenced by conflicts," and conduct an annual examination of each NRSRO. Id. As of September 30, 2011, this office was not established due to the failure of the House and Senate Appropriations Subcommittees on Financial Services and General Government to provide the necessary approval. See SEC SEPT. 2011 SUMMARY REPORT, supra note 47, at 8. The proposed CRA Board under the Franken Proposal, by contrast, would likely not have such funding problems since it would be permitted to levy fees periodically from the QNRSROs and QNRSRO applicants. See Franken Proposal, supra note 95, § 15E(w)(2)(D).


224 See Dodd-Frank § 936.

225 Id. § 934.

226 Id. § 935.

227 Id. § 938.
Rapid Ratings,\footnote{228}{See History, \textit{RAPID RATINGS: FIN. HEALTH RATINGS}, http://www.rapidratings.com/page.php?25 (last visited Jan. 22, 2012) (“Rapid Ratings may or may not apply for [NRSRO] designation, depending on how the SEC moves forwards on a variety of rule amendment recommendations put forth in June 2008.”).} have so far refrained from doing so.\footnote{229}{See Neumann, \textit{supra} note 142 (Rapid Ratings “decided not to seek SEC approval because of compliance and administrative costs. . . . DBRS [] said it is worried that the costs of complying with tougher disclosures if the rules go through could discourage upstarts from seeking SEC approval.”).} A current NRSRO, DBRS, suggestively points out that two current NRSROs, JCR and R&I, have withdrawn their registrations with respect to structured finance products and have thus stopped rating such securities due to the new costs, while another NRSRO, A.M. Best, has curtailed the expansion of its rating activities as a result.\footnote{230}{See DBRS Letter, \textit{supra} note 189, at 8 nn.33-34.} Yet another NRSRO, Kroll, advocates for the expansion of the Equal Access Rule, which provides an NRSRO-specific benefit, to include all CRAs.\footnote{231}{See Letter from James Nadler, President of Kroll Bond Rating Agency, Inc. on Release No. 34-64,456, File No. 4-629, to Elizabeth M. Murphy, Sec'y, Sec. Exch. Comm’n 8 (Sept. 13, 2011) [hereinafter Kroll Letter], \url{available at http://www.sec.gov/comments/4-629/4629-15.pdf}.} This suggests that Kroll may want the ability to take advantage of this benefit without subjection to the legal burdens of the regulatory regime. Along the same lines, Kroll has described the new disclosure rules as a “disincentive to any credit rating agency considering becoming an NRSRO.”\footnote{232}{See Letter from Jules B. Kroll, Chairman & Chief Exec. Officer, Kroll Bond Rating Agency, Inc., to Elizabeth M. Murphy, Sec'y, Sec. Exch. Comm’n 11 (Aug. 8, 2011), \url{available at http://www.sec.gov/comments/s7-18-11/s71811-36.pdf}.} Similarly, A.M. Best and Egan-Jones have described the new disclosure rules as “disproportionately hurt[ing]” and “counter-productive for” the smaller NRSROs.\footnote{233}{See Letter from Larry G. Mayewski, Exec. Vice President, A.M. Best Co., to Elizabeth M. Murphy, Sec'y, Sec. Exch. Comm’n 2 (Aug. 8, 2011), \url{available at http://www.sec.gov/comments/s7-18-11/s71811-39.pdf}; Letter from Sean J. Egan, Egan-Jones Ratings Co., to Elizabeth M. Murphy, Sec'y, Sec. Exch. Comm’n 1 (Aug. 5, 2011), \url{available at http://www.sec.gov/comments/s7-18-11/s71811-27.pdf}.}

In summation, NRSRO status has become much less attractive than it used to be. Of the three main financial benefits from earlier eras—lack of competition, regulatory licensing power, and the negligence shield—only a fraction of the regulatory licensing power remains and the Equal Access Rule provides a benefit that no NRSROs have yet used. In addition, a reputational benefit remains by virtue of the message sent to the market that such NRSROs are regulated. Nonetheless, the market continues to listen to the opinions of the Big Three and mostly ignores the seven smallest NRSROs,
showing that NRSRO status confers no meaningful seal of approval. Finally, the legal burdens of NRSRO status have substantially increased through heightened regulation. While the tipping point towards NRSRO Nullification has not been reached for any of the club’s current members, the evidence suggests that today’s NRSROs are paying careful attention to the benefit/burden calculation. Depending on the resolution of today’s unresolved items, withdrawal from the club appears to be a realistic possibility.

III. THE IMPLICATIONS OF TODAY’S UNRESOLVED ITEMS

This part identifies and discusses the financial and reputational implications that today’s most critical unresolved items—the Franken Proposal and the new standards of creditworthiness—will have on the NRSROs’ decision to withdraw from Ratings Reform. It then assesses the impact that today’s existing distinctions in concert with such unresolved items will have on this decision and the legal implications of NRSRO Nullification. This part also assesses the extent to which such items and their alternatives will promote accurate and reliable ratings. It concludes that certain proposals for resolving these items would be more likely to tip the balance toward NRSRO Nullification, while other proposals would be more likely to prevent this result, and that the smaller seven are more likely to opt out of Ratings Reform than the Big Three.

A. Financial and Reputational Implications

1. The Franken Proposal and its Alternatives

a. The Franken Proposal

Mechanically, the Franken Proposal would require the SEC to create a self-regulatory organization (which could also be a public or private utility),\(^\text{234}\) known as the CRA Board, whose directors would be controlled by—and therefore aligned with—the investor community.\(^\text{235}\) The CRA Board would


\(^{235}\) See Franken Proposal, supra note 95, § 15E(w)(2)(C) (“The [CRA] Board shall initially be composed of an odd number of members selected from the industry...[N]ot less than a majority shall be representatives of the investor industry who do not represent issuers...not less than 1 member should be a representative of the issuer industry...not less than 1 member should be a
determine the QNRSRO club’s membership through an application process whereby they would assess, among other things, the institutional and technical capacity of each NRSRO to issue the required ratings.\textsuperscript{236} Unlike the high barrier to entry that prevented non-NRSROs from becoming NRSROs prior to 2006, the QNRSRO barrier is intended to impose an exceedingly low threshold for entry.\textsuperscript{237} Its goal of divorcing issuer selection from issuer payment would mitigate the issuer-pays conflict and the ratings-shopping problem. The CRA Board would seek to achieve this goal by matching QNRSROs and issuers with respect to the initial rating for a structured finance product that the issuer would presumably purchase in advance.\textsuperscript{238} The match would be made through a random assignment system, such as a lottery or rotation, and the CRA Board would rate the best performers and then increase or decrease a QNRSRO’s market share in such matches based on accurate performance over time.\textsuperscript{239} At least once each year, the CRA Board would assess the performance of the QNRSROs.\textsuperscript{240} Issuers would still be free to shop for and purchase second ratings on the open market,\textsuperscript{241} and the Equal Access Rule would continue to permit nonhired NRSROs to compete through unsolicited ratings.

This proposal has been met with substantial criticism from the majority of today’s NRSROs.\textsuperscript{242} In particular, there is a concern that the CRA Board would be poorly suited to determine which NRSROs should be in the QNRSRO club and how assignments should be allocated among them. The smaller NRSROs, for whom the proposal was purportedly developed, are concerned that they will be shut out of the benefits by new conflicts of interest that may favor the Big Three as well as high costs and bias based on their lesser developed track

\begin{footnotes}
\item[236] See id. § 15E(w)(3).
\item[237] See supra note 181.
\item[238] See Franken Proposal, supra note 95, § 15E(w)(5).
\item[239] See id. (“The [CRA] Board shall . . . evaluate a number of selection methods, including a lottery or rotating assignment system . . . . In evaluating a selection method . . . . the Board shall consider . . . . a mechanism which increases or decreases assignments based on past performance.”).
\item[240] See id. § 15E(w)(7) (“The [CRA] Board shall . . . evaluate the performance of each [QNRSRO], . . . at a minimum, [once per year].”).
\item[241] See id. § 15E(w)(9) (“Nothing in this section shall prohibit an issuer from requesting or receiving additional credit ratings with respect to a debt security, if the initial credit rating is provided in accordance with this section.”).
\item[242] See supra note 187.
\end{footnotes}
records. Such a club could simply reentrench the Big Three. The Big Three are concerned that their dominance could be eroded in a nonmeritorious way and that the CRA Board may be plagued by investor biases.

None, however, appears to take issue with the CRA Board’s rating function—only its ability to use this function to allocate ratings business. In fact, many of the NRSROs would seem to support the rating function, because it would provide investors with a new tool to decide which NRSROs to hold in highest esteem.

While some of these concerns may be overstated and could be addressed in the final rule, the widespread concern over the allocating function, especially from the smaller NRSROs, suggests that this cure may be worse than the disease.

243 See, e.g., DBRS Letter, supra note 189, at 8, 14 (discussing CRA Board’s conflicts and arguing that the cost of participating in the Franken Proposal’s system would discourage all but the largest rating agencies from even trying); see Kroll Letter, supra note 231, at 2 (“[W]e question how the Board will evaluate the performance of a new NRSRO, such as [Kroll], in a way that enables it to obtain [rating] assignments based on merit.”).

244 See Kroll Letter, supra note 231, at 2 (“May further entrench the largest incumbents”); Coffee, supra note 1, at 258 (“If the Board were to prefer established raters with a demonstrated history of rating accuracy, this would largely perpetuate the existing oligopoly of the Big Three and might subject the Board to criticism for failing to encourage greater competition.”).

245 See Letter from Patrick Milano, Exec. Vice President, Operations, Standard & Poor’s Ratings Servs., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 3 (Sept. 13, 2011) [hereinafter S&P Letter], available at http://www.sec.gov/comments/4-629/4629-19.pdf (“The [Franken Proposal] would effectively treat ratings as a commodity, presuming that all ratings on structured finance products . . . are of equal quality and utility.”); see also Bai, supra note 50, at 50 n.9 (quoting Sen. Christopher Dodd: “Not all the rating agencies are equal . . . . [T]here are different companies of different sizes and needs, and to be choosing rating agencies based on arbitrary choice without considering whether or not the rating agency can actually do the job is my concern.” (citations omitted)). Others have expressed similar concerns. See Coffee, supra note 1, at 258.

246 See S&P Letter, supra note 245 (“Those ultimately responsible for selecting NRSROs to perform work may have their own biases. . . .”).

247 See, e.g., DBRS Letter, supra note 189, at 2 (“These goals are better served by measures that . . . give investors the tools they need to make informed choices about which credit ratings to employ in making their investment decisions.”); Letter from Michel Madelain, President & Chief Operating Officer, Moody’s Investors Serv., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 6 (Sept. 13, 2011) [hereinafter 2011 Moody’s Letter], available at http://www.sec.gov/comments/4-629/4629-23.pdf (“Because investors demand credible (i.e. objective, predictive and relatively stable) ratings, . . . issuers are motivated to seek ratings from CRAs that have the best reputations among investors. . . .”); S&P Letter, supra note 245, at 11-12 (“By selecting a capable, well-regarded rating agency with a reputation for independence and objectivity, issuers can increase the marketability of their securities.”); Morningstar Letter, supra note 186, at 4 (“A Section 15E(w) System that examines ratings accuracy and timeliness in the future assignment of ratings will encourage competition among NRSROs to provide the most accurate and timely ratings in order to ensure that they will continue to secure additional business under the Section 15E(w) System.”).
The first benefit of the Franken Proposal is that QNRSROs would get a guaranteed percentage of the structured finance market for initial ratings. This guaranteed market share would prove invaluable to any CRA since structured finance products have the proven potential to outpace every other type of debt security and, due to their complexity, such products command justifiably higher prices. On the other hand, exclusion from the QNRSRO club could impose a substantial burden.

The Franken Proposal does not explain how the CRA Board would initially allocate the guaranteed portion of initial ratings business nor how exclusion would be decided. In the free market, as of year-end 2010, the eight NRSROs registered with respect to structured finance products shared the market as follows: S&P (38.98 percent), Moody’s (33.57 percent), Fitch (21.34 percent), DBRS (3.34 percent), Morningstar (2.75 percent), A.M. Best (0.02 percent), Egan-Jones (0.00 percent), Kroll (0.00 percent). While these figures show that the Big Three commanded almost 94 percent of this market, such dominance does not seem manifestly disproportionate when taking into account the disproportionately large staffing that the Big Three have compared with the smaller NRSROs. Were we to treat the total staff members at these eight agencies (3,855) as devoted to structured finance ratings only, a proportionate initial allocation of assignments by the CRA Board (assuming all eight were admitted as QNRSROs) vis-à-vis staffing would require dividing each agency’s total staff devoted

---

248 See Enhancing Investor Protection, supra note 52, at 55 (“[R]ating structured finance products generated much higher fees than rating similar amounts of corporate bonds. For example, rating a $350 million mortgage pool could justify a fee of $200,000 to $250,000, while rating a municipal bond of similar size justified only a fee of $50,000.” (citations omitted)); see also Abu Dhabi Comm. Bank v. Morgan Stanley & Co., 651 F. Supp. 2d 155, 167 (S.D.N.Y 2009) (“[T]he Rating Agencies each received fees in excess of three times their normal fees for rating the Cheyne SIV . . . .”).

249 Eight of today’s ten NRSROs are registered with the SEC with respect to structured finance products. See Solicitation of Comment, supra note 95, at 7. Yet only a subset of this group might benefit from the guaranteed business afforded by the CRA Board’s system to the extent such NRSROs are not selected to be QNRSROs. See supra note 236 and accompanying text.


251 Id. at 8.

252 The actual number of credit analysts and supervisors devoted only to structured finance is not available. Disclosure on Form NRSRO only requires the total number of credit analysts and supervisors and does not distinguish among the five ratings categories: financial institutions, insurance companies, corporate issuers, asset-backed securities (i.e., structured finance), and government securities. See id.; see also Form NRSRO, supra note 168, Exhibit 8.
to structured finance products by the total staff of all eight agencies devoted to the same. This would produce the following initial allocations: S&P (34.89 percent), Moody’s (31.23 percent), Fitch (27.20 percent), DBRS (2.46 percent), Morningstar (0.62 percent), A.M. Best (3.11 percent), Egan-Jones (0.13 percent), Kroll (0.34 percent). Accordingly, the Big Three’s market share, which would equal approximately 93 percent, would not be meaningfully diluted. Such an allocation, however, would give more work to firms like A.M. Best, Egan-Jones, and Kroll, which would help them establish track records for accuracy.

Since the Franken Proposal does not explain how the CRA Board would initially allocate the guaranteed portion of initial ratings nor how it would exclude NRSROs, today’s NRSROs understandably have concerns. One potential solution to this problem would be to equate QNRSROs with those NRSROs that are registered to rate structured finance products and require that initial allocations be made in proportion to staffing resources. Form NRSRO could be amended to require the disclosure of the credit analysts and supervisors devoted to each rating category. The institutional and technical capacity test could also be objectified to confirm that it is only meant to weed out NRSROs that are patently unqualified to rate structured finance products. While the smaller NRSROs would likely favor these solutions, the Big Three might actually prefer allowing CRA Board discretion since it could provide a way for investor representatives to filter out NRSROs that the market would not trust. In spite of the Big Three’s poor performance during the financial crisis, the market has continued to trust the Big Three, so there is practically no danger that the Big Three would face exclusion from the QNRSRO club. The exclusion of the smaller NRSROs would impose severe financial and reputational burdens, since they would get no market share in initial ratings, which would translate into fewer jobs in the market for second ratings.

The Big Three may also object to allocations in proportion to staffing. Even though this would not meaningfully dilute their current market share, it would encourage smaller NRSROs to hire more credit analysts and supervisors in order to take a larger slice of the pie. This would potentially erode the Big Three’s market share in a

\[253\] Cf. supra note 252.

\[254\] See supra notes 146-47 and accompanying text.
nonmeritorious way. Any initial allocation, however, would be nonmeritorious since there is not yet an accepted way to measure good performance. Merit-based track records for accuracy will take time to develop.255

The second benefit of the Franken Proposal for those in the QNRSRO club is that the best-performing QNRSROs would be rewarded with bonus assignments on initial ratings. This mechanism would directly reward accurate performance and would permit any subset of the QNRSROs to dominate the field by virtue of their merit.

While initial allocations based on proportionate staffing would not be merit-based, the CRA Board’s ability to adjust ratings business in favor of the best performers based on past performance addresses the fundamental problem of non-merit-based allocations. This also addresses any concerns that guaranteed business would remove the incentives of the NRSROs to work hard and produce rigorous and competitive ratings.256 Thus, merit-based allocations would become the new norm over time, which would confer a significant benefit on the best-performing QNRSROs. While it would also impose a burden for the poorer performing QNRSROs, it would provide the proper incentives for such QNRSROs to improve their performance or be penalized.

There is legitimate concern, however, that such allocations will be unfair due to conflicts of interests or plain incompetence. To the extent the CRA Board has subjective discretion to allocate ratings, commentators have pointed out that at least two new conflicts of interest would be introduced: bias in favor of QNRSROs willing to produce more conservative investor-friendly ratings (whether or not justified) since the CRA Board would be controlled by investors;257 and bias in favor of the Big Three since the CRA Board will depend on fees from the deepest pockets among the QNRSROs for its continued existence.258 While the final rules could mitigate such new

255 See Letter from Senator Al Franken & Senator Roger F. Wicker, to Honorable Mary Schapiro, Chairman, Sec. & Exch. Comm’n 4 (Sept. 14, 2011), available at http://www.sec.gov/comments/4-629/4629-28.pdf (“Of course, these track records would take time to develop. But, over time, it would be possible to effectively judge NRSROs on track records of accuracy.”).

256 See 2011 Moody’s Letter, supra note 247, at 8 (“Such a system may create incentives to conduct the least amount of work and innovation possible to remain in the lottery or rotation system.”); see also S&P Letter, supra note 245, at 2.

257 See S&P Letter, supra note 245.

258 See DBRS Letter, supra note 189, at 8 (“It is likely to cost at least $300-400 million to operate the Board. Under the best of circumstances, this expense would
conflicts by requiring the metrics for such allocations to be publicly disclosed and based on objectively verifiable performance data, the determination of such metrics is not obvious. The CRA Board would undermine its legitimacy if it appeared to be plagued by new conflicts of interest.

Many in the market are concerned, most saliently, that the CRA Board would be unqualified to discern what accurate performance means and thus could not develop any system to reward or penalize QNRSROs based on past performance. For example, the Investment Company Institute asked, "[W]hat will be the criteria used for determining the ‘best performer’ for purposes of assigning a rating agency to a new issue? Is an ‘A1’ rating more correct than an ‘A’ rating?"

The Financial Services Roundtable points out that there are no accepted performance measures, and if investors’ preferences and the CRA Board’s preferences deviate, investors would be stuck with bonds rated by NRSROs they do not like. S&P altogether “rejects the notion that credit ratings can be ‘accurate’ or ‘inaccurate.’" To the extent the CRA Board is not able to produce a definition of accuracy that the market accepts, the goal of empowering this board to make merit-based allocations would be completely compromised.

At present, some efforts have been made to develop methodologies for defining accuracy. Professor Lynn Bai has provided one example of standards for discerning accurate performance that may be instructive in measuring the performance of the various NRSROs and Professor Coffee has provided another suggestion. In addition, Moody's has

have to be divided among only four or five rating agencies, and would be in addition to the costs of complying with whatever extra layer of regulation Qualified NRSRO status might entail. At the end of the day, the cost of participating in the Section 15E(w) System would be so high that it would discourage all but the largest rating agencies from even trying." (footnote omitted)).


See Bai, supra note 50, at 103 (“The default ratio, ‘fallen angels’ ratio, rating change ratio, and large rating change ratio are used to measure rating agencies’ performance.”); see also Senate Turmoil Hearings, supra note 49 (Professor Coffee has suggested that “[t]he SEC should define what ‘default’ or ‘impaired’ means so as to include delayed payments, then calculate these rates over five year cumulative periods, and publish its results on its own website. This would enable consumers to engage in a simple, one stop comparison. For smaller institutions (e.g., small pension funds, or
identified two methods of measuring ratings accuracy: one ordinal (meaning relative performance) known as Average Position, and the other cardinal (meaning absolute performance) such as the “investment-grade default rate and the average rating of defaulters prior to default.”

Bai indicates that ratios measuring default, migration from investment-grade to non-investment-grade, and stability are universally accepted, but other metrics are used without a consensus over their value. Morningstar has suggested that calibrated qualifications “could eliminate any undue influence or the necessity for the CRA Board to make any individual decisions with regard to particular securitization transactions.” While such transparency may minimize concerns over conflicts of interest, any system seeking to define accuracy in a business based on future predictions will necessarily come under careful scrutiny, especially by those found to be poor performers. This would suggest that, before there is a consensus in the market over how to measure relative accuracy, any system including an allocating function based on accuracy ratings would be premature.

The final benefit of the Franken Proposal is reputational and indirectly financial. Unless the market has widespread reason to disagree with the rating abilities of the CRA Board, the best-performing QNRSROs would be in a better position over time to compete in the market for second ratings than excluded NRSROs and non-NRSROs by virtue of the reputational capital they would develop from the CRA Board’s rankings. By publicly rewarding accuracy and thereby communicating to the market a view regarding the agencies that would be most likely to produce accurate initial ratings, such signals would provide a benchmark for accuracy against which the accuracy of second ratings could be tested. This would influence the preferences of investment fiduciaries in relying on NRSROs, since it would provide insight regarding the best performers that is currently hard to identify in the market. These preferences, in turn, would influence the hiring

---

263 2011 Moody’s Letter, supra note 247, at annex i. “Our principal measure of ratings accuracy is the Average Position (AP) of defaulters. Bounded between 0 and 1, AP measures where in the distribution of ratings defaulters were located relative to non-defaulters.” Id.

264 See Bai, supra note 50, at 79-80.

265 Morningstar Letter, supra note 186, at 3.
practices of issuers since interest rates on debt securities would be affected in their negotiations with underwriters.

If investment fiduciaries believed in the accuracy of a rating, then they would accept an interest rate commensurate with the normal range for such rating. If, on the other hand, they believed that a rating may be inflated, they would justifiably demand a premium on the interest rate above the normal range. For example, if an issuer were to hire the worst-performing QNRSRO (or, alternatively, an excluded NRSRO or non-NRSRO) to give a second rating on its securities when it could have hired one of the best-performing QNRSROs, investment fiduciaries would demand a higher interest rate to compensate them for the risk of inflated ratings. They would communicate these preferences to underwriters prior to pricing and such preferences would get incorporated into the pricing negotiations. To avoid this dilemma, issuers would come to prefer hiring the best-performing QNRSROs to give second ratings since this would undermine any such argument about the propriety of a ratings inflation risk premium.

In this way, the CRA Board’s rating function would realign the financial and reputational interests of all QNRSROs to favor accuracy. Since the CRA Board would send a signal of credibility into the market, all QNRSROs would be encouraged to produce the most accurate ratings possible to benefit from this signal. Although the poor performers would experience a reputational burden, this would only further incentivize them to improve their performance. Investment fiduciaries, moreover, would generally trust this message since it would come from an investor-controlled board. Since only QNRSROs would be evaluated, the market would encourage non-NRSROs and excluded NRSROs to join the NRSRO and QNRSRO clubs to get this reputational benefit. Opting out of either status, by contrast, would mean sacrificing such reputational benefit. So the rating function would work to keep NRSROs voluntarily regulated (since only NRSROs can be QNRSROs) while at the same time promoting accurate and reliable ratings.
b. Alternatives to the Franken Proposal

As required by Congress, the SEC is also considering five alternative proposals to the Franken Proposal to create incentives for NRSROs to compete over accuracy.266

i. Rely Only on the Existing Equal Access Rule

The first alternative is to rely only on the existing Equal Access Rule implemented in November 2009.267 This rule created a mechanism for non-hired NRSROs in the structured finance context to gain equal access to the information that issuers provide to the NRSROs they hire so that such non-hired NRSROs can provide unsolicited second ratings.268

While this rule allows non-hired NRSROs to compete in the market for second ratings, and thereby ostensibly helps the lesser known NRSROs develop track records for accuracy, it does not provide a financial patron for such activities.269 Thus, there is no guarantee that any NRSRO will find it financially rational to take advantage of the Equal Access Rule. Indeed, it appears that no NRSRO has taken advantage of the rule to date.270 Moreover, they would not be encouraged to do so, since this regime would not reward accuracy. Issuers would continue to provide the primary source of revenue for ratings under this system and NRSROs would be encouraged to please such issuers even in the unsolicited context so that they could potentially gain business from them in the future. Unless this rule were coupled with a rating mechanism, such as that provided by the Franken Proposal, it would only perpetuate today's race to the bottom, since issuers would still prefer to hire NRSROs willing to inflate their ratings, and most investment fiduciaries would not have the ability to forcefully challenge such preferences.

Moreover, the enhancements proposed by the various supporters of this proposal generally do not attempt to solve the problem that non-hired NRSROs cannot afford to give

---

266 See supra note 95.

267 See supra note 144.

268 See supra note 144.

269 See supra note 144.

270 See Morningstar Letter, supra note 186, at 6 (“We believe that the absence of unsolicited initial ratings primarily results from the costs of providing these unsolicited ratings without adequate compensation and a lack of interest by arrangers and investors in these ratings.”).

271 Id.
unsolicited ratings for free, nor do they propose how market actors should distinguish the best-performing NRSROs from the worst, which is a task “far beyond the means of many insurance companies, pension plans, and other small institutional investors.” The proposed enhancements are generally focused on two points: relaxing a rule that non-hired NRSROs must rate one out of every ten deals they review and allowing the disclosed information, which is currently available only to NRSROs, to be made available to all market participants (most pertinently, non-NRSROs).

Reliance only on the Equal Access Rule is effectively a do-nothing approach that does not offer new financial benefits to the NRSROs nor a mechanism to promote accurate and reliable ratings. Moreover, it imposes a financial burden as a condition for competing since lesser established NRSROs would need to find separate funding to provide such ratings. One way to indirectly get separate funding would be to leave the regulatory regime and put such costs savings into unsolicited ratings instead. This might be the thinly veiled motivation behind the suggestion from some NRSROs that equal access should be expanded to non-NRSROs. Thus, adopting this proposal may encourage NRSRO Nullification. This proposal, subject to the suggested enhancements described above, is

---

271 Morningstar appears to be the only exception: it proposes applying the Franken Proposal’s framework to the Equal Access Rule such that “NRSROs could be selected on a rotational basis to provide an unsolicited rating.... [and] NRSROs could be compensated on a market-value basis that could represent the average compensation paid by the issuer to the Commission or other organization to the other credit rating agencies rating the same transaction.” See id. at 7.


273 17 C.F.R. § 240.17g-5(e) (2011). See Kroll Letter, supra note 231, at 3 (suggesting eliminating requirement that one in ten deals reviewed must be rated); Letter from Karrie McMillan, Gen. Counsel, Investment Co. Inst., to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 6-7 (Sept. 13, 2011), available at http://www.sec.gov/comments/4-629/4629-22.pdf (suggesting increasing the number of free peeks from ten to twenty-five since a new entrant might find, for example, that it is not qualified to review the first twelve issuances it reviews); Letter from Richard A. Dorfman, Managing Dir., Head of Securitization, & Christopher B. Killian, Vice President, Securitization Grp., Sec. Indus. & Fin. Mkts. Ass’n (SIFMA), to Elizabeth M. Murphy, Sec’y, Sec. Exch. Comm’n 23 (Sept. 13, 2011), available at http://www.sec.gov/comments/4-629/4629-9.pdf (suggesting that only one in twenty deals reviewed be required to be rated).

274 See Kroll Letter, supra note 231, at 3 (suggesting equal access should be expanded to all market participants, not just NRSROs); 2011 Moody’s Letter, supra note 247, at 2 (same).
supported by Moody’s, S&P, Fitch, and Kroll as well as a number of other non-CRA entities.275

ii. Create an Investor-Owned NRSRO

The second alternative is to create an investor-owned NRSRO that would compete with the existing NRSROs. By law, issuers would have to obtain initial credit ratings from the investor-owned NRSRO and a second issuer-chosen rating from one of the existing NRSROs. While this solution divorces issuer selection from issuer payment in the initial credit rating context by presumably forcing the issuer to pay for a rating from the investor-owned NRSRO, it does not do so with respect to second ratings. Thus, a monopoly would be created for the investor-owned NRSRO in the market for initial ratings while a fierce competition would be set into motion in the market for second ratings.

Since the market for initial ratings would be monopolized, today’s NRSROs would be denied their current market share of initial ratings representing the loss of an important financial benefit. Since this model has no mechanism to reward accuracy, the market for second ratings would be plagued by the same race to the bottom that currently exists for all ratings. Issuers (and perhaps investors), moreover, would likely distrust the ratings from the investor-owned NRSRO for a number of reasons: it may have investor biases since investors own it, it may lack incentives to do high quality work since it would have a monopoly, and it would lack an established track record as a newcomer. As a result, such ratings would not provide a reliable benchmark for accuracy against which the accuracy of second ratings could be tested. So this model would remove an important financial benefit from today’s NRSROs (the ability to compete for a share of the initial credit ratings market), and like the first alternative approach, it would not address the conflicts of interest at the heart of the rating-agency question. In the comment letters sent to the SEC, no commentators supported this proposal.276

275 See Comments on Solicitation, supra note 187.
276 See id.
iii. Split the Bill Among Participants

The third alternative does not clarify whether it would permit the issuer to select its preferred NRSRO for initial credit ratings. It only focuses on fees by stating,

an NRSRO would be compensated through transaction fees imposed on original issuance and on secondary market transactions. Part of the fee would be paid by the issuer or secondary-market seller and the other portion of the fee by the investors purchasing the security in either the primary or secondary markets.\footnote{See Solicitation of Comment, supra note 95, at 44.}

It appears that this model intends to split the bill among the issuer, investors, and secondary-market sellers, which would presumably mitigate some of the loyalty that the NRSROs feel toward the issuer as their sole client. While this would partially divorce issuer selection from issuer payment, it would not satisfy Congress’s requirement that the issuer and issuer-aligned parties may not select the NRSRO that will determine and monitor initial credit ratings.\footnote{See supra note 189.}

Moreover, although this proposal partially mitigates the issuer-pays conflict, it does not address the ratings-shopping problem if the issuer would ultimately decide which NRSRO to choose. Thus, there would still be a race to the bottom since this model proposes no mechanism to reward accuracy. As a result, the misaligned incentives and today’s \textit{status quo} would persist. This model proposes no other benefits to the NRSROs. In the comment letters sent to the SEC, no commentators supported this proposal.\footnote{See Comments on Solicitation, supra note 187.}

iv. Institute a User-Pays Model

The fourth alternative involves the institution of a user-pays model. Under this model, the issuer would not pay for any ratings and the full fee would be allocated to users, defined as “any entity that included a rated security, loan, or contract as an element of its assets or liabilities as recorded in an audited financial statement.”\footnote{See Solicitation of Comment, supra note 95, at 46.} Such users would be required to enter into a contract with an NRSRO and pay for its rating services.\footnote{See id.}
While forcing investors to buy NRSRO ratings would guarantee business to the NRSROs by effectively reinstituting a new form of NRSRO-dependent regulatory license, the failure of a mechanism to reward accuracy would mean that this model would suffer from today’s conflicts of interest in reverse. Although it eliminates the issuer-pays conflict, it introduces in its place a user-pays conflict and merely reverses the ratings-shopping problem. Under this model, some users would shop to select the NRSROs most willing to deflate (rather than inflate) ratings at the lowest prices. While this might benefit investors in the short term by driving up interest rates on relatively safe securities, this type of artificial ratings deflation could significantly raise the cost of capital for issuers, which could have unforeseen macroeconomic consequences in much the same way that artificial ratings inflation did during the financial crisis of 2008. Since this model would only offer new financial benefits to those NRSROs most willing to please investors rather than those most able to produce accurate ratings, it would simply flip today’s problems instead of address them. In the comment letters sent to the SEC, no commentators supported this proposal, although one commentator provided a detailed alternative user-fee system.

v. Institute an Investor Designation Model

The fifth alternative would be an investor designation model. Under this model, the security holders would select the NRSRO or NRSROs that would get paid and the amount of payment while the issuer would pay the NRSRO or NRSROs according to such directions. This would successfully divorce issuer selection from issuer payment in the market for initial ratings in keeping with Congress’s mandate. Instead of the Franken Proposal’s CRA Board making rating assignments in advance to a specific NRSRO, however, the NRSROs would choose to do the work before knowing whether they would be compensated and, if so, how much. After the work is done, the security holders would make the compensation decision “based on their perception of [the] research underlying the ratings.”

---

282 See Partnoy, supra note 62, at 12 (Some institutions “might press the rating agencies for lower ratings in hopes of receiving higher returns.”).
284 See Solicitation of Comment, supra note 95, at 45.
They would vote in proportion to their holdings and the payment would be made according to such instructions from a deposit (presumably equal to the cost of one rating) that the issuer would have placed in advance with a third party administrator as opposed to the CRA Board under the Franken Proposal. There does not appear to be any requirement that the security holders would have to consider long-term performance data in allocating such payment.

This model would not only fail to reward accuracy, but would also perpetuate the same financial burden that plagues the Equal Access Rule: the prospect of uncompensated work. Not only would there be no guaranteed minimal level of business, but the NRSROs would be asked to spend time and resources to produce ratings without any guarantee of payment in whole or in part. Moreover, as with the user-pays model, this model reverses the ratings-shopping problem by putting the payment decision into the hands of interested security holders, even though the payment source would be the issuer. Since such security holders would not have to consider the long-term accuracy of each NRSRO’s rating performance when deciding which NRSROs to pay, they would be encouraged to pay the NRSROs that have given the most deflated ratings in such offering since that would translate into higher interest rates benefiting the security holders in the short term. The familiar race to the bottom would persist, financial benefits would only accrue to those NRSROs able to win the affection of the most security holders and one new financial burden (uncompensated work) would be imposed. In the comment letters sent to the SEC, no commentators supported this proposal.285

In summary, none of the alternatives to the Franken Proposal that the SEC is currently considering would offer NRSRO-specific benefits that are comparable to those offered by the Franken Proposal: none would guarantee today’s NRSROs any particular level of business, reward accurate performance through bonus business or provide an investor-controlled signal of credibility. Nonetheless, most NRSROs are concerned that the allocating function would do more harm than good, though all would appear to support the rating function. While the Equal Access Rule has broader support than the Franken Proposal, such support lacks a strong foundation since this rule, like the other alternatives proposed,

285 See Comments on Solicitation, supra note 187.
would not promote accurate and reliable ratings or work to prevent NRSRO Nullification without being paired with a rating feature. While it cannot be known if the adoption of any of these alternatives in place of the Franken Proposal would cause a tipping point towards NRSRO Nullification, all would fall short of the Franken Proposal in preventing it.

2. The New Standards of Creditworthiness

Since the SEC’s proposals for the new standards of creditworthiness do not require, but only permit, reliance on NRSRO credit ratings as part of a broader professional judgment analysis, the SEC’s proposed new standards would succeed in eliminating the NRSROs’ power to license under the federal securities laws since the NRSROs would no longer have a mandate to determine creditworthiness. Instead, their ratings would become one factor, if a factor at all, in an investment fiduciary’s separate credit determination. This would effectively shift the responsibility for such credit determinations from the NRSROs (to the extent they were previously responsible) to the appropriate investment fiduciary in each case. Although some market participants have argued that the federal agencies would be defying Congress’s intent by incorporating any new form of reliance on NRSRO credit ratings, voluntary reliance by fiduciaries is not prohibited. Congress’s mandate would only prevent the SEC from incorporating a “requirement of reliance.” \footnote{286} Since requiring reliance would violate Congress’s intent by effectively reinstating regulatory licenses through the back door, I analyze the federal agencies’ three other options with respect to the incorporation of NRSRO credit ratings into the new standards of creditworthiness: prohibiting reliance on NRSRO credit ratings, permitting reliance on only NRSRO credit ratings, and permitting reliance on any credit ratings.

The first option, advocated by some, would be to prohibit reliance on NRSRO credit ratings altogether under the new standards of creditworthiness. \footnote{287} While the federal agencies could hypothetically prohibit such reliance, they would need to have an adequate substitute to replace credit ratings. While the Free Market Camp has put forward some creative possibilities, such as thirty-to-ninety day rolling averages of

\footnote{287} See Better Markets Letter, supra note 194, at 7.
credit spreads and credit-default-swap spreads, a chorus of commentators appears to agree that such market measures would not be a sufficient replacement. This likely explains the shift in the Free Market Camp to a professional judgment analysis that would include permissive partial reliance on both market measures and credit ratings.

The second option would be to permit voluntary reliance on only NRSRO credit ratings as part of a professional judgment analysis. This would raise the question of whether such reliance should be limited to NRSRO credit ratings or should apply equally to all credit ratings. If such reliance were limited to NRSRO credit ratings only, this would provide a significant benefit for the NRSROs, because investment fiduciaries seeking to rely in part on credit ratings would be limited to only the NRSRO pool. On the downside, however, such a policy would discourage upstart CRAs that might have the potential to outperform today’s NRSROs. In effect, it would reinstate the chicken-and-egg problem that characterized the NRSRO designation process prior to 2006. In addition, this solution might unduly limit the ability of fiduciaries to fulfill their legal duties in accordance with their own professional judgment. On the one hand, investment fiduciaries would be told to take ultimate responsibility for their investment decisions; on the other hand, to the extent they rely on ratings, they would be told to only rely on NRSRO ratings. This would seem inconsistent with the new law’s approach of heightening the investment fiduciary’s responsibility. While this solution would work to prevent NRSRO Nullification, it would not be in sync with the alternative answer to the rating-agency question: professional judgment.

The third option would permit voluntary reliance on any credit rating as part of a professional judgment analysis. This would achieve the law’s purpose of placing greater responsibility on the professional judgment of investment fiduciaries and less on NRSROs. By not limiting fiduciaries to

---

288 See supra note 75 and accompanying text.
289 See, e.g., Hill, supra note 17, at 143-44 (“Something else is to replace these references. . . . The problem is that there is no ready alternative.”); see also Coffee, supra note 1, at 233 (“Alternatives to credit ratings, such as credit default swap spreads, provide at best only a partial substitute.”); SEC REPORT OF 2003, supra note 9, at 39 (“[T]he volatility of credit spreads, their backward-looking nature, and the fact that their use would be limited to liquid securities, make them an inferior alternative to credit ratings.”).
290 See Partnoy, supra note 67, at 16.
291 See supra note 105 and accompanying text.
only NRSRO ratings, such fiduciaries would be free to ignore credit ratings altogether or rely in part on either NRSRO or non-NRSRO ratings as part of their professional judgment analysis. Since no adequate substitute for ratings appears to exist, most would likely opt to rely in part on the ratings that would help produce the highest return and best support for such fiduciary’s legal duties.

This comports with the SEC’s general proposal, which is to permit fiduciaries to rely in part on any credit ratings, whether or not they are NRSRO credit ratings. This levels the playing field for NRSROs and non-NRSROs by allowing both to compete to produce credit ratings that fiduciaries would want to incorporate into their analyses. This would solve the chicken-and-egg problem and would be in sync with the new law’s approach of heightening the investment fiduciary’s responsibility.

Such an approach, however, has two potential downsides: it might encourage NRSRO Nullification and could lead to certain moral hazards. The former could be a potential problem because this approach would not distinguish between NRSROs and non-NRSROs. To the extent a CRA’s ratings would be equally in demand as an NRSRO or non-NRSRO, there would be no reason to stay regulated. So any solution must counter this incentive. The latter could be a problem since it presents the potential moral hazard that investment fiduciaries may choose to “cover themselves” by relying blindly or purposefully on poor performing NRSROs or non-NRSROs to maximize returns. For example, if a fiduciary, such as a money market fund manager, is required to invest in only safe securities (i.e., highly rated securities), such a fiduciary, in breach of his duties, might buy securities with higher yields that it believes or should believe are riskier so long as they have the desired safety rating given by some CRA. In good times, if such securities do not default, a fund manager would appear to be outperforming his competitors. In bad times, however, such a fund manager might experience large losses and seek to deflect blame under the cover of a “diligent analysis” and partial reliance on credit ratings commensurate with safety. While a professional judgment test is supposed to prevent excessive risk taking by requiring reliance on other factors in addition to credit

---

292 See supra note 196.
293 See Coffee, supra note 1, at 259 (“[T]here is the even more sinister danger that many institutions (in particular, money market funds) wanted inflated ratings so that they could earn the higher returns from riskier securities.”).
ratings, it is not supposed to end risk taking altogether. Proving a breach of fiduciary duty under these circumstances would be very hard since the principals would have no way of checking the professional judgment analysis of their fiduciaries and courts would not want to second-guess private decision making fraught with inherent risks.294

Interestingly, the rating function from the Franken Proposal could provide a potential solution to both NRSRO Nullification and the moral hazard problems described above. To the extent that the rater only evaluated the disclosed performance statistics of NRSROs (and not non-NRSROs) to tell fiduciaries and ultimate investors which NRSROs are the best performers, this would work against NRSRO Nullification by causing such fiduciaries to prefer NRSRO over non-NRSRO ratings since the accuracy of the former would be better known to the market. To the extent non-NRSROs would not be rated, they would not benefit from the rater’s reputational signal. This would work to negate NRSRO Nullification and encourage NRSRO registration since any perceived incentive to leave the regulatory regime would be undermined by the reputational benefit a CRA could get through a rating.

Such a rater could also mitigate the moral hazard of investment fiduciaries choosing to rely on poor performing NRSROs and non-NRSROs to maximize their returns. Since the rater would be opining about which NRSROs are the best performers, investment fiduciaries choosing to ignore such opinions may strike up suspicion among their ultimate investors. Certifying agreement with the rater’s methodology for defining accuracy and acting in accordance with such opinions, by contrast, would send the opposite message to investors. However, an investment fiduciary eager to show it is not blindly relying on either credit ratings or the rater may

294 Traditionally the business judgment rule has protected the investment decisions of fiduciaries relying on investment grade ratings, though this reliance defense may be weakened now. See id. at 266 (“Today, if a money market fund's board suffers a major loss on an investment, it will very likely be protected by the business judgment rule (and not be held liable) if an NRSRO ratings agency gave the flawed security an investment grade rating.”). But see PARTNOY, supra note 62, at 6 (“The accountability of NRSROs has deteriorated so much that institutional investors now are vulnerable if they rely on credit ratings in making investment decisions. To the extent rating agencies are not subject to liability, an institutional investor's defense of reliance on ratings is weakened, because constituents can argue that ratings are less reliable when rating agencies are not accountable for fraudulent or reckless ratings.”). Nonetheless, so long as they have done some private due diligence in addition to relying on credit ratings and have a rational basis for their decision, proving a breach of the duties of care or loyalty would likely be hard to do.
have good reason to ignore such signal. In the case of disagreement with the rater’s methodology, the SEC could require that investment fiduciaries explain their reasons for relying on ratings from non-NRSROs and ratings from those NRSROs that the rater considers to be poor performers. This would create a market-based countermechanism to check the methodology of the rater and would work against blind reliance on the opinions of such a rater. A requirement that agreement with such methodology must be certified would similarly work against blind reliance. By combining freedom to rely on any ratings with a requirement that agreement with the rater’s methodology for determining accuracy be certified and disagreements with the rater’s methodology or reliance on non-NRSROs be explained, the law’s purpose of heightening the investment fiduciaries’ responsibility would be achieved. It would also mitigate the potential moral hazard because investment fiduciaries would have to justify their reliance on either the rater’s pronouncements or their own dissenting conclusions in order to satisfy their legal duties.

In sum, while the SEC’s proposals for the content of the new standards of creditworthiness would substantially eliminate an important NRSRO-specific financial benefit on the federal level (NRSRO-dependent regulatory licenses), permissible reliance on such ratings as part of a broader professional judgment analysis preserves the opportunity for NRSROs to still capture much of this benefit, even if investment fiduciaries can equally rely on non-NRSRO ratings. Moreover, there appear to be ways to heighten the responsibility of investment fiduciaries without encouraging NRSRO Nullification or otherwise encouraging fiduciaries to take excessive risks that would be contrary to the interests of their principals.

3. Existing Distinctions

At present, there is a sufficient equilibrium between the benefits and burdens of NRSRO status that has kept today’s NRSROs from surrendering their NRSRO status, even though some non-NRSROs have chosen not to join the club as a result of the current balance. In assessing whether the impact of today’s most critical unresolved items will cause a tipping point toward NRSRO Nullification, the existing benefits and burdens of NRSRO status must be taken into account. The two primary NRSRO-specific benefits that remain are reputational, and financial benefits by virtue of the message an NRSRO sends to
the market in being regulated and the persistence of NRSRO-specific regulatory licenses. The Equal Access Rule also provides an informational benefit to NRSROs. In addition, two primary NRSRO-specific burdens also remain in the form of the costs of being regulated and heightened liability exposure as a result of the SEC’s enhanced penalty powers and exposure to private rights of action. While these benefits and burdens appear to sufficiently cancel each other out at the present, they could become increasingly important in concert with the resolution of today’s most critical unresolved items.

First, all ten NRSROs derive an intangible reputational benefit from the message that regulation sends to the market. This benefit, however, would be more valuable to the Big Three than the other seven NRSROs because Ratings Reform was put into place on account of the Big Three’s past performance. Moreover, the Big Three, unlike the other seven, have been subject to widespread negative media attention as a result of their role as major culprits in the lead-up to the financial crisis of 2008. Ratings Reform therefore signals the return of the watchdog, which has likely restored a good deal of investor confidence in the Big Three. Therefore, were any of the Big Three to leave the regulatory regime, this would produce headline news that might lead to a public outcry. Since such an outcry could severely damage or destroy the Big Three’s reputations, it would be unlikely that they would risk NRSRO Nullification even if the costs of remaining regulated were high. Thus, for the Big Three, the reputational benefits of regulation weigh heavily against NRSRO Nullification.

By contrast, Ratings Reform was not put into place for the other seven, nor have they received much criticism or media attention. If any of the other seven opted out of Ratings Reform, the public would be unlikely to react strongly, if at all. The public may not appreciate the important role that the other seven play in mitigating the importance of the Big Three. As a result, the reputational benefits of remaining regulated would weigh considerably less in any calculation favoring NRSRO Nullification by the other seven NRSROs than it would for the Big Three.

Second, all ten NRSROs have the opportunity to derive financial benefits from all pre-Dodd-Frank regulatory licenses through July 2012 and such remaining regulatory licenses (which are mostly on the state and private levels) thereafter. This benefit too, however, is more valuable to the Big Three than the other seven. In spite of the Big Three’s poor
performance during the recent financial crisis, “sticky” market practices suggest that the Big Three will continue to be the primary beneficiaries of such regulatory licenses unless a rater is put in place to send clear signals that better performers should be hired instead. In the absence of a rater or the Franken Proposal’s CRA Board, there would be few new opportunities for the smaller seven to gain market share. To the extent the smaller seven could do comparably good business outside of the regulatory regime, the existence of some fraction of today’s NRSRO-specific regulatory licenses may not stop them from opting out.

Third, all ten NRSROs have the opportunity to benefit from the Equal Access Rule. This rule, however, is meant to favor the smaller seven over the Big Three. Since it provides equal information to non-hired NRSROs, it is supposed to level the playing field for the smaller seven since they are hired less frequently than the Big Three. In reality, it has only provided a marginal benefit, if any benefit, since taking advantage of the rule has not been cost effective.  

Fourth, all ten NRSROs must pay the cost of complying with Dodd-Frank’s heightened regulatory regime, whereas non-NRSROs are exempt from these costs. This burden also applies differently to the Big Three and other seven NRSROs since the Big Three, as bigger companies, have larger economies of scale to afford such expenses. For the smaller NRSROs, such costs are therefore relatively higher. As a result, NRSRO Nullification would be more attractive in respect of the compliance cost savings for the smaller seven NRSROs than for the Big Three. So this factor, too, would more readily drive the other seven to NRSRO Nullification than any of the Big Three.

On balance, the current equilibrium suggests that the seven smaller NRSROs are in a more fragile position than the Big Three with respect to NRSRO Nullification. While many of these seven have fought hard for many years to gain NRSRO status, it was more valuable in the past than it is now. Although a sufficient equilibrium currently exists to keep such NRSROs regulated, the resolution of the most critical unresolved items could tip today’s balance, especially with respect to the seven smallest NRSROs, if the burdens continue to increase and the benefits continue to decrease.

\[\text{See Hill, supra note 17, at 144.}\]
\[\text{See supra note 186.}\]
B. Legal Implications

In addition to the financial and reputational implications discussed above, the NRSROs would not withdraw from the regulatory regime unless it made legal sense. There are two potential legal implications of NRSRO Nullification that may work to deter its exercise: forced registration and regulation under the Advisers Act, and new legislation to force registration and regulation under Ratings Reform.

To the extent that any subset of the NRSROs left the Ratings Reform regime, the SEC could force registration and regulation under the Advisers Act without requiring new legislation. This, however, would be undesirable. First, prior to CRARA, the SEC and the case law both took the position that CRAs were not required to be registered as investment advisers. Second, while CRARA only granted an exemption from the Advisers Act to NRSROs, the SEC has not mandated that any non-NRSROs register under the Advisers Act. Third, the Advisers Act was almost completely ineffective in terms of regulating the NRSROs prior to CRARA. Thus, even though withdrawal from NRSRO status might open the door to capturing such newly minted non-NRSROs under the term “investment adviser,” doing so would not replicate the intended benefits to society of Ratings Reform—accurate and reliable ratings—and therefore would not be desirable.

A second possibility is forced registration and regulation under Ratings Reform. While this would require new legislation, it would be a more desirable option than Advisors Act regulation. Congress could supplement the NRSROs’ voluntary withdrawal right and the CRAs’ voluntary registration right by adopting a mandatory NRSRO registration requirement based on external criteria. Such a requirement could force the CRAs meeting specified criteria to register (or stay registered) and thereby capture such agencies within the web of Ratings Reform. This would mean that NRSROs would not only have to

297 See supra notes 134-36 and accompanying text.
298 Based on multiple investment adviser searches on the SEC website, it appears that non-NRSROs do not register as investment advisers. For example, neither Rapid Ratings, Pacific Credit Ratings, Global Credit Ratings, nor Dagong are registered as investment advisers. The keyword “rating” produces no results corresponding to any of today’s non-NRSROs. See Investment Adviser Search, SEC. & EXCH. COM’N, http://www.adviserinfo.sec.gov/(S(feasnj50jnhzmaljbhw4goio))/IAPD/Content/Search/iapd_Search.aspx (last visited Jan. 18, 2012).
299 See supra notes 135-37 and accompanying text.
300 See supra note 176.
accept the financial, reputational and legal burdens that exist today, but they might also be prevented from avoiding new burdens adopted in the future. This could serve to reverse or prevent NRSRO Nullification.

But how far can the government reach in forcing regulation upon CRAs that are non-NRSROs? Under Congress’s commerce power, it can regulate “the use of the channels of interstate commerce . . . the instrumentalities of interstate commerce, or persons or things in interstate commerce, . . . [and] those activities having a substantial relation to interstate commerce.”

Thus, Congress could not reach (nor would it want to reach) all of the approximately seventy-six global non-NRSROs, since most are foreign entities that operate primarily outside of the jurisdiction of the United States. For those non-NRSROs that rate debt securities traded within the United States, however, Congress would likely have wide latitude in imposing Ratings Reform if it chose to do so, especially since there would be no First Amendment concerns. Whether or not ratings themselves propose commercial transactions, they become indelibly linked with commercial transactions (the sale of debt securities) that form a part of the core of American capitalism. Especially in light of the role of credit ratings in the financial crisis of 2008, it is likely beyond debate that credit ratings substantially affect interstate commerce. Moreover, Congress has explicitly found that “the activities of credit ratings agencies are fundamentally commercial in character.” While the courts do not always afford deference to Congress’s findings with respect to

---

302 For example, Peru’s Pacific Credit Ratings operates primarily in Latin America. See PCR: PAC. CREDIT RATING, http://www.ratingspcr.com/acerca.php (last visited Aug. 14, 2011) (“We operate in Latin America . . . .”). South Africa’s Global Credit Ratings Co. operates primarily in Africa, though it is seeking to expand to other non-U.S. markets. See Global Credit Rating Co., FIN. MKTS. DIRECTORY, http://www.fmd.co.za/data/M01732/M01732.htm (last visited Jan. 18, 2012) (“Having firmly established a market leadership position in Africa, a major thrust has been to establish a similar position in South America, Eastern Europe, Asia and the Middle East, via a combination of acquisitions, alliances, and organic growth.”).
303 Ratings Reform does not reach the content of CRA speech. See supra note 174.
304 Some have argued in the commercial speech context that credit ratings are merely opinions about commercial transactions that do not themselves propose commercial transactions. See, e.g., Letter from Laurence H. Tribe & Thomas C. Goldstein, Legal Consultants, to Moody’s Investors Serv., supra note 123, at 2 (“NRSRO ratings . . . are independent evaluations that do not propose any transaction.”).
commerce, any CRA that rates debt securities traded within the United States would be fighting an uphill battle to challenge prospective legislation on the basis that its ratings are noneconomic activity (i.e., merely speech) that do not substantially affect interstate commerce.

Furthermore, the securities laws already include precedents for such mandatory requirements. For example, every issuer engaged in interstate commerce, in a business affecting interstate commerce, or whose securities are traded by any means or instrumentality of interstate commerce can be captured by the public company reporting rules by the mere existence of two external criteria: total assets exceeding $1 million and a class of equity securities held by five hundred or more holders. Similarly, beneficial owners of equity securities of public companies can be forced to make certain disclosures by merely exceeding a 5 percent ownership threshold.

By analogy, Congress could devise similar legislation based on one of these two precedents to similarly capture CRAs in the web of Ratings Reform based on external criteria. For example, Congress could put into place a ratings market share threshold expressed as a percentage of total outstanding ratings in each ratings class (similar to the 5 percent beneficial ownership threshold) or a fixed number based on total ratings outstanding in each ratings class (similar to the public company reporting threshold). This way, the requirement would only force registration and regulation for those CRAs producing more than a de minimis number of ratings in each class.

Currently, the SEC has identified five different rating classes: financial institutions, insurance companies, corporations, asset-backed securities (i.e., structured finance products), and government, sovereign, and municipal securities. As of year-end 2010, each of the Big Three rated

---

306 See United States v. Morrison, 529 U.S. 598, 617-18 (2000) (limiting Congress’s ability to regulate noneconomic activity based on congressional findings that it created a substantial effect on interstate commerce).


308 See id. § 78m(d).

309 See id.

310 See id. §§ 78l(g), 78m(a).

311 See id. § 78c(a)(62)(B).
more than 1,600 debt securities in each ratings class and held at least 8 percent of the market share of each ratings class. While A.M. Best had approximately 26 percent of the market in insurance company ratings, no other NRSRO had more than 8 percent of the market in any other ratings class and only two other NRSROs (DBRS and Kroll) held more than 7 percent in a different rating class. In each ratings class, however, at least five NRSROs rated one hundred or more debt securities, and at least four NRSROs rated one thousand or more debt securities.

These statistics, however, do not include non-NRSRO ratings and therefore do not reflect the entire universe of CRA credit ratings. Thus, to the extent Congress seeks to adopt such a mandatory NRSRO registration requirement based on external criteria, it would need to gather information about total outstanding ratings in each class from all CRAs engaged in interstate commerce or affecting interstate commerce. From these figures, it could devise a threshold meant to require at least four or five CRAs in each class to register as NRSROs such that there would be sufficient competition and comparative data available in each ratings class to dampen the impact that the Big Three could have on any segment of the market.

Using the figures from the SEC’s most recent summary report on NRSROs and treating them as the entire universe of CRA credit ratings for illustrative purposes, a percentage threshold of outstanding ratings in each class would need to be set at 0.05 percent for each ratings class in order to capture five CRAs in each class and 0.06 percent for each ratings class in order to capture four CRAs in each class. If a fixed number threshold is used instead, one hundred credit ratings in each ratings class would capture five CRAs in each class, while one thousand credit ratings would capture four in each class.

While such a threshold would close the voluntary registration and withdrawal loophole by preventing or reversing NRSRO Nullification, it would not address the underlying goal of Ratings Reform to encourage accurate and reliable ratings. Therefore, to the extent Congress considers adopting a mandatory registration requirement based on external criteria, it must also couple any such measures with a mechanism, such

---

312 See SEC SEPT. 2011 SUMMARY REPORT, supra note 47, at 6-7.
313 Id.
314 Id.
315 See id.
as the rating function of the Franken Proposal’s CRA Board, to promote accurate and reliable credit ratings.

In summation, the most critical items that remain unresolved—the Franken Proposal and the new standards of creditworthiness—in concert with existing distinctions based on NRSRO status could lead to a tipping point in favor of NRSRO Nullification for at least some of today’s NRSROs to the extent these measures continue to increase NRSRO-specific burdens without providing NRSRO-specific benefits. The resolution of these items could also prevent NRSRO Nullification if they provide sufficient benefits to avoid the tipping point. Congress could prevent or reverse NRSRO Nullification by adopting a mandatory NRSRO registration requirement based on external criteria that could force regulation upon a sufficient segment of the CRAs to keep Ratings Reform intact, but this would only be a partial solution since it would not, by itself, promote accurate and reliable ratings.

IV. RECOMMENDATIONS

In this final part, I present my central recommendation for resolving the Franken Proposal and new standards of creditworthiness in consonance with the dual goals of promoting accurate and reliable ratings and preventing NRSRO Nullification: the SEC should adopt a refined version of the rating function suggested by the Franken Proposal—but not the allocating function. In addition, the SEC should adopt its current proposal to permit partial reliance by investment fiduciaries on any credit ratings under the new standards of creditworthiness, subject to one additional requirement: investment fiduciaries seeking to rely on NRSROs that the rater deems good performers must certify agreement with the rater’s methodology for defining accuracy, while investment fiduciaries seeking to rely on non-NRSROs or NRSROs that the rater deems poor performers must publicly explain their disagreement with the rater’s methodology or show why certain non-NRSROs, when compared with NRSROs, produce ratings of equal or better quality. Under this proposal, it will not be necessary to close the voluntary registration loophole through a mandatory registration requirement.
A. Create a Rater of the NRSROs, Not a Rater and an Allocator

My central recommendation is that the SEC should create a rater of the NRSROs, not a rater and an allocator. The Franken proposal, by contrast, would put both functions in place. This would avoid the difficult problems created by the allocating function, which function is generally opposed by the precise constituents it is meant to support—the smaller NRSROs. These problems include how to choose the QNRSRO club’s membership and how to allocate on a meritorious basis without becoming subject to conflicts of interest or choosing a system that the market thinks is wrong. The rating function, on the other hand, which serves the same underlying goals, would likely find widespread acceptance.

To the extent that the allocating function may eventually be appropriate, there is one primary reason that it is currently premature: there is no consensus on how to measure the relative accuracy of one NRSRO’s performance against another’s. The Franken Proposal’s CRA Board, however, would be well positioned to credibly develop a set of such metrics and to test their acceptance in the market. This is because, as a board controlled by investor representatives, it would be naturally aligned with the interests of investment fiduciaries and ultimate investors.

A distinction nonetheless must be drawn between using such metrics to influence the market and using them to make the market. The rating function would have the effect of doing the former, while the allocating function would have the effect of doing the latter. Since it is predictable that some investment

---

316 It must be acknowledged that this recommendation has also been advanced by Professor Lynn Bai and Senators Carl Levin and Tom Coburn. See Bai, supra note 50, at 101 (“[T]he Franken Proposal should be modified in a way such that the primary function of its board would be not to allocate rating jobs for the credit rating industry, but to closely monitor and rank the performances of its players and make this information freely accessible to the investment community.”); see also CARL LEVIN, CHAIRMAN & TOM COBURN, RANKING MINORITY MEMBER, U.S. PERMANENT SUBCOMM. ON INVESTIGATIONS: COMM. ON HOMELAND SEC. & GOV’T AFFAIRS, WALL STREET AND THE FINANCIAL CRISIS: ANATOMY OF A FINANCIAL COLLAPSE 316 (Apr. 13, 2011) available at http://www.ft.com/cms/fc7d55e8-661a-11e0-9d40-00144feab49a.pdf (“The SEC should use its regulatory authority to rank the Nationally Recognized Statistical Rating Organizations in terms of performance, in particular the accuracy of their ratings.”). To this author’s knowledge, the ratings inflation risk premium hypothesized in this article that would motivate issuers to prefer hiring the better performing NRSROs on account of the rater’s pronouncements is a novel rationale for this recommendation.
fiduciaries will disagree with whatever metrics are proposed by the CRA Board, such disagreements should also be given room to be tested in the market. Over time, there may be a convergence over the best available methods for ranking NRSROs based on accurate ratings. Since none yet exists, the allocating function should not yet be adopted.

The rating function, on the other hand, should be adopted because, standing alone, it would nonetheless promote accurate and reliable ratings. It would do this by putting investment fiduciaries in a stronger position to influence the hiring decisions of issuers. Although only the allocating function would directly address the issuer-pays conflict by separating issuer payment of the NRSRO from issuer selection,\(^\text{317}\) the rating function would put information into the market that does not currently exist about which NRSROs are the best performers.\(^\text{318}\) Investment fiduciaries, through underwriters, could use this information during pricing negotiations to penalize issuers for choosing poor performing NRSROs or non-NRSROs. They could demand an interest rate premium to compensate them for ratings inflation risk. To avoid paying such a premium, issuers would come to prefer hiring the most accurate performers since this would undermine any argument that such a ratings inflation risk premium would be appropriate. This would change the game because rating accurately, instead of rating generously, would become good business.

The rating function would also work to prevent NRSRO Nullification so long as the rater only rated NRSROs and not non-NRSROs, took into account all NRSRO ratings (i.e., solicited and unsolicited) and made such accuracy ratings, including the metrics used, publicly available. Such refinements could easily be

\(^\text{317}\) Cf. supra note 189.

\(^\text{318}\) It must be observed that, independent of whether the SEC creates a rater to interpret performance data, it is requiring the disclosure of such data. See supra note 219 and accompanying text. This will allow investors to analyze the same data as any such rater and thereby develop competing accuracy rankings, which is likely a daunting and meticulous task. This author hypothesizes that the interplay between the rater and the market’s check and balance on the rater is the best way to create a consensus over the meaning of relative accuracy. In the absence of a rater and a requirement that investors double-check the rater’s methodologies, investors may not have sufficient motivation to develop in-house views about which CRAs are the best performers in spite of the new disclosure requirements. Even if sufficient motivation exists for investors to compete among themselves to create accuracy rankings, the various conclusions of innumerable investors may be too scattered to meaningfully influence the hiring decisions of issuers. A rater would help focus the debate about which CRAs are the best performers by providing an objective benchmark on behalf of all investors.
built into the Franken Proposal’s current architecture. This would add significant value to becoming an NRSRO because it would allow agencies to credibly distinguish themselves based upon their superior ability to rate accurately. This would produce more market visibility and more business for the best performers. Opting out of NRSRO status, by contrast, would mean losing this important benefit.

Since the other alternatives that the SEC is considering have no mechanism to promote accurate and reliable ratings nor would they work to prevent NRSRO Nullification, a modified form of the Franken Proposal adopting only this refined rating mechanism is justified.

B. Permit Partial Ratings Reliance Under the New Standards

In addition, the SEC should adopt its current proposal to permit partial reliance by investment fiduciaries on any credit ratings under the new standards of creditworthiness subject to one additional requirement: investment fiduciaries seeking to rely on NRSROs that the rater deems good performers must certify agreement with the rater’s methodology for defining accuracy, while investment fiduciaries seeking to rely on non-NRSROs or NRSROs that the rater deems poor performers must publicly explain their disagreement with the rater’s methodology or show why certain non-NRSROs, when compared with NRSROs, produce ratings of equal or better quality.

Partial reliance on credit ratings is necessary because there is no adequate substitute in the market to replace credit ratings. Investment fiduciaries, moreover, should not be limited to considering only NRSRO ratings because this would undermine the law’s purpose in shifting responsibility for determining creditworthiness from the NRSROs to the relevant investment fiduciary in each case. Since the rater I propose would only rate NRSROs, and not non-NRSROs, this limitation should negate any perverse incentives by the NRSROs to opt out of the regulatory regime simply because investment fiduciaries would be permitted to rely on non-NRSRO ratings.

My proposal would also promote accurate and reliable ratings. By requiring investment fiduciaries to certify agreement with the rater’s methodology for defining accuracy or to explain disagreements with the rater’s methodology or reliance on non-NRSROs, a countermechanism would be introduced into the market to check such rater’s pronouncements. This would directly
address the Franken Proposal’s most serious flaw: the allocation of ratings business based on an accuracy standard that the market does not accept. This “certify or explain” rule would harness the professional judgment of investment fiduciaries in the service of accuracy. While there is currently no accepted definition of accuracy in the market, the collective effort of the rater and the community of investment fiduciaries will help bring one into existence. At the same time, it will work against the moral hazards that investment fiduciaries may be relying blindly on the opinions of the rater or ignoring its signals with impunity.

C. Consider Closing the Loophole

Finally, given this article’s focus on the threat posed by the NRSROs’ voluntary withdrawal right, it is necessary to consider whether this voluntary registration and withdrawal loophole should simply be closed. While I assess this possibility elsewhere, under the central recommendation I advance, closing the voluntary registration and withdrawal loophole by adopting a mandatory registration requirement is not necessary because a rater will produce sufficient financial and reputational benefits for today’s NRSROs to prevent NRSRO Nullification. While closing the loophole would be an even more definitive way of preventing NRSRO Nullification, it would not work to promote accurate and reliable ratings by itself and would therefore only be a partial solution. In addition, closing the loophole would require new legislation by Congress, whereas my central recommendation only requires action on the part of the SEC. There is value, in any event, to leaving this particular loophole open: it sends a signal to the NRSROs to stay voluntarily regulated because mandatory regulation is always looming as an option.

319 See supra Part III.A.3.
INTRODUCTION

Clarence Ray Allen was seventy-six years old when he was executed in California in 2006. The State of California disputed Allen’s claims that he was handicapped and greatly diminished by the infirmities of age. He used a wheelchair, had endured both a heart attack and a stroke, was diabetic, and claimed to be legally blind. The press coverage of his execution made prominent mention of the fact that Allen did indeed walk to the death chamber although supported by guards, inviting the inference that he was a malingerer and had exaggerated the toll of age and ill health. A deeper look into the record reveals that he was escorted to the death chamber by four burly guards, with whose support a paraplegic could also have covered the short distance without other aid.

The United States has a growing elderly death row population; they are beginning to trickle into the execution chamber. The Supreme Court has several times rebuffed efforts
to gain Eighth Amendment protection from execution for the long-serving condemned irrespective of their age, denying certiorari to the Lackey claim. The Lackey claim urges that the combination of long confinement in anticipation of death and execution constitutes cruel and unusual punishment prohibited by the Eighth Amendment. The chilly reception of the Lackey claim by the Supreme Court is best explained not by its lack of merit but rather by the devastating impact its recognition would have on capital punishment. Execution in the United States follows condemnation on average by more than a dozen years. Hundreds of death row inmates have not had their cases finally resolved twenty and thirty years after sentences were pronounced. The ranks of the long serving are steadily growing. Recognizing the Lackey claim would take the United States a long way down the road to abolition. The Supreme Court has been inhospitable to total abolition but willing to reform capital punishment by trimming back the types of crimes and criminals eligible for capital punishment. Unlike the general Lackey claim, Lackey-for-the-Elderly is another such modest reform. Lackey-for-the-Elderly is therefore more likely to succeed than the wider claim. Its adoption would bring an end to a practice that the Eighth Amendment ought not tolerate. It would spare us the spectacle of the elderly being carried or wheeled to the execution chamber after decades of growing old in death row confinement in the name of American justice.

My exploration of the case for an Eighth Amendment bar against executing the long-serving elderly will begin with a review of the representation of the elderly on America’s death rows and a survey of the very limited avenues of relief currently available to them on the basis of age. I will then discuss the attribution problem by asking at whose door should “fault” for long delays between condemnation and consummation of a capital sentence be laid—the prisoner, the state, or the working through of due process? For many jurists, attribution of fault is critical to resolving the question of whether the long serving of any age should be permitted to exit death row alive. I will then argue that the long-serving elderly should be relieved of both death row confinement and the continuing threat of execution.
I. **LACKEY-FOR-THE-ELDERLY: A PROPOSAL FOR AN EIGHTH AMENDMENT BAR TAILORED FOR ELDERLY DEATH ROW INMATES**

A prohibition against execution of the long-serving elderly would amount to an age-specific version of the *Lackey* claim. The *Lackey* claim takes its name from the eponymous Clarence Lackey. In *Lackey v. Texas*, Lackey argued that after seventeen years on death row his execution would be cruel and unusual punishment forbidden by the Eighth Amendment.\(^4\) The Supreme Court denied his petition for certiorari in 1995. Justice Stevens wrote a memorandum to the denial of certiorari stating that the question Lackey raised warranted review, but should be allowed to percolate through the lower courts before certiorari was granted.\(^7\) Since *Lackey*, the Supreme Court has rebuffed a handful of similar petitions over dissents from Justice Breyer from the denial of certiorari and renewed statements by Justice Stevens that the Court should in due course grant a *Lackey* petition and consider the issue on its merits.\(^8\) The *Lackey* claim, although a staple in end-game capital litigation, has yet to prevail in any U.S. court under the Eighth Amendment or its analogs in the constitutions of the thirty-four states that retain capital punishment.\(^9\) The *Lackey* claim may well roil both abolitionists and retentionists. Abolitionists may fear a time limit would portend a rush to execute before the deadline. Retentionists may fear that a time limit would be so great a curb on executions as to amount to abolition, given a national average in excess of twelve years to accomplish execution. The apprehension of the two camps will be addressed in this article.

---

\(^7\) *Id.*
\(^9\) In *Knight v. Florida*, Justice Thomas notes that federal and state courts considering *Lackey* claims since *Lackey v. Texas* “have resoundingly rejected the claim.” *Knight*, 528 U.S. at 992 (Thomas, J., concurring). *But see People v. Anderson*, 493 P.2d 880, 894-95 (Cal. 1972) (holding capital punishment to be unconstitutional under the California Constitution in part because of delays in carrying out sentences).
Lackey claims ripen when a prisoner confronts an execution date. The Lackey petitioner then argues that the imposition of execution on the heels of the long death row confinement constitutes excessive punishment prohibited by the Eighth Amendment. These petitions also urge that long-term death row confinement may itself constitute cruel punishment. Justice Stevens and Justice Breyer are clearly in sympathy with both theses. Some international courts and the constitutional courts of some nations have held that a lengthy period awaiting execution constitutes cruelty and have required that prisoners be spared exposure to extended periods of time under sentence of death quite apart from the imminence of the threat of execution. Because of their inevitable frailties, constitutional questions about long-term confinement and subsequent execution arise in an acute form when we focus on the aged of death row. No attempt will be made here to specify precisely the geriatric threshold that would trigger the protection of a Lackey-for-the-Elderly regime—whether chronological age or deterioration and vulnerability associated with aging processes. Such practical considerations can be left for the time at which the Eighth Amendment questions raised here have gained the ear of the American bench.

The question of whether the Eighth Amendment should afford shelter to the aged of death row can be parsed into two related inquiries. First, is it prohibited cruelty to confine persons beyond a certain chronological age, or those exhibiting the deterioration associated with old age, in death row conditions? And second, is it prohibited cruelty to execute aged condemned prisoners? Of these two issues, the first may perhaps be more readily acknowledged as raising a valid Eighth Amendment issue: whatever conclusion one might ultimately reach on the question, the proposition that death row conditions exact a greater toll on the physically and mentally frail or infirm does not appear controversial. I will argue that the Eighth Amendment should be construed to

10 Ceja v. Stewart, 134 F.3d 1368, 1371 (9th Cir. 1998) (Fletcher, J., dissenting).
relieve the elderly condemned from both death row incarceration and the long-serving elderly condemned from the continued threat of execution—thus, Lackey-for-the-Elderly.

II. THE AGED OF DEATH ROW BY THE NUMBERS

It will be useful to begin by examining the facts and statistics that reveal the extent of the elderly population on death row and the reasons for its continuing growth. This exercise will demonstrate that there are good reasons to expect that courts and prison administrations will be confronting what to do with the aged condemned not as an occasional anomaly but as a recurrent issue in death-penalty law and practice.

Clarence Ray Allen was the fourth septuagenarian to be executed since 2004.\footnote{The other over-seventy prisoners executed were James Hubbard, age seventy-four, executed August 5, 2004, in Alabama; John Nixon, age seventy-seven, executed December 14, 2005, in Mississippi; and John Boltz, age seventy-four, executed June 1, 2006, in Oklahoma. \textit{Death Penalty Information Center, Execution Database, Death Penalty Info. Center,} \url{http://www.deathpenaltyinfo.org/executions} (last visited Apr. 16, 2011).} A fifth was executed in 2010.\footnote{Gerald Holland, age seventy-two, was executed in Mississippi on May 20, 2010. \textit{Id.}} Three men sixty-five or older have also been executed since 2002.\footnote{Id.} Prior to 2002 only one person sixty-five or older, a sixty-six-year-old, had been executed in the entire history of the modern death-penalty era commencing with the reinstatement of capital punishment in 1976. That execution occurred in 1984.\footnote{Id.}

These executions of persons in their late sixties and seventies reflect the graying of death row. In 2000, only 2.3 percent of death row prisoners were sixty or older; 11.1 percent were fifty to fifty-nine.\footnote{\textit{Bureau of Justice Statistics, U.S. Dep’t of Justice, Capital Punishment 1999,} at 9 tbl.8 (2000) [hereinafter \textit{Capital Punishment 1999}], \url{available at http://bjs.ojp.usdoj.gov/content/pub/pdf/cp99.pdf}.} In 2009, 2.6 percent were sixty-five or older and 5.6 percent between sixty and sixty-four; 21.1 percent were age fifty to fifty-nine.\footnote{Id.}
Table 1: Percentage of Death Row Prisoners over Age Fifty in the United States

![Bar chart showing percentage of death row prisoners over age 50 in 1999 and 2009.]

Table 2: Number of Death Row Prisoners over Age Fifty in the United States

![Bar chart showing number of death row prisoners over age 50 in 1999 and 2009.]

---


40 Capital Punishment 2009, supra note 18, at tbl.5.
The reason for this change is not a wave of capital crimes by the elderly. Rather, it is one awkward consequence of the contemporary capital punishment regime. A capital sentence marks the beginning of a potentially decades-long incarceration. Final resolution of capital cases—whether by execution or sentence reduction and removal to general prison population—can take decades, and for many long-serving inmates, resolution has not yet come. In this system, a modest 15 percent of persons sentenced to die since 1977 have been executed, while more than 40 percent of those sent to death row at any time from 1977 forward are still there and growing older.

### Table 3: Average Elapsed Time from Sentence to Execution (Years)

<table>
<thead>
<tr>
<th>Year</th>
<th>Time on Death Row</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>14</td>
</tr>
<tr>
<td>1999</td>
<td>12</td>
</tr>
<tr>
<td>2000</td>
<td>10</td>
</tr>
<tr>
<td>2001</td>
<td>8</td>
</tr>
<tr>
<td>2002</td>
<td>6</td>
</tr>
<tr>
<td>2003</td>
<td>4</td>
</tr>
<tr>
<td>2004</td>
<td>2</td>
</tr>
<tr>
<td>2005</td>
<td>0</td>
</tr>
<tr>
<td>2006</td>
<td>2</td>
</tr>
<tr>
<td>2007</td>
<td>4</td>
</tr>
<tr>
<td>2008</td>
<td>6</td>
</tr>
<tr>
<td>2009</td>
<td>8</td>
</tr>
</tbody>
</table>

---

21. *Id.* This table shows that while 29.3 percent of prisoners on death row in 2009 were over fifty, these prisoners represented only 14.4 percent of the additions to death row that year.

22. *Id.* at 12 tbl.9 & 14 tbl.12. Table 9 shows that in 2009, the average death row prisoner had been on death row for 152 months, or approximately 12.6 years. Table 12 shows that the average time between sentencing and execution in 2009 was 169 months, or 14.1 years.

23. *Id.* at 16 tbl.14.

Almost all the states that lead the country in number and frequency of executions as well as the less execution-prone death-penalty states have elderly death row inmates. These inmates include people condemned in their fifties and sixties, but the majority of these older death row inmates have been on death rows for twenty, twenty-five, and thirty years. Thus, state and federal courts and clemency authorities can expect to confront the issue of whether to proceed with the execution of aged inmates long incarcerated on death row with increasing frequency in the years to come.

Old age comes early to prison populations, because the population is unhealthy at entry and prison conditions are generally inimical to physical and mental health. In an era when “sixty is the new forty” for free Americans, prison health experts treat the onset of old age in prison as commencing as early as fifty. No legislature has enacted capital punishment as a sentence expressly to include decades spent ripening into old age on death row, but death sentences in almost every death-penalty jurisdiction now have that potential—and for a growing number of inmates, that reality. Relative to the conditions of the general prison population, the more rigorous conditions of confinement on death row take a greater toll on

---


26 As of December 31, 2009, 196 of California’s 684 death row prisoners (or approximately 28.7 percent) had been under sentence of death for more than twenty years. In Florida, 117 out of 389 (or approximately 30 percent) death row prisoners had been on death row for more than twenty years. CAPITAL PUNISHMENT 2009, supra note 18, at 19 tbl.18.


the minds and bodies of these inmates and therefore exacerbate the decline and distress of older prisoners. These conditions include severely limited opportunities for exercise or work, social isolation, lack of stimulation and contact with the world beyond prison gates, and, in particular, lack of contact with family and friends. To these conditions must be added the background condition of living with the prospect of execution.

III. CURRENTLY AVAILABLE RELIEF

An aged death row inmate has at the present time two potential avenues of relief that address his or her age. If he is far into senility, he can argue—in the parlance of the common law—that he is no longer of sound memory. *Ford v. Wainwright* constitutionalized this common law rule by holding that the execution of a person incapable of understanding that he is being executed for committing a heinous crime is forbidden by the Eighth Amendment. *Ford* was argued in habeas petitions by two of the executed septuagenarians. Although a *Ford* claim has yet to prevail, if States continue to execute persons of such advanced age, some *Ford* claims will eventually succeed. The Ninth Circuit Court of Appeals offers a second avenue of relief, available to death row prisoners befuddled by age but not lost in advanced senility. In *Rohan ex rel. Gates v. Woodford*, the court recognized a statutory right to be competent to assist counsel in capital habeas cases. A death row inmate who cannot rationally communicate with habeas counsel may stay habeas proceedings and execution. The

---

32 Kate McMahon, Dead Man Waiting: Death Row Delays, the Eighth Amendment and What Courts and Legislatures Can Do, 25 BUFF. PUB. INT. L.J. 43 (2007) (discussing “death row phenomenon”).
34 Clarence Ray Allen and James Hubbard both presented this argument in their habeas corpus petitions. See *Allen v. Ornoski*, 435 F.3d 946 (9th Cir. 2006); *Hubbard v. Campbell*, 379 F.3d 1245 (11th. Cir. 2004).
35 *Rohan ex rel. Gates v. Woodford*, 334 F.3d 803 (9th Cir. 2003).
Rohan decision makes it possible to remove some elderly from death row before a Ford claim would be ripe and on a less demanding standard. In his opinion in Rohan, Judge Kozinski in effect reduced the standard on which a prisoner can stay execution from the very demanding Ford standard to the Dusky standard of incompetent to assist counsel. 36 The stay would necessarily be permanent in the case of a habeas petitioner with irreversible loss of mental function. 37 For example, Leroy Nash, the oldest inmate on death row at ninety-four years old until his death in 2010, won a stay when the court of appeals extended the right to be competent to assist counsel in capital habeas cases to appeals from denials of habeas. 38 In Nash v. Ryan, the Ninth Circuit Court of Appeals clarified the standard to be applied, namely, “whether rational communication with the petitioner is essential to counsel’s ability to meaningfully prosecute an appeal.” 39 Leroy Nash subsequently left death row and died in a medical facility. 40 Nash may be the first death row inmate to exit death row because he was too old to execute.

However, both Ford and Rohan can afford relief only to the elderly seriously compromised in their mental function. What of the more mentally fit elderly marking off the decades on death row? What are the prospects for a constitutional categorical exclusion of the elderly from execution after long incarcerations on death row? 41

36 Id. at 816-17.
37 To date, no other federal circuit has either followed or rejected Rohan. However, in Holmes v. Buss, Judge Posner bolstered the Rohan analysis although he found it unnecessary to either accept or reject Rohan because its “validity has throughout these proceedings been assumed rather than litigated.” 506 F.3d 576, 578 (7th Cir. 2007). He elaborated Judge Kozinski’s defense of the standard of rational communication with counsel as appropriate in habeas, and discussed prosecutorial misconduct and ineffective assistance of counsel as examples of where a lay defendant’s recollections may well assist habeas counsel. Id. at 579-80.
38 Nash v. Ryan, 581 F.3d 1048 (9th Cir. 2009).
39 Id. at 1054.
41 The exclusion under discussion would not bar capital adjudication of persons who kill at an advanced age. Such a bar would be analogous to that recognized for youthful murderers in Roper v. Simmons. Rather, the nature of the exclusion explored in this article is analogous to that in Ford v. Wainwright prohibiting the execution of a prisoner convicted and capitally sentenced who subsequently becomes incompetent to execute by virtue of insanity. Unlike the Ford prohibition, there would be no possibility that the aged prisoner would be restored to fitness and executable, unless medical advances allow the processes of aging to be reversed.
IV. THE ATTRIBUTION PROBLEM: WHO IS RESPONSIBLE FOR DELAY IN CONSUMMATING EXECUTIONS?

A proponent of any type of Lackey claim must address the attribution problem that has bedeviled discussions of the issue (for death row prisoners of any age) in domestic, foreign, and international courts; for many readers attribution would be the essential starting point of a Lackey discussion, if not the heart of the matter. At whose door should responsibility for delay be placed? Should it be charged to legal maneuvering by the prisoner to delay execution of a sentence, the normal course-of-review processes, or dilatory (or worse) conduct by the state? Judges throughout the world have wrestled with the attribution question. They have differed sharply as to the proper attribution of responsibility for delay and the inferences to be drawn as to whether the human or constitutional rights of prisoners have been violated. Justice Stevens took the position on attribution in Lackey—to which he and Justice Breyer have subsequently adhered—that prisoners should not be held responsible for delays caused by state “negligence or deliberate action,” or “a petitioner’s legitimate exercise of his right of review.” However, these justices would debit delay caused by “abuse of the judicial system by escape or repetitive, frivolous filings” in calculating time on death row that may require relief.

Justice Thomas rejected this analysis in his concurrences in the Lackey cases, in which a spirited serial debate has transpired between pro-Lackey Justices Stevens and Breyer and anti-Lackey Justice Thomas. Justice Thomas is among those judges who lose patience when asked to consider delay as cruelty visited on prisoners that may support an Eighth Amendment claim. Particularly when there has been a confession, Justice Thomas is disposed to argue that the prisoner could avoid delay by submitting to his sentence. Perhaps the emphasis on conceded guilt relieves the justice of any acknowledged need to consider the weight of delay in supporting a petition when the State has been negligent or

---

43 Id.
44 Thompson v. McNeil, 129 S. Ct. 1299, 1301 (2009); Foster v. Florida, 537 U.S. 990, 991 (2002). Having noted that Foster was a confessed murderer who could have avoided delay by submitting to his sentence, Justice Thomas adds, “Moreover, this judgment [on Foster by the people of Florida] would not have been made had petitioner not slit Julian Lanier’s throat, dragged him into the bushes, and then, when petitioner realized he could hear Lanier breathing, cut his spine.” Foster, 123 S. Ct. at 471.
deliberately caused delay; at any rate, Justice Thomas did not address delay attributable to the state. He did, however, address delay caused by time for the judicial review available to a death row inmate to run its course: Thomas found a “mockery of our system of justice . . . for a convicted murderer”\(^\text{45}\) to claim delay “renders his sentence unconstitutional”\(^\text{46}\) when the postponement is a result of “his own interminable efforts of delay.”\(^\text{47}\) He noted, “I remain ‘unaware of any support in the American constitutional tradition or in this Court’s precedent for the proposition that a defendant can avail himself of a panoply of appellate and collateral procedures and then complain when his execution is delayed.’”\(^\text{48}\) There is little doubt that if the issue is to be resolved by Eighth Amendment precedent, Justice Thomas’s position prevails. Further, for those who like him see only the machinations of heinous criminals and abolitionists’ interminable ploys, the response to the suggestion that delay compromises the constitutional integrity of the sentence is moral outrage.\(^\text{49}\)

Justices Stevens and Breyer treat state-caused delay as building towards some number of years, which, if exceeded, ought to constitutionally prohibit execution. Justice Stevens points to the “staggering”\(^\text{50}\) error rate in capital trials, more than “30 percent of death verdicts overturned.”\(^\text{51}\) Justice Breyer takes issue with Justice Thomas for failing to distinguish between delay caused by “constitutionally defective death-penalty procedures for which petitioner was not responsible”\(^\text{52}\) and delay that is petitioner’s “fault.”\(^\text{53}\)

Neither “delay” nor “fault” attributable to the actions of either state actors or the defendant and his abolitionist lawyers adequately explains the dozen and more years it takes on average to deliver the condemned to the execution chamber in the United States. The complexity of the system and the demands for legal and judicial resources to carry cases through

---

\(^{45}\) Thompson, 129 S. Ct. at 1301 (Thomas, J., concurring in denial of certiorari) (quoting Turner v. Jabe, 58 F.3d 924, 933 (4th. Cir. 1995) (Luttig, J., concurring)).

\(^{46}\) Id.

\(^{47}\) Id.

\(^{48}\) Id. (quoting Knight v. Florida, 528 U.S. 990 (1999)).

\(^{49}\) Turner, 58 F.3d at 933 (Luttig, J., concurring) (quoted with approval by Thomas in Thompson and in Knight; has both criminals and abolitionists in his sights when he characterizes the defendant's Lackey claim as "an affront to lawabiding citizens").

\(^{50}\) Thompson, 129 S. Ct. at 1300.

\(^{51}\) Id.

\(^{52}\) Id. at 1303 (Breyer, J., dissenting).

\(^{53}\) Id. (Thomas, J., concurring).
the many appellate and postconviction stages are “responsible” for as much or more delay within the system taken as a whole than the machinations, misconduct, or errors of actors within the system. “Delay” is not the most apt term for the years between sentence and execution in many retentionist countries across the globe; with the benefit of contemporary standards of due process in death cases, retentionist countries cannot execute in “days or weeks” as our forbearers did, but many years or even decades after conviction and sentence. The most salient factor is unlikely to be delay attributed wholly to the defendant or the state but “delay” attributed to the complexity of, and resource limitations in, the death-penalty review system. Thus, the response of a judge to time consumed by the operation of the many tiered capital punishment regime is especially important in his or her analysis of the attribution problem.

Justice Thomas attributes what could be termed “systemic delay” to prisoners in their efforts to delay or prevent execution. It would be a “mockery of justice” from the perspective of many retentionists to allow prisoners who have failed again and again to persuade courts of errors in their convictions or sentences to then claim immunity from execution. The logic of this position was captured by the Ninth Circuit in Judge O'Scannlain's opinion in McKenzie v. Ray: “It would indeed be a mockery of justice if the delay incurred during the prosecution of claims that fail on the merits could itself accrue into a substantive claim to the very relief that had been sought and properly denied in the first place.” Justice La Forest made the same point powerfully in Kindler v. Canada: “It would be ironic if delay caused by the appellant’s taking advantage of the full and generous avenue of the appeals available to him should be viewed as a violation of fundamental justice.”

54 The delays in the appellate system have been discussed by many authors. In his 2007 law review article, Judge Arthur Alarcon points to twenty institutionally created delays in the postconviction process, including delays in appellate and habeas counsel, delays in reporter transcription, and delays in the issuance of orders and certifications. Arthur L. Alarcon, Remedies for California’s Death Row Deadlock, 80 S. CAL. L. REV. 697 (2007).


56 Knight, 528 U.S. at 993 (Thomas, J., concurring).

57 McKenzie v. Day, 57 F.3d 1493, 1494 (9th Cir. 1995).

58 Knight, 528 U.S. at 998 (quoting Kindler v. Minister of Justice, [1991] 2 S.C.R. 779, 779, 838 (Can.).)
However, what some see as reprehensible maneuvering others see as legitimate appeals to test conviction and sentence. Thus, Justice Breyer explains, “I do not believe that a petitioner’s decision to exercise his right to seek appellate review of his death sentence automatically waives a claim that the Eighth Amendment proscribes a delay of more than 30 years.”

Judges who regard contemporary enhanced due process standards of review of capital cases favorably are also inclined to regard the efforts of prisoners to avail themselves of these processes as manifestations of human nature, the will to live. In *Soering v. United Kingdom*, the European Court of Human Rights opines that, “just as some lapse of time between sentence and execution is inevitable if appeal safeguards are to be provided to the condemned person, so it is equally part of human nature that the person will cling to life by exploiting those safeguards to the full.”

Mde. Christine Chanet quotes this passage from *Soering* to support her dissent in *Barrett and Sutcliffe v. Jamaica*, in which she wrote, “Without being at all cynical, I consider that the author cannot be expected to hurry up in making appeals so that he can be executed more rapidly.”

Judges have observed that the very human desire to live renders a prolonged period in which a prisoner contests his sentence intolerably inhumane. Thus, Lord Scarman and Lord Brightman state in their *Riley v. Attorney General of Jamaica* dissent, “In truth, it is this ineradicable human desire which makes prolongation inhuman and degrading.”

Agreeing with the dissenters in *Riley* that it is human nature to fight to live, the Supreme Court of India repudiates the attribution question itself:

> We think that the cause of delay is immaterial when the sentence is death. Be the cause for the delay, the time necessary for appeal and consideration of reprieve or some other cause for which the accused himself may be responsible, it would not alter the dehumanizing character of the delay.

At first blush, the presentation of the attribution issue by Justice Stevens in *Lackey* suggests that a calculus might be employed by judges to determine whether, in a given case, the period of death row incarceration had exceeded the humane

---

limit. However, because—like the rest of us—judges tend either to attribute systemic delay to prisoners or decline to do so, there is no neutral attribution calculus for deciding *Lackey* petitions. In the bulk of cases, the delay attributable to egregious conduct on the part of the state or the prisoner will be dwarfed by that attributable to the operation of the contemporary death-penalty system. The attribution question devolves into a choice between two views of the condemned in the toils of the contemporary death penalty with its robust due process protections: the manipulator, who showed no pity in taking life and now parades his own suffering; or the condemned human being compelled by human nature to fight off annihilation year on year, in degradation of the law and his own humanity. The choice between these two views must fall to the voting strength of their adherents on constitutional courts or in legislative chambers. The same fault lines of judicial orientation on this issue are observable among U.S. jurists and those serving on foreign and international courts. While the U.S. Supreme Court is famously divided on the question of whether foreign and international judicial opinions should have even persuasive force in U.S. constitutional law, there are among these extranational sources opinions that resonate for both pro- and anti-*Lackey* justices, should they choose to pay heed to them.

As for the elderly of death row—the particular subject of this article—within the terms of the attribution debate, their circumstances are much like those of other long-serving death row prisoners. Whatever the quantity or quality of their distress—like their younger peers—some will see them as manipulators and others as suffering intolerably inhumane treatment. There is, however, one salient difference in that at least some of the elderly may no longer be capable of strategizing. Thus, if charged with any delay beyond that point, it would result from the fact that they are constructively accountable for the wiles of their attorneys.

---

64 For example, in *Roper v. Simmons*, holding that murderers less than eighteen years of age when they killed are categorically exempt from capital punishment, Justice Kennedy, writing for himself and four other justices, relied extensively on the persuasive force of foreign and international opinions. 543 U.S. 551, 567-68, 575-79 (2005). Justice Scalia vehemently contested this reliance in a dissent joined by Chief Justice Rehnquist and Justice Thomas. *Id.* at 622-28 (Scalia, J., dissenting). Justice O'Connor, who dissented on separate grounds, made no objection to the majority's reliance on offshore law. *Id.* at 604 (O'Connor, J., dissenting).
V. The Eighth Amendment Does Not Tolerate Death Row Confinement for Elderly Prisoners—Or Won’t Soon

Let us turn to the first question raised by the proposal for Lackey-for-the-Elderly: should the Eighth Amendment be understood to forbid death row incarceration of persons who reach the stipulated age or exhibit the infirmities and limitations of old age?

The long confinement of prisoners on death row is a historical novelty. Traditionally, a prisoner was sent to death row in transit to eternity; death row stays were measured in days, weeks, and months, not decades. The prisoner who had no earthly future was to devote the death row interlude to making peace with his or her Maker as best he or she could. The justification for the bleak and barren nature of death row confinement is therefore linked to the liminal status of the condemned; their needs resembled those dying in freedom more than prisoners serving terms of years. The transitory nature of death row is no longer available as a justification for confinement in death row conditions for persons of any age: the transition justification for imposing these conditions cannot be sustained for twenty or thirty years.

Contemporary prison administrators, however, like their predecessors in earlier periods, confront security challenges in managing death row populations: there are dangerous and desperate inmates among their charges. Death row conditions are justified as necessary to provide security dealing with dangerous prisoners with little to lose. The Eighth Amendment’s cruel and unusual punishment ban has been consistently construed to afford prison officials wide discretion to achieve security, order, and discipline and to offer little relief.

---

65 By the mid-twentieth century, prior to the reforms initiated in 1976 in Gregg v. Georgia, executions took place within months or (a very few) years after condemnation. Of the 227 executions carried out in the United States from 1956 to 1959, almost half (107) occurred within one year of the sentence—thirty-eight within six months—and two-thirds occurred within two years of the sentence. Only fifteen executions occurred more than four years after the sentence. WILLIAM LUNDEN, BD. OF CONTROL OF STATE INSTS., THE DEATH PENALTY, AN ANALYSIS OF CAPITAL PUNISHMENT AND FACTORS RELATED TO MURDER 12 (1960).

66 Many jurisdictions prohibit death row inmates from contact visits. They are not given access to educational or occupational training. Death row inmates spend between twenty-one and twenty-three hours a day in their cells (most of which have solid doors that impede human contact), and many are not given access to any group recreation time. See Death Row Facts, DEATH PENALTY INFO. CENTER, http://www.deathpenaltyinfo.org/death-row (follow “Death Row Conditions” hyperlink) (last visited Mar. 26, 2011).
to prisoners contesting prison conditions in light of these security considerations. To be sure, the Eighth Amendment does protect prisoners from inhumane treatment and conditions. However, in *Wilson v. Seiter*, the Court recognized that, while the Eighth Amendment forbad serious deprivation of prisoners' basic human needs, it did so only when prison personnel acted with deliberate indifference to those needs. *Wilson* presents a severely circumscribed notion of what constitutes the basic needs of prisoners: "food, warmth or exercise" and medical care.

In addition to the narrowness of the conception of human need and the burden of establishing that prison personnel were at least reckless in contemplating the deprivation, the Court has rendered any prisoner's efforts to redress prison conditions more difficult by adopting a balancing test under which security interests weigh heavily against even the most basic needs of prisoners. Thus Justice O'Connor, in *Whitley v. Albers*—a case whose facts turned on a prisoner's being shot during the quelling of a prison disturbance—weighed the interest of a prisoner in his physical security against the broader security interests for which prison officials are responsible: when security interests are challenged, the "deliberate indifference standard" must give way to a simple good-faith standard. The following year, in *Turner v. Safley*, Justice O'Connor summarized and reinforced the holdings of previous prison conditions cases: "when a prison regulation impinges on inmates' constitutional rights, the regulation is valid if it is reasonably related to legitimate penological interests."

Security is a paramount interest of prison

---

67 *Whitley v. Albers*, 475 U.S. 312, 312 (1986) (holding that infliction of pain in the course of prison security measures was an Eighth Amendment violation only if inflicted unnecessarily and wantonly, and finding that the shooting of a prisoner during a riot without prior verbal warning did not violate the Eighth Amendment); *Rhodes v. Chapman*, 452 U.S. 337, 349 (1981) (holding that the Constitution does not mandate comfortable prisons, and that prisons which house inmates convicted of serious crimes cannot be free of discomfort); *Estelle v. Gamble*, 429 U.S. 97, 97 (1976) (holding that while deliberate indifference to a prisoner's serious illness or injury constitutes cruel and unusual punishment in violation of Eighth Amendment, inadvertent failure to provide adequate medical care to prisoner would not violate the Eighth Amendment).

68 *Rhodes*, 452 U.S. at 347 (holding that prison conditions that involved the wanton and unnecessary infliction of pain, or deprived inmates of basic needs or the minimal civilized measure of life's necessities, violated the Eighth Amendment).


70 Id. at 304.

71 Id. at 300.


administration, and administrators must be given wide discretion to determine how to provide it.\textsuperscript{74}

Once the fiction of transition gives way to the recognition of the reality of decades of confinement, the future of death-row-style confinement must depend uniquely on the security justification. Decades of idleness, isolation, and close confinement are unjustifiable as transitional for persons of any age. Potent as the security rationale may be in providing an Eighth Amendment blessing for death row conditions, security is an inapposite justification for death row confinement for the elderly. A prisoner who is dependent on the kindness of guards and fellow prisoners to comb his hair or walk to the shower does not require death row level security. The security rationale collapses when confronted with the realities of confining seventy- and eighty-year-old prisoners in death row conditions.

The needs of the condemned elderly, as well as prison security needs, suggest that these elderly should be housed and cared for much as the growing legions of geriatric general prison population inmates created by the current sentencing regime\textsuperscript{75} are housed and cared for.\textsuperscript{76} Increasingly, the general population’s elderly are segregated for appropriate care for reasons of both efficiency and humanity.\textsuperscript{77} The trend towards providing more age appropriate conditions for elderly prisoners implicates the Eighth Amendment prohibition of cruel and unusual punishment with respect to the aged of death row. Should age appropriate care advance to the status of a norm of prison administration, neither the crabbed conception of basic human needs expressed in \textit{Rhodes} and \textit{Wilson} nor the deliberate indifference standard enshrined in those cases would be a barrier to the recognition that the Eighth

\textsuperscript{74} See also Procunier v. Martinez, 416 U.S. 396, 404-05 (1989) (supporting deference to security judgments of prison officials in recognition of institutional competence and also holding that federalism requires deference on the part of federal courts to state prison authorities).

\textsuperscript{75} The era of long mandatory sentences and the curtailment of parole has produced a large and growing geriatric general population in prison. See Joana Brown Morton, \textit{Implications for Corrections of an Aging Prison Population}, 5 CORRECTIONS MGMT. Q. 78, 78-88 (2001); see also Ronald H. Aday, \textit{Golden Years Behind Bars: Special Programs and Facilities for Elderly Inmates}, 58 FED. PROBATION 47, 9-11 (1994).

\textsuperscript{76} Bruce Gross, \textit{Elderly Offenders: Implications for Correctional Personnel}, \textit{Forensic Examiner}, Spring 2007, at 59-61 (describing prisons as designed for young offenders and as accelerating the deterioration in mental and physical fitness of the aged, and also their inability to walk fast enough, to see and hear well enough, to negotiate prison life, their susceptibility to being preyed upon by younger prisoners, their increased health care needs, and the challenges they present to prison administration).

\textsuperscript{77} Aday, \textit{supra} note 27, at 144-67.
Amendment commands appropriate geriatric care for the elderly of death row as well as general population elderly. A brief review of the constitutional standards imposed on prison conditions will facilitate the argument.

In *Rhodes v. Chapman*, the Court sought for the first time to establish the limits that the Eighth Amendment imposes on prison conditions that do not blatantly and unanswerably violate the commands of the Eighth Amendment. At issue in *Rhodes* was the practice of double celling. Whatever the merits of this practice, double celling does not rise to the level of atrocious maltreatment that had been held to violate the Eighth Amendment’s prohibition of cruelty in earlier cases (e.g., whipping prisoners” and intentionally or recklessly denying necessary medical care”). What then of a more general nature applicable to prison conditions does the Eighth Amendment teach? To answer this question, Justice Powell returned in *Rhodes* to *Gregg v. Georgia*’s analysis of the history and development of the Court’s Eighth Amendment jurisprudence: Justice Powell relied upon the teaching of *Gregg* that the Eighth Amendment’s requirements are not static; rather, evolving standards of decency prohibit not only the “physically barbarous” but also “the unnecessary and wanton infliction of pain.” Unnecessary and wanton pain includes inflictions of pain “totally without penological justification.” Comparing the practice of double celling with deprivation of medical attention and the systematically brutal, life imperiling conditions held to violate the Eighth Amendment in *Hutto v. Finney*, Justice Powell concluded that, unlike those conditions, double celling does not “deprive inmates of the minimal civilized measures of life’s necessities,” such as “essential food, medical care, or sanitation.” Justice Powell also sounds the theme of deference to prison officials in matters of security”

---

82 Rhodes, 452 U.S. at 346.
83 Id.
84 Id.
85 Id. at 347-48
86 Id. at 349 n.14 (“Moreover, a prison’s internal security is peculiarly a matter normally left to the discretion of prison administrators.”).
and prison administration.\textsuperscript{87} Justice Powell thus equated the apparently more liberal \textit{Gregg} standard, unnecessary and wanton infliction of pain without penological justification, with the more stringent standard, life’s minimal necessities. The bridge between these two apparently disparate standards is supplied by Justice Powell in \textit{Rhodes} by deference to the institutional competence of prison administrators, who face resource limitations and are responsible for prison security. Such deference in effect means that Eighth Amendment review by courts must be limited to the most blatant and dire privations.

\textit{Wilson v. Seiter} presented an Eighth Amendment challenge to a congeries of prison conditions that were alleged to be systematically inhumane.\textsuperscript{88} Justice Scalia, writing for a majority of the Court, explained that even when alleging systematically inhumane conditions, the prisoner-petitioner must establish the deliberate indifference of prison officials.\textsuperscript{89} Justice White offered a vigorous rebuttal in an opinion concurring in the judgment but protesting the imposition of the deliberate indifference standard in cases that do not involve “specific acts or omissions directed at individual prisoners.”\textsuperscript{90} He argued that the standard was virtually impossible to meet in a case involving a challenge to systemically inhumane conditions, and one that could leave prisoners in constitutionally unacceptable conditions without recourse.\textsuperscript{91} He expressed concern that lack of resources would become an excuse for both constitutionally unacceptable conditions and the failure of courts to intervene.\textsuperscript{92} Justice White relied upon the \textit{Gregg} standard that prisoners were not to be subjected to unnecessary pain without penological justification; Justice Scalia’s opinion for the Court was devoid of reference to that more demanding formulation, but rather relied solely on the basic needs, minimal necessities standard of \textit{Rhodes}.

Let us suppose the day arrives when age appropriate conditions of confinement for geriatric prisoners advances from

\textsuperscript{87} Id. at 351-52 n.16 (“For all of those reasons, courts are ill equipped to deal with the increasingly urgent problems of prison administration and reform. Judicial recognition of that fact reflects no more than a healthy sense of realism.”).

\textsuperscript{88} Prisoners were subjected to “overcrowding, excessive noise, insufficient locker storage space, inadequate heating and cooling, improper ventilation, unclean and inadequate restrooms, unsanitary dining facilities and food preparation, and housing with mentally and physically ill inmates.” Wilson v. Seiter, 501 U.S. 294, 296 (1991).

\textsuperscript{89} Id. at 303.

\textsuperscript{90} Id. at 309.

\textsuperscript{91} Id. at 310.

\textsuperscript{92} Id. at 311.
a trend to a norm of prison management: *Rhodes* and *Wilson* will not stand in the way of recognizing that the Eighth Amendment requires relief from death row confinement and even substantial parity of treatment with the general population elderly, provided that security concerns are recognized to recede for an aged population. The “deliberate indifference” standard of *Rhodes* and *Wilson* would not protect prison officials who failed to conform to established norms for carceral care of elderly inmates. Prison administrators cannot claim ignorance of the standards and norms of their profession. The Eighth Amendment will command laggards to comply. Further, the reasons for the current trend toward providing geriatric care in prisons suggests that the gap between the minimalist basic needs standard of Justice Scalia in *Wilson* and that of avoidance of unnecessary pain in the older cases would not doom the claim under the more stringent standard. If, as some contemporary experts and practitioners maintain, age appropriate care is in fact more efficient, then resource scarcity does not provide a practical barrier to a conception of basic needs at least robust enough to avoid physical suffering. Collecting the elderly in facilities designed to meet the needs of the aged prisoner could prove both more humane and more efficient. Appropriate prison geriatric care would then find a home in the Eighth Amendment and the distinction between death row and general population in this regard would prove untenable.

The issue before the U.S. Supreme Court in the *Lackey* cases has not been long death row confinement but execution of prisoners long on death row. Let us now examine the case for an Eighth Amendment bar to the execution of the long incarcerated elderly.

VI. **DOES THE EIGHTH AMENDMENT TOLERATE THE EXECUTION OF THE CONDEMNED LONG-SERVING ELDERLY?**

There are two daunting hurdles confronting proponents of any type of *Lackey* claim. One hurdle is pragmatic, in that its recognition produces unacceptable consequences. For if the general *Lackey* claim were recognized, it could bring us perilously close to total abolition.

The second hurdle is the apparent inability of a *Lackey* claim to satisfy the demands of the methodology the Supreme Court employs in determining whether execution constitutes cruel and usual treatment. This is the doctrinal, or merits,
problem facing proponents of the proposition that the Eighth Amendment should recognize a *Lackey* claim. The difficulty in sum is that the Court requires evidence that contemporary standards of decency no longer tolerate execution. The Court’s method has relied heavily on evidence that legislation in the states manifests the development of a consensus against the practice of executing the class in question. *Lackey* claims therefore face an apparently insuperable hurdle: no American legislation has endorsed the claim.

Whether or not the general *Lackey* claim must fall before these barriers, I will argue that *Lackey*-for-the-Elderly may have a less arduous course. Let us first examine the pragmatics and then the doctrinal Eighth Amendment support for barring the execution of the long-serving elderly.

A. Pragmatics

1. *Lackey*-for-the-Elderly Is Consonant with the Supreme Court’s History of Limited, Piecemeal Abolition

Long imprisonment on death row is the norm—not the exception—for those eventually executed. The Supreme Court has gradually narrowed the reach of capital punishment but these reforms do not imply a willingness to trench so deeply as to challenge the retention of capital punishment as would the institution of a *Lackey* regime. A further pragmatic consideration is the fear that any *Lackey*-like restrictions on execution will spur a rush to execution.93 Although the general *Lackey* claim may be a bridge too far for the Supreme Court at this juncture, *Lackey*-for-the-Elderly may be within the bounds of the achievable. Its modesty comports with the gradual, piecemeal reformist trajectory of the Supreme Court’s Eighth Amendment capital jurisprudence since the 1976 *Gregg* decision that inaugurated the contemporary capital punishment regime. For while the numbers of the aged on death row are increasing, they will remain a relatively small segment of the long-serving condemned.94 Their numbers will become large enough for the

---

93 Justice Thomas observes, “The [*Lackey*] claim might . . . provide reviewing courts a perverse incentive to give short shrift to a capital defendant’s legitimate claims so as to avoid violating the Eighth Amendment right . . . .” Knight v. Florida, 528 U.S. 990, 992 (1999).

94 At year end 2009, 2.6 percent of death row was sixty-five or older and 8.2 percent was sixty or older. CAPITAL PUNISHMENT 2009, supra note 18, at 9 tbl.5.
issue of their execution to be salient in the experience of courts and prisons, but not so large that sparing them will strike a mortal or near mortal blow to capital punishment.

In the 1970s the United States Supreme Court considered total abolition of the death penalty but opted instead for permitting retention provided novel constitutional strictures were respected. In lieu of total abolition, the Court embarked upon reform. It required enhanced or “super due process” for death cases, and constricted the reach of capital punishment. The Court gradually augmented a list of crimes and criminals that were ineligible for the capital sanction. Jurisdictions wishing to retain capital punishment were required to devise statutes that would reserve capital punishment for the worst of the worst murders and murderers. Only homicides—and among them only the most aggravated murders—remain eligible for capital punishment. Nor are persons who “did not commit and had no intention of committing” homicides any longer eligible for capital punishment, eliminating capital liability for persons who take part in a felony where murder was committed by an accomplice. Recently the mentally retarded and persons less than eighteen when they killed were categorically exempted. Previously those less than sixteen years of age when they killed had been exempted. All the while, abolitionists and retentionists vied for political support and for supremacy in the

---

95 Margaret Jane Radin coined this apt phrase in her article, Cruel Punishment and Respect for Persons: Super Due Process for Death, 53 S. Cal. L. Rev. 1143 (1980).
96 See, e.g., Kennedy v. Louisiana, 554 U.S. 407 (2008) (holding that the Eighth Amendment prohibits the death penalty for the rape of a child); Roper v. Simmons, 543 U.S. 551 (2005) (holding that it is unconstitutional to execute any offender who is under eighteen at the time the crime was committed); Atkins v. Virginia, 536 U.S. 304, 311, 321 (2002) (holding that executions of mentally retarded criminals are “cruel and unusual punishments” prohibited by Eighth Amendment); Thompson v. Oklahoma, 487 U.S. 815 (1988) (holding that defendants who are less than sixteen at the time of their crime cannot be executed); Ford v. Wainwright, 477 U.S. 399 (1986) (holding that it is a violation of the Eighth Amendment to execute the insane); Enmund v. Florida, 458 U.S. 782 (1982) (holding that the death penalty for those who neither killed nor intended to kill in the course of a felony constitutes cruel and unusual punishment); Coker v. Georgia, 433 U.S. 584 (1977) (holding that the imposition of the death penalty for rape of an adult woman violates the Eighth Amendment).
97 Enmund, 458 U.S. at 801.
98 Id.
99 Atkins, 536 U.S. at 304.
100 Roper, 543 U.S. at 551.
101 Thompson, 487 U.S. 815.
The product of these vectors has been partial abolition by attrition. An unintended consequence of the enhanced due process regime and the contestation of capital punishment is that the condemned await final resolution of their cases for historically unprecedented periods. This long gestation period gives rise to the Lackey issue—whether it is cruel and unusual punishment to convert a sentence of death into a sentence of decades on death row followed by execution.

Relief for the elderly condemned would be another exclusion of a limited and discrete class. Relief for the entire class of the long-serving condemned would be a far more consequential abolitionist step. Indeed, because our capital punishment regime takes so long to produce executions, a Lackey rule would be close to a mortal blow to capital punishment rather than one more in a series of modest exclusions. The trouble with the Lackey claim is that it breaks with the established practice of the Supreme Court’s paring back and chipping away at capital punishment and instead trenches deeply. By the end of 2009, there were well over six hundred prisoners who had been on death row for twenty years or longer, comprising very nearly 20 percent of the total death row population. More than two hundred had been on death row for twenty-five years or longer. Almost 80 percent of long-serving inmates were admitted to death row in their twenties and thirties, with the balance divided almost equally between nineteen-year-olds and persons forty and older. By comparison, the numbers of death row elderly are relatively modest. There were approximately eighty individuals sixty-five or older on death row at the end of 2009. A Lackey-for-the Elderly rule would be consonant with the scope and pace of gradual retrenchment that has thus far found acceptance in the Supreme Court. Further, Lackey-for-the-Elderly mimics the recent exclusion of those under eighteen years of age at the time they committed a capital crime and the mentally

---

103 CAPITAL PUNISHMENT 2009, supra note 18, at 19 tbl.18.
104 Id.
105 Only 11.1 percent of death row prisoners were forty or older when arrested; 10.5 were nineteen or younger. CAPITAL PUNISHMENT 2008, supra note 24, at 10 tbl.7.
106 On December 31, 2009, 2.6 percent of 3173 total prisoners on death row were over the age of sixty-five. CAPITAL PUNISHMENT 2009, supra note 18, at 9 tbl.5.
in that there is an essentially objective, even arithmetic, definition available for demarcating the class of the excluded, relieving concerns about manipulation and undeserving claims. Lackey-for-the-Elderly is therefore a more practical goal: it would, if embraced by the Supreme Court, be consistent with its inclinations to date and offer relief to a cohort whose continued life under the shadow of the gallows is acutely misaligned with Eighth Amendment values.

2. Would a Lackey-for-the-Elderly Regime Produce a Rush to Execution?

Would rules sparing the elderly—or, for that matter, all the long-incarcerated condemned—lead to accelerated rates of executions for either of these classes of prisoners? Would a Lackey regime result in “speed rather than accuracy” in capital litigation, a consequence that would dismay defenders of due process whether or not they are of abolitionist sympathies? Could the adoption of such protection reverse the trend of secular decline in executions or propel more elderly and long-serving prisoners into the ranks of those actually executed? I will argue that these fears are not well founded.

Most execution-prone states share with less active death-penalty states populations of long-serving and elderly prisoners. Among the ten most execution-prone states, only Virginia lacks prisoners who have been on death row at least twenty-five years. Texas has nineteen. Florida is the leader in this class with more than sixty. The most execution-prone states, with the exception of Virginia, do not outperform the less execution-prone death-penalty states dramatically in the

---

107 To qualify for the protection of Atkins, an offender must have an IQ no higher than seventy to seventy-five. Atkins v. Virginia, 536 U.S. 304, 309 n.5 (2002).

108 McKenzie v. Day, 57 F.3d 1461, 1467 (9th Cir. 1995).

109 Judge Kozinski expresses the concern about the consequences of a Lackey regime, By and large, the delay in carrying out death sentences has been of benefit to death row inmates, allowing them to extend their lives, obtain commutation or reversal of their sentences or, in rare cases, secure complete exoneration. Sustaining a [Lackey] claim . . . would, we fear, wreak havoc with the orderly administration of the death penalty in this country by placing a substantial premium on speed rather than accuracy.

Id. at 1467.

110 CAPITAL PUNISHMENT 2009, supra note 18, at 19 tbl.18.

111 Id.

112 Id.
average number of years prisoners have been on death row. The national average time in 2009 was 12.7 years.\textsuperscript{113} For California, a state that rarely executes, the average was 14.2,\textsuperscript{114} but it was 13.9 for Florida.\textsuperscript{115} Georgia did not do much better at 13.3.\textsuperscript{116} Texas, the leader in executions performed, achieved 10.8.\textsuperscript{117} Other than Virginia’s stellar 5.4 years,\textsuperscript{118} the best records achieved by the ten most execution-prone states were in South Carolina and Oklahoma, which managed to get below 10 years, at 9 and 8.9 respectively.\textsuperscript{119}

I base my skepticism on what I consider to be the dim prospects for efficiency reforms of the death-penalty system. The number of years to resolve cases has increased over a decade in which there has been a pronounced secular decline in the number of executions.\textsuperscript{120} The ratio, as it were, of systemic effort per resulting execution has steadily increased. The two most plausible avenues of reform are money to move cases more quickly and competently through the system and the streamlining of the appellate and postconviction process.\textsuperscript{121} To date, such measures have either not made a difference in reversing the secular trend or have failed to garner sufficient support to be introduced. It is vanishingly unlikely that the well-documented shortage of qualified trial and appellate counsel will be addressed any time soon by cash-strapped states that did not take this step even when their resources were greater. The prospects for streamlining the process are no better. The passage of the Anti-Terrorism and Effective Death Penalty Act of 1996,\textsuperscript{122} for example, which restricts habeas relief for death row prisoners, has not reversed the trends to longer delays and fewer executions. To the frustration of some retentionists, the underlying cause of the multitiered and time-consuming legal processes is the so-called super due process for

\begin{flushright}
\textsuperscript{113} Id.
\textsuperscript{114} Id.
\textsuperscript{115} Id.
\textsuperscript{116} Id.
\textsuperscript{117} Id.
\textsuperscript{118} Id.
\textsuperscript{119} Id.
\textsuperscript{120} Despite a steep drop in both the number of executions and the number of admissions to death row over the decade, time under sentence of death for those executed increased from eleven years and eleven months in 1999 to fourteen years and one month in 2009. See CAPITAL PUNISHMENT 2009, supra note 18; see also CAPITAL PUNISHMENT 1999, supra note 17.
\textsuperscript{121} The Ninth Circuit Court of Appeals has recently proposed both as solutions to California’s backlog of death row cases. Alarcon, supra note 54, at 698.
\end{flushright}
death required by the Supreme Court coupled with the matching zeal of pro- and anti-death-penalty litigators. To date, the political will to execute has not been sufficient to result in spending public money and court resources on death cases to reverse the trend toward more time to produce fewer executions. It is doubtful that the recognition that a Clarence Ray Allen is too old to execute would galvanize more public and retentionist reaction than, for example, the Supreme Court’s recent decision to prohibit execution of child rapists or the increasingly more publicized fact that only a small fraction of those condemned since 1976 have actually been executed.

Let us look at the reaction to the introduction of a Lackey-like regime in the Caribbean nation of Jamaica for possible lessons about the impact of such a regime in the United States. The Anglophone Caribbean nations retain capital punishment; the death penalty enjoys strong public and political support as a consequence of the high rates of murder and other violent crime these countries suffer. Until recently the Judicial Committee of the Privy Council has been the high court for all former British dependencies in the Caribbean, a relationship that is in the early stages of being replaced by a regional court, the Caribbean Court of Justice. In 1993, the Judicial Committee ruled in Pratt and Morgan v. The Attorney General of Jamaica that, under the Jamaican Constitution, it would constitute inhuman and degrading punishment to execute a person imprisoned under sentence of death for more than five years. The decision resulted in the commutation of approximately two hundred death sentences in the region. The decision stimulated anti-imperialist and nationalist resistance and increased momentum for the creation of a regional court to displace the Privy Council. There were predictions that the decision would inspire “hanging courts.” There have been no executions in Jamaica since Pratt and

---

125 Pratt & Morgan v. Jamaica, 3 SLR 995, 2 AC 1, 4 All ER 769 (Privy Council 1993) (en banc).
Morgan. The Pratt and Morgan decision has been a much criticized decision by retentionists, but has been criticized by abolitionist critics as well who fear “hanging courts.”

Would prisoners who might otherwise be left to languish in obscure corners of death row be victims of Lackey-for-the-Elderly? The sluggishness of the death-penalty system in the United States, Jamaica, and other countries where stringent due process standards are respected suggests that fears of a rush to the gallows are misplaced. My argument is necessarily speculative, as there is no record to consult for the counterfactual Lackey scenario in the United States. My argument relies on both the natural experiment produced by Pratt and Morgan in Jamaica and the history of protracted multitiered litigation in death cases necessary to satisfy contemporary due process and human rights standards.

B. On the Merits: The Case that the Eighth Amendment Bars Execution of the Long-Serving Elderly

The case for Lackey-for-the-Elderly is in effect a specialized and augmented version of the Lackey claim. Support for the Elderly version therefore rests in part on the strength of the case for the general Lackey claim. Likewise, understanding the additional merits of the Elderly claim requires review of the case for the general Lackey claim.

1. In Recent Cases, the Supreme Court’s Eighth Amendment Methodology Has Become More Favorable to the Success of Lackey Claims

At first blush, the chances for success of the Lackey claim look bleak. Justice Thomas pointed out that he was “unaware of any support [for it] in the American constitutional tradition or in this Court’s precedent” and equally that no state or federal court has recognized a Lackey claim since the first denial of certiorari in Lackey v. Texas. His point must be conceded, for it was accurate when made in 1999 and remains true at this writing. However, a methodological shift in recent Supreme Court cases considering whether the Eighth

---

128 HOOD & HOYLE, supra note 124, at 106.
130 Id. at 461.
Amendment permits execution for particular classes of offenders suggests that Lackey claims may not be as outside the pale as its critics would prefer.

The Eighth Amendment standard by which the constitutionality of a form of punishment is to be judged was established in *Trop v. Dulles*: whether the punishment comports with “evolving standards of decency that mark the progress of a maturing society.” The Supreme Court has barred the execution of several classes of persons who were traditionally subject to capital punishment because the punishment for them would be cruel and unusual in light of contemporary standards. Invoking the Eighth Amendment’s Cruel and Unusual Punishment Clause, the Court has excluded the mentally retarded, and youthful murderers, and those who have been convicted because, during the course of committing a felony, an accomplice took life.

In two of the recent cases, *Atkins v. Virginia* and *Roper v. Simmons*, the Court relaxed the requirements for exclusion, to the dismay of the dissenting justices. The Court relied more heavily than in earlier cases on its “own judgment” or “independent judgment” and less on objective indicia of an evolution in public values. Although also continuing to rely on “objective” indicia of a national consensus to be found in legislative enactments and the practice of the states, the Court’s majorities were satisfied with less robust “objective” evidence of the development of a national consensus. The Court’s reliance on its “own judgment,” as will be seen, is deeply rooted in its capital Eighth Amendment precedents, and bodes well for the future reception of Lackey claims.

Writing for the Court in *Roper*, Justice Kennedy reviewed the method by which a determination is to be made as to whether offenders who kill before their eighteenth birthday

134 Enmund v. Florida, 458 U.S. 782, 801 (1982); see also Tison v. Arizona, 481 U.S. 137 (1987) (refining the Enmund standard and holding that reckless indifference is sufficient to satisfy intent if the defendant is a major participant in a felony that results in murder).
135 Atkins, 536 U.S. 304.
136 Roper, 543 U.S. 551.
137 In the most recent capital exclusion case, *Kennedy v. Louisiana*, 554 U.S. 407 (2008), discussed infra at note 165, Justice Kennedy’s majority opinion is yet more assertive of reliance on normative judgment at the expense of the weight accorded legislative enactments.
should be categorically exempt from capital punishment. The three sources of support recognized in previous cases are “legislative enactments,” 138 “state practice,” 139 and “our own judgment.” 140 Additionally, the Court looks to foreign and international law as instructive but not controlling. 141 In the earlier cases, the Court accorded preeminence to statutes. 142 In Coker v. Georgia, for example, the Court relied on the finding that Georgia was the only jurisdiction whose law authorized capital punishment for the rape of an adult woman as support for a constitutional bar against capital punishment for rapists of adult women. 143 The Court again emphasized the importance of statutes as evidence of the crystallization of consensus in Enmund v. Florida. 144 There the record was less “compelling” 145 but nonetheless strong: a large majority of states no longer permitted the execution of those who neither contemplated nor intended killing in the course of a felony, such as robbery during which an accomplice killed. 146 But when the Court barred execution of the mentally retarded in Atkins, only eighteen of the then thirty-eight death-penalty states, and the federal jurisdiction, exempted the mentally retarded by statute. 147 In Roper in 2005, the Court conceded that the objective indicia embodied in statutes was weaker than in the earlier cases in both Atkins and in the instant case exempting persons who were not yet eighteen when they committed murder. 148 Indeed, a bare majority of the death-penalty states had no statutory prohibition against the execution of the mentally retarded or murderers who were seventeen years of age but not yet eighteen when they killed. 149 The majority opinions in the two cases reasoned that recent additions to the ranks of prohibiting states, prohibitory statutes of more long standing, and the addition of the bloc of states that had abolished capital punishment entirely established that we had arrived at a

138 Roper, 543 U.S. at 563.
139 Id.
140 Id.
141 Id. at 575-78.
144 458 U.S. 782 (1982).
145 Id. at 793.
146 Id. at 792.
147 Atkins, 536 U.S. at 313.
149 Id. at 564.
national legislative consensus against the execution of the mentally retarded and youthful murderers.\textsuperscript{150} Inevitably, to compensate for the comparative weakness of the “objective” case relative to its strength in previous cases establishing categorical exclusions, the arguments for the Court’s independent judgment must bear greater weight in the two recent cases.

The dissenter’s verdict is that the \textit{Atkins} and \textit{Roper} majorities have substituted their own subjective moral judgments for the objective social consensus demanded by \textit{Trop v. Dulles} and the Court’s previous Eighth Amendment jurisprudence of capital exemption. Justice Scalia’s dissents in both cases are scathing, for he found that the Court has abandoned reliance on social consensus as the touchstone of evolving Eighth Amendment values in favor of “the subjective views of individual Justices.”\textsuperscript{151} In his \textit{Atkins} dissent, Justice Scalia declared, “Seldom has an opinion of this Court rested so obviously upon nothing but the personal views of its Members.”\textsuperscript{152} In \textit{Roper}, he again dissented, “Because I do not believe that the meaning of our Eighth Amendment . . . should be determined by the subjective views of five Members of this Court . . . .”\textsuperscript{153}

Plausible as this criticism may appear initially, it is misleading to characterize the majorities’ reliance on their independent judgment as either merely “subjective” or as a jurisprudential departure. The Court in \textit{Atkins} and \textit{Roper} reverts to a well-established method of normative analysis employed in Eighth Amendment cases, with roots that go back at least as far as the foundation of the contemporary capital regime (i.e., to \textit{Gregg v. Georgia}). Chief Justice Rehnquist and Justice Scalia, dissenters in \textit{Atkins} and \textit{Roper}, rejected this approach. Their view prevailed temporarily in \textit{Stanford v. Kentucky}.

In that case the Supreme Court, in an opinion authored by Justice Scalia, repudiated the previously established reliance on the independent judgments of the justices as an ingredient in determinations of contemporary standards of decency.\textsuperscript{154} But in \textit{Atkins}, this holding of \textit{Stanford}
proved to be a spur, a single shoot without progeny. The Court in \textit{Atkins} and \textit{Roper} reverted to reliance on the independent judgment of the justices as relevant to its determination of the Eighth Amendment propriety of capital punishment. Their judgment was inevitably normative, but this does not entail that it was merely subjective. The \textit{Atkins} and \textit{Roper} majority opinions apply normative standards regularly employed in the Court’s capital Eighth Amendment precedents.

In \textit{Coker v. Georgia}, the first of the categorical exclusion cases, the Court held that “the Constitution contemplates that in the end our own judgment will be brought to bear on the question of the acceptability of the death penalty under the Eighth Amendment.”\textsuperscript{156} The Court arrived at that judgment by inquiring into whether capital punishment serves the constitutionally required goals of punishment:

Under \textit{Gregg}, a punishment is “excessive” and unconstitutional if it (1) makes no measurable contribution to acceptable goals of punishment and hence is nothing more than the purposeless and needless imposition of pain and suffering; or (2) is grossly out of proportion to the severity of the crime. A punishment might fail the test on either ground.\textsuperscript{157}

The goals of punishments, as reiterated in all the Supreme Court cases mandating capital exclusion, are deterrence and retribution: “The death penalty is said to serve two principal social purposes: retribution and deterrence of capital crimes by prospective offenders.”\textsuperscript{158}

Since \textit{Coker}, and relying on \textit{Gregg},\textsuperscript{159} the Court has held that unless capital punishment advances the penal goals of

\textsuperscript{156} \textit{Coker}, 433 U.S. at 597 (quoted in \textit{Atkins}, 536 U.S. at 341 and \textit{Roper}, 543 U.S. at 590).

\textsuperscript{157} \textit{Id.} at 592 (White, J.). The sentence preceding that quoted in the text above details the precedents for this holding:

In sustaining the imposition of the death penalty in \textit{Gregg}, however, the Court firmly embraced the holdings and dicta from prior cases, \textit{Furman v. Georgia}, [408 U.S. 238]; \textit{Robinson v. California}, 370 U.S. 660 (1962); \textit{Trop v. Dulles}, 356 U.S. 86 (1958); and \textit{Weems v. United States}, 217 U.S. 349 (1910), to the effect that the Eighth Amendment bars not only those punishments that are “barbaric” but also those that are “excessive” in relation to the crime committed.

\textit{Id.} at 591-92.


\textsuperscript{159} And indeed on \textit{Furman v. Georgia}, 408 U.S. 238, 239-40 (1972) (holding that the imposition of the death penalty on offenders convicted of rape would violate the Eighth Amendment).
retribution or deterrence, it is barred by the Eighth Amendment. In debates about the Lackey issue, retribution is to the fore and deterrence recedes in that the plausibility of additional deterrent value in execution after decades of incarceration in the shadow of the gallows is difficult to defend. The Court’s Lackey debates therefore turn on whether decades-plus-death is excessive retribution offensive to the Eighth Amendment. In Atkins and Roper, the Court held that in its independent judgment, the deficiencies in judgment and self-control attributable to mental disability and immaturity rendered the retarded and youthful murderers less culpable than normal and mature murderers. Thus, these classes were not among the most culpable murderers for whom capital punishment was retributively justified, a judgment reflected in a large and growing number of statutory prohibitions.

Critics of the Lackey claim will of course hasten to note that the Lackey claim has not been endorsed in a single American statute or adopted by any American court. How then could the Lackey claim prevail within the framework established in the capital exclusion cases? Justices Stevens and Breyer adumbrated the available Eighth Amendment arguments in their demurrals from the denial of certiorari in the Lackey cases. There are two crucial lines of argument necessary to support Lackey claims. One is of course the standard Eighth Amendment argument that decades-plus-death is retributively excessive punishment. The second breaks new Eighth Amendment ground. Previously, cases acknowledging capital exclusion have tested whether traditionally accepted practices had outlived their social mandate. The Lackey claim, by contrast, questions whether an emergent practice is permitted by the Eighth Amendment.

The Lackey claim is directed against a novel form of cruel punishment alien to tradition—persons serving decades on death row, perhaps half a lifetime, perhaps into old age. It is directed against a cruel innovation rather than a cruel relic. The model of social history implicit in Trop v. Dulles is progressive. It deserves

---

160 In Lackey v. Texas Justice Stevens writes, “additional deterrent effect from an actual execution now, on the one hand, as compared to 17 years on death row followed by the prisoner’s continued incarceration for life, on the other, seems minimal.” 514 U.S. at 1046. This proposition resounds through subsequent Lackey cases. In Knight v. Florida, Justice Thomas apparently concedes as much in noting that Justice Breyer’s criticism of execution after long delay for lack of additional deterrent effect would be remedied by reverting to something like our earlier and sprightlier system. 528 U.S. 990, 990 n.1 (1999).

161 Roper, 534 U.S. at 552.

162 Atkins, 536 U.S. at 321; Roper, 543 U.S. at 571.
the old-fashioned designation “whiggish” in that it assumes that society continually improves, becomes ever more humane. It can be argued, therefore, that the Court’s Eighth Amendment jurisprudence of capital exclusion is restricted by *Trop v. Dulles* to delivering the *coup de grace* to decaying practices. The alternative is to recognize the limitations of the *Trop v. Dulles* model when confronted by historically novel forms of cruelty. Indeed, the Supreme Court in *Atkins*, and more starkly in *Roper*, without fully acknowledging the fact, relied less on a history of American progress and more on normative judgment to ban practices permitted by the majority of death-penalty states. In *Roper* especially, the majority also turned to the very strong support for barring the execution of persons under eighteen in international and foreign law.\textsuperscript{163} Indeed, whether or not one shares the dissenting justices’ pejorative view of the majority opinions in *Atkins* and *Roper*, those justices are correct that the majority opinions rely on normative judgments and foreign and international law to a far greater extent than previous Eighth Amendment capital exclusion cases. The *Lackey* claim, like the bar on the execution of youthful offenders, enjoys robust support in international and foreign law.

In the most recent of the capital exclusion cases, *Kennedy v. Louisiana*, the Court delivered yet another blow to the vaunted importance of state legislative consensus in its Eighth Amendment capital exclusion methodology.\textsuperscript{164} Justice Kennedy considered an objection to the Court’s holding that a national consensus had developed against the execution of rapists of children, based on the small number of states that authorized their execution and the fact that no child rapist had been executed by any state in forty years.\textsuperscript{165} The objection was that the Court was “blocking” the development of a national consensus.

\begin{itemize}
  \item \textsuperscript{163} Thus, Justice Kennedy writes,
  \begin{quote}
    The overwhelming weight of international opinion against the juvenile death penalty is not controlling here, but provides respected and significant confirmation for the Court’s determination that the penalty is disproportionate punishment for offenders under 18 . . . . The United States is the only country in the world that continues to give official sanction to the juvenile penalty. It does not lessen fidelity to the Constitution or pride in its origins to acknowledge that the express affirmation of certain fundamental rights by other nations and peoples underscores the centrality of those same rights within our own heritage of freedom.
  \end{quote}

  \textsuperscript{164} *Roper*, 543 U.S. at 554.
  \textsuperscript{165} Id. at 407 (2008).
  \textsuperscript{165} Id. at 409.
\end{itemize}
rather than relying on such a consensus.\textsuperscript{166} Louisiana argued that the \textit{Coker} decision had been read broadly to ban execution for nonlethal child rape rather than specifically and only the rape of an adult woman. \textit{Kennedy v. Louisiana} clarified any ambiguity about \textit{Coker}. \textit{Coker} prohibited execution for the rape of an adult woman, not nonlethal child rape.\textsuperscript{167} Thus, the argument goes, legislatures in the states did not appreciate that they were free to permit or forbid execution for child rapists. The question of whether a national consensus existed against the practice was therefore untested in the states.\textsuperscript{168} The Court should wait upon the day when legislatures have affirmatively banned execution for child rapists to declare a national consensus rather than act in the face of legislative inaction.\textsuperscript{169} To ban execution of child rapists under the Eighth Amendment would stifle the development of a consensus. To act in the absence of legislative bans would violate the Court’s methodological commitment to look to state legislatures for evidence of national consensus.

Justice Kennedy’s rejoinder to this criticism would seem to diminish further the relative weight of the “objective” component in Eighth Amendment capital exclusion methodology. He associated “evolving standards of decency” with the Court’s duty to restrain the use of capital punishment to insure its use only for crimes that are judged to be among the worst, apparently at the expense of the importance of the record of legislative enactments. He wrote,

These concerns overlook the meaning and full substance of the established proposition that the Eighth Amendment is defined by “the evolving standards of decency that mark the progress of a maturing society.” Confirmed by repeated, consistent rulings of this Court, this principle requires that use of the death penalty be restrained. The rule of evolving standards of decency with specific marks on the way to full progress and mature judgment means that

\textsuperscript{166} \textit{See id.} at 446 (“Our determination that there is a consensus against the death penalty for child rape raises the question whether the Court’s own institutional position and its holding will have the effect of blocking further or later consensus in favor of the penalty from developing. The Court, it will be argued, by the act of addressing the constitutionality of the death penalty, intrudes upon the consensus-making process. By imposing a negative restraint, the argument runs, the Court makes it more difficult for consensus to change or emerge. The Court, according to the criticism, itself becomes enmeshed in the process, part judge and part the maker of that which it judges.”).

\textsuperscript{167} \textit{Id.} at 426-27, 429.

\textsuperscript{168} \textit{Id.}

\textsuperscript{169} \textit{Id.} at 446.
resort to the penalty must be reserved for the worst of crimes and limited in its instances of application.\textsuperscript{170}

In sum, in its three most recent capital exclusion cases, the Supreme Court has tempered its reliance on the tally of state enactments and reasserted its reliance on its own normative judgment. Normative analysis of whether capital punishment for a crime or a class of offenders is excessive punishment has been a factor in the Court’s capital jurisprudence since \textit{Gregg v. Georgia}. In the recent cases, \textit{Atkins} and \textit{Roper}—and in a more limited and subtle way \textit{Kennedy v. Louisiana}—the Court has recalibrated the importance of normative analysis relative to the so-called “objective indicia” of evolution of standards. The Supreme Court could travel further down the road already taken and recognize the \textit{Lackey} claim—or the more modest \textit{Lackey-for-the-Elderly} claim.

2. Decades-Plus-Death Is Excessive Punishment Forbidden by the Eighth Amendment

Justices Stevens and Breyer consider at least three arguments in favor of the general \textit{Lackey} claim in their dissents from denial of certiorari in \textit{Lackey} cases. The touchstone argument is that decades-plus-death is retributively excessive punishment and therefore lacks an Eighth Amendment justification. The justices also argue that decades-plus-death is sanctioned neither by our legal tradition nor by considered or express political recognition. Additionally, Stevens and Breyer find support for \textit{Lackey} in the jurisprudence of foreign constitutional courts and international law.\textsuperscript{171}

\textit{a. Decades-Plus-Death Is Retributively Excessive Punishment}

The capital exclusion cases hark back to \textit{Gregg v. Georgia} for the framework for understanding retributive

\textsuperscript{\textit{170}} \textit{Id. at 446-47} (citations omitted).
\textsuperscript{\textit{171}} From the retirement of Justice Stevens in 2010 to the present writing, no other justice has publicly endorsed granting certiorari to a \textit{Lackey} claim. Should the day arrive when certiorari is granted, a stock of arguments for granting the petition is in hand in the dissents from denial of certiorari written by Justices Stevens and Breyer. Indeed, Justice Stevens’s memorandum to the denial of certiorari in \textit{Lackey v. Texas} in 1995 supplies jurisprudential seed sufficient to the task. \textit{Lackey v. Texas}, 514 U.S. 1045 (1995) (Stevens, J., mem. respecting denial of cert.).
justification and retributive excess. Justice White wrote in *Coker v. Georgia*,

Under *Gregg*, a punishment is “excessive” and unconstitutional if it (1) makes no measurable contribution to acceptable goals of punishment and hence is nothing more than the purposeless and needless imposition of pain and suffering; or (2) is grossly out of proportion to the severity of the crime. A punishment might fail the test on either ground.¹⁷²

“*Gregg instructs,*” Justice Kennedy reiterated in *Kennedy v. Louisiana*, that “the two distinct purposes served by the death penalty” are “retribution and deterrence of capital crimes.”¹⁷³ Justices Stevens and Breyer in their *Lackey* opinions explored arguments for the thesis that this novel imposition of decades-plus-death is not justified by a gain in deterrence¹⁷⁴ and exceeds the limits of justified retribution. Punishment that is justified neither as deterrence nor as an appropriate level or type of retribution is “gratuitous infliction of suffering” and as such not tolerated by the Eighth Amendment.¹⁷⁵

This doctrine that constitutionally legitimate punishment must respect limits is well grounded in the Court’s Eighth Amendment precedents; yet the bite of the doctrine seems to elude some critics of the *Lackey* claim, perhaps in part because judges sympathetic to *Lackey* claims use emotionally and morally charged terms to describe long tenure on death row—terms like “dehumanizing,”¹⁷⁶ “horrible,”¹⁷⁷ and “frightful.”¹⁷⁸ These usages may invite dismissal as merely expressing personal moral repugnance. Justice Thomas, for example, responds dismissively to what he takes to be a misplaced focus on the suffering of the long incarcerated condemned rather than the gruesome suffering inflicted by the justly condemned.¹⁷⁹ But

¹⁷³ *Kennedy*, 554 U.S. at 441.
¹⁷⁶ *Lackey*, 514 U.S. at 1046 n.* (Stevens, J.) (quoting People v. Anderson, 493 P.2d 880, 894 (Cal. 1972)).
¹⁷⁷ Id. (quoting *In re Medley*, 134 U.S. 160, 172 (1890)).
¹⁷⁸ Id. (quoting *Furman v. Georgia*, 408 U.S. 238, 288-89 (1972)).
¹⁷⁹ Justice Thomas gives a detailed description of the extensive torture inflicted on a murder victim by *Lackey* petitioner William Lee Thompson and his codefendant:

Justice Stevens altogether refuses to take into consideration the gruesome nature of the crimes that legitimately lead States to authorize the death
the Lackey claimant is not contesting the constitutionality of capital punishment as retribution for cruelly taking life. Justice Thomas’s rejoinder misses that the Lackey claimant is arguing, and to this extent correctly, that the Eighth Amendment imposes limits on punishment even on those most deserving of punishment. The Eighth Amendment would not, for example, allow a state to inflict drawing and quartering, or the burning of entrails while alive, as the prelude to or the method of execution regardless of how hideous the crime or how cruel the criminal. To meet Lackey proponents on the ground on which they are arguing, the proper rejoinder must include an argument that decades-plus-death is not excessive punishment for the worst murders or the worst murderers in the light of contemporary standards of decency.

Another example of a misleading rejoinder is that of Judge Kozinski in the Ninth Circuit’s Mackenzie v. Day, a case that raises a Lackey claim. Judge Kozinski wrote, “By and large, the delay in carrying out death sentences has been of benefit to death row inmates, allowing them to extend their lives,” and pursue various forms of legal relief or commutation. Intuitively, there is certainly something to be said for remaining in life as well as for the chance for relief from a death sentence. However, Judge Kozinski implicitly compared death plus delay with the prospect of prompt execution. Justice Stevens has concluded that delay in executions is “inescapable” in our death-penalty system. If Stevens is correct—and history to date certainly bears him out—the proper way to frame the Eighth Amendment issue is

penalty and juries to impose it. The facts of this case illustrate the point. On March 30, 1976, petitioner and his codefendant were in a motel room with the victim and another woman. They instructed the women to contact their families to obtain money. The victim made the mistake of promising that she could obtain $200 to $300; she was able to secure only $25. Enraged, petitioner’s codefendant ordered her into the bedroom, removed his chain belt, forced her to undress, and began hitting her in the face while petitioner beat her with the belt. They then rammed a chair leg into her vagina, tearing its inner wall and causing internal bleeding; they repeated the process with a nightstick. Petitioner and his codefendant then tortured her with lit cigarettes and lighters and forced her to eat her sanitary napkin and to lick spilled beer off the floor. All the while, they continued to beat her with the chain belt, the club, and the chair leg. They stopped the attack once to force the victim to again call her mother to ask for money. After the call, petitioner and his codefendant resumed the torture until the victim died.

Thompson v. McNeil, 129 S. Ct. 1299, 1302 (2009) (Thomas, J., concurring).\textsuperscript{180} McKenzie v. Day, 57 F.3d 1461 (9th Cir. 1995).\textsuperscript{184} Id. at 1467.\textsuperscript{182} Thompson, 129 S. Ct. at 1300.
not as a choice between dispatch and delay, but whether or not decades-plus-death (the system we have had for thirty years) is excessively retributive. Perhaps for many of the condemned, periodic torture would be preferable to certain and immediate death, but that does not render torture plus death a sentence that would survive Eighth Amendment review. Again in the Ninth Circuit, Judge Fletcher, sympathetic to the Lackey claim, evokes the life on death row of a Lackey claimant who has served twenty-three years on death row:

For twenty-three years, Ceja has suffered the anxiety of impending death and the greatly restricted activity allowed death row inmates. During that time, Ceja has had an execution date set at least five times: February 8, 1978; September 24, 1980; May 11, 1983; December 19, 1984; and January 21, 1998. For 23 years, Ceja has lived in solitary confinement, much of it in the typical death row cell on Cell Block 6 at the Arizona State Prison in Florence. Those cells are little more than a 7' x 10' windowless concrete box with a metal sink and toilet and a concrete slab for a bed. Activity outside that cell is typically limited to 3 one-to-two hour periods per week in which the inmate may shower or exercise. Visitations and phone privileges are much more limited than those for the general prison population. Many of a death row inmate’s neighbors are deeply disturbed men responsible for some of the most notorious murders in Arizona.183

When Justices Stevens or Breyer called attention to such severe privations and anxieties, it would miss the Eighth Amendment point to dismiss their sympathy for the Lackey claim as merely an expression of their personal repugnance contemplating the suffering of the condemned. In Lackey v. Texas, Justice Stevens sketched the argument that after seventeen years awaiting death, the length of Lackey’s death row incarceration, there was little or no additional deterrent or retributive value to be achieved by executing him and hence no Eighth Amendment justification for execution.184 Let us suppose for the sake of argument that Justice Stevens underestimated the deterrent or retributive value and overestimates the suffering, exacted from Lackey in seventeen years, such that the commutation of his sentence to life imprisonment would cheat justice and the hangman. Suppose a prisoner were sentenced to die at twenty and executed at sixty after having spent forty years on death row. Suppose a person sentenced at thirty is executed at seventy-five. It is difficult to resist, once it is acknowledged that the Eighth Amendment’s Cruel and Unusual Punishment Clause

183 Ceja v. Stewart, 134 F.3d 1368, 1368-69 (9th Cir. 1998).
is a doctrine of limitation, that a limit beyond which deterrence is exhausted and retribution is excessive has been reached at some number of years under sentence of death. The method of calculating the Eighth Amendment limit may not be easily agreed upon. Indeed, the line drawn may well be to some extent arbitrary, if no more arbitrary and debatable than maximum sentences for noncapital crimes. An unavoidable degree of arbitrariness and disagreement does not relieve the Supreme Court of the duty to set limits, which the longest serving death row inmates have surely exceeded.\textsuperscript{185}

In addition to the critical matter of excessive punishment, the opinions of Justices Stevens and Breyer in the \textit{Lackey} cases advance two further arguments for hearing and indeed granting \textit{Lackey} petitions that the Court has endorsed in its capital exclusion cases.

\textit{b. The Long Delay Departs from the Traditional Practice Sanctioned by the Constitution}

Capital punishment is sanctioned by the U.S. Constitution\textsuperscript{186} and is enshrined in its text.\textsuperscript{187} However, to the extent that the institutions and practices of the late eighteenth century remain guides to constitutionality today,\textsuperscript{188} no such provenance can be claimed for periods of a decade or more awaiting execution or reduction of sentence. The practice of executing within “days or weeks”\textsuperscript{189} cannot justify the contemporary national average of twelve years (or even that of the most efficient state, Virginia at 5.4 years\textsuperscript{190}), much less within two and three decades. Of course, the anti-\textit{Lackey} jurist replies that contemporary review processes produce these

\begin{flushleft}
\footnotesize
\textsuperscript{185} \textit{Lackey} petitioner William Lee Thompson, for example, had been on death row for thirty-two years when his petition was denied in 2009. \textit{Thompson}, 129 S. Ct. at 1303-04.

\textsuperscript{186} \textit{Gregg v. Georgia}, 428 U.S. 153, 169 (1976) (“We now hold that the punishment of death does not invariably violate the Constitution.”).

\textsuperscript{187} \textit{U.S. Const. amend. V} (“No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any person be subject for the same offense to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law.”).


\textsuperscript{189} \textit{Knight v. Florida}, 528 U.S. 990, 995 (Breyer, J.).

\textsuperscript{190} \textit{Capital Punishment 2009, supra} note 18, at 19 tbl.18.
\end{flushleft}
delays and that prisoners exploit them. But the bare fact that capital punishment has been so constrained by due process as to require long intervals between condemnation and execution does not satisfactorily answer the question of whether such delays should be considered unconstitutionally cruel. They cannot be justified as sanctioned by tradition imported from England, or by the worldview of the Founders, or by American practice prior to the late twentieth century. Decades of uncertainty and waiting were unknown within the tradition.

c. Long Delay Lacks the Legitimacy of Legislative Enactment

A related argument is that the decades-plus-death sentence lacks the legitimacy of legislative enactment. No American legislature has ever authorized this penalty. A rejoinder to this argument is that no legislature has enacted legislation prohibiting the execution of the long serving. In Ceja v. Stewart, Judge Fletcher offers an explanation for this lack of positive endorsement: "There has never been such a sentence imposed in this country—or any other, to my knowledge. Neither Arizona nor any other state would ever enact a law calling for such a punishment." The argument about constitutional tradition and the argument about legislation are related in that both criticize decades-plus-death on the grounds that it is an artifact of the contemporary death-penalty system devoid of the legitimacy that emanates from deliberate choice or acknowledgement within political processes.

d. The Repudiation of Execution Long Delayed by Foreign and International Courts

The Supreme Court’s openness to the persuasive force of the decisions of foreign and international courts is a thread that runs through its Eighth Amendment capital jurisprudence. In the most extensive discussion of foreign and international law in its Lackey cases, Justice Breyer canvasses the “growing number of courts outside the United States... that accept or assume the lawfulness of the death penalty” that have held that delay is a factor which may

---

191 Ceja v. Stewart, 134 F.3d 1368, 1369 (9th Cir. 1999).
192 Knight, 528 U.S. at 995-97.
193 Id. at 995.
render execution inhuman, degrading and cruel. “[P]articularly instructive” are the opinions of nations that share our legal traditions. Famously, the Judicial Committee of the Jamaican Privy Council imposed a strict five-year limit on the length of detention after which execution was no longer legal. Justice Breyer also noted that the Supreme Court of India requires that delay must be taken into account in sentencing and the Supreme Court of Zimbabwe bans execution outright in the case of delay. He expressed concern that the decision of the European Court of Human Rights prohibiting extradition because of long delay, the so-called “death row syndrome,” is cruel treatment prohibited by the European Convention on Human Rights will be followed by other international and national courts as periods of delay grow longer on U.S. death rows. This proved a prescient concern in that Canada’s Supreme Court subsequently ruled in part because of long death row delays that Canada would no longer extradite persons facing the death penalty to the United States.

International law and foreign law, particularly European law and the law of the former Commonwealth nations with which we share a legal tradition, offer a measure of support to the Lackey claim in that the three most recent capital exclusion cases, Atkins, Roper, and Kennedy, all find foreign and international jurisprudence helpful in resolving Eighth Amendment questions.

VII. LACKEY-FOR-THE-ELDERLY

Justices Stevens and Breyer have shown that sound Eighth Amendment jurisprudence supports the Lackey claim. Nevertheless, the Lackey claim has not won further overt support on the Court in the more than fifteen years since Clarence Lackey’s petition for certiorari was denied. The Court has refused certiorari to petitioners with nearly twice Lackey’s

---

194 Id. at 997.
195 Pratt & Morgan v. Attorney General for Jamaica, [1993] 1 A.C. 1, 29 (P.C.) (appeal taken from Jam.).
198 Knight, 528 U.S. at 995-96.
200 Knight, 528 U.S. at 996.
seventeen-year tenure on death row. Perhaps it is time for the more modest proposal of Lackey relief for the elderly. A petition for an elderly inmate marshals all the support for the general Lackey claim and builds on the special claims of the aged. It trades on the modest scale of the reform it requires.

It should be clear at the outset that the issue raised by the elderly claim is not the death sentence meted out upon conviction but whether it is constitutional to continue to subject persons who have achieved old age on death row to the threat of execution after long delay from the time of condemnation. If properly convicted and sentenced after the commission of a capital crime, it is assumed that the elderly of death row were culpable and eligible for capital punishment. What then distinguishes the elderly of death row from their younger peers?

I have argued that the elderly ought to be relieved of continuing to live in death row conditions whether or not they remain under threat of execution. Whether or not relieved of death row conditions of confinement, the elderly will have logged long detention in all but the most extraordinary cases of late-life conviction. If we continue to permit death row conditions for the elderly, their fragility due to aging processes and death row incarceration argue that their excess suffering renders their execution an acute violation of Eighth Amendment retribution norms. If they have been spared some years of death row incarceration, the prolongation of the ordeal of waiting for execution should be sufficient to exceed tolerable retributive standards. This argument rests on the proposition that frailty makes such forms of stress too intense to pass Eighth Amendment muster. The decision to spare the elderly from execution would be an application of a recognized protective norm systematically respected in society. The elderly are exempted wholly or in part from social obligations such as labor and military service on the basis of their frailty.

We do not expect persons in the decline of old age to meet the challenges posed to those in the prime of life; we exempt and protect the old as we do the young. Leroy Nash at ninety-three or Clarence Ray Allen at seventy-six were guilty of heinous crimes and served many years before their deaths. Yet the

---

202 See, e.g., Johnson v. Bredesen, 130 S. Ct. 541 (2009) (denying certiorari to petitioner who had been on death row for twenty-nine years); Knight, 528 U.S. 990 (denying certiorari to petitioner who had been on death row for twenty-five years).

specter of executing the demented and the multiply disabled aged shames us by violating a protective norm. They may deserve execution but they are no longer fit for execution. The Eighth Amendment significance of old age under condemnation is therefore not about acknowledging an evolution in values but rather recognizing that familiar values have become salient because of changes in the institution of capital punishment.

CONCLUSION

The long delays between pronouncement of sentence and execution, and the considerable uncertainty about whether any condemned man or woman will be executed in our system of capital punishment, have given rise to a new form of cruelty unknown to our ancestors. Delay is not aberrant but normal. It cannot be purged from the system without doing unacceptable violence to constitutionally mandated due process. It cannot be reduced without money for representation and court resources that have not been allocated to this purpose and will not be forthcoming. If the general Lackey claim is a victim of its consequences, perhaps we can manage a modest special case, Lackey-for-the-Elderly.
NOTES

Finding a Better Analogy for the Right of Publicity

INTRODUCTION

The right of publicity is the simple idea that there “is [an] inherent right of every human being to control the commercial use of his or her identity.” Although conceptually straightforward, it has been the subject of significant commentary and debate. Neither courts nor scholars have accepted a uniform theoretical foundation for the right of publicity. Consequently, it has developed into a disjointed doctrine. Scholars invariably analogize to more grounded concepts in an attempt to rationalize and set limits on the right. When either justifying the right of publicity’s existence or resolving a doctrinal issue, writers have argued that the right of publicity should mirror copyright law, trademark law, and other legal doctrines.

1 J. THOMAS MCCARTHY, THE RIGHTS OF PUBLICITY AND PRIVACY § 1:3 (2d ed. 2011).

2 See, e.g., K.J. Greene, Intellectual Property Expansion: The Good, the Bad, and the Right of Publicity, 11 CHAP. L. REV. 521, 521 (2008) (“Is there really anything left to say about this topic, given the proliferation of writing on it in the last ten to fifteen years? A lot has been said about the right of publicity, most of it negative.”).

3 See Michael Madow, Private Ownership of Public Image: Popular Culture and Publicity Rights, 81 CALIF. L. REV. 125, 238 (1993) (“[T]he affirmative case for publicity rights is at best an uneasy one. Individually and cumulatively, the standard justifications are not nearly as compelling as is commonly supposed.”).


5 See, e.g., Stacey L. Dogan & Mark A. Lemley, What the Right of Publicity Can Learn from Trademark Law, 58 STAN. L. REV. 1161, 1163-64 (2006) (“This approach turns the right of publicity into a new form of IP right, one based explicitly on analogies to and justifications for real property. Thinking about the right of publicity by analogy to IP law may indeed be helpful.”).

6 Randall T.E. Coyne, Toward a Modified Fair Use Defense in the Right of Publicity Cases, 29 WM. & MARY L. REV. 781, 782 (1988) (arguing that the right of publicity should be structured in the same manner as the fair use doctrine under...
or trademark dilution. Using tools from other academic disciplines, other scholars have argued for or against the right of publicity. Indeed, whenever scholars encounter the right of publicity, their first instinct is to compare it to something else.

While these analogies often provide novel and insightful critiques, the commentaries often ignore the significant differences between whatever perspective they are arguing from and the right of publicity. Indeed, at least one commentator has noted that legal issues created by the right of publicity cannot be resolved by “automatic invocation of a ready-made framework.”

One analogy that has been overlooked, which parallels the right of publicity and is far more practical than others frequently offered, is private contracting in the commercial and entertainment industries. In many ways, comparing this sort of private contracting is not an analogy at all, but a reference to business custom. Industry collective bargaining agreements protect similar interests as the right of publicity, and operate in many of the same ways. Specifically, both the right of publicity and entertainment collective bargaining agreements developed out of the same social and technological changes. Collective bargaining agreements confer rights—the most important of which is the right to residual compensation—to actors and performers to control and compensate those individuals for use of their personas within the contractual relationship. The right of publicity ensures individuals' control of their personas from the world at large. The one discernible difference between collective bargaining and violations of the right of publicity is that the

---

7 See generally Dogan & Lemley, supra note 5.
9 See generally Alice Haemmerli, Whose Who? The Case for a Kantian Right of Publicity, 49 DUKE L.J. 383, 411-30 (1999) (arguing that the theories of Immanuel Kant justify the right of publicity); Madow, supra note 3 (arguing from a Cultural Studies perspective).
11 See infra Part II.B.
12 See infra Part I.
13 See infra notes 203-09 and accompanying text.
14 See infra notes 198-204 and accompanying text.
15 1 M CAR THY, supra note 1, § 1.3.
usage in contracts is with consent of the individual and usage that violates the right of publicity is without consent. 16

This note suggests a new method to analyze the right of publicity. Voluntary contracts within the entertainment industry provide an analytical tool to assess both the underlying policy justifications for the right of publicity and the doctrinal rules within it. First, in terms of an underlying justification for the right of publicity, reference to contracts and business practice shows that the right of publicity is not a half-baked intellectual property right with little justification. Rather, this analytical framework supports unjust enrichment and natural rights theory justifications for the right of publicity—not labor theory and diminution-in-value justifications, as some scholars have suggested. 17 Second, contractual structure in entertainment contracts, which is an industry standard determined through collective bargaining and protects similar interests as the right of publicity, 18 provides a tool to analyze doctrinal rules within the right of publicity. Comparing entertainment contracts to the right of publicity supports extending the right of publicity to non-celebrities.19 Additionally, entertainment contracts provide lawmakers and courts with a benchmark to determine whether unauthorized usage of an individual’s image or persona is incidental to, and therefore not infringing on, the right of publicity. 20

Given that unions dominate the commercial and entertainment industries, 21 the most appropriate place to find standard entertainment contracts is coordinated collective bargaining agreements. Since right-of-publicity infringements are most prevalent in advertising due to First Amendment limitations, 22 this note focuses primarily on the Commercial

17 See infra Part III.
18 See infra Part II.
19 See infra Part IV.A.
20 See infra Part IV.B.
21 See infra notes 165-68 and accompanying text.
22 See 2 J. Thomas McCarthy, The Rights of Publicity and Privacy § 8:16 (2d ed. 2011) (“Today, advertising is labeled as ‘commercial speech’ which is within the First Amendment, but it enjoys a lower level of constitutional protection than does ‘news’ or ‘entertainment.’ In some cases, its level of First Amendment protection seems so attenuated as to be practically nonexistent.”); Peter L. Felcher & Edward L. Rubin, Privacy, Publicity, and the Portrayal of Real People by the Media, 88 YALE L.J. 1577, 1597-99 (1979) (arguing that news and entertainment have higher levels of constitutional protection and that commercial speech “is not regarded as being of constitutional proportions”). However, the right of publicity can prevail against even the strongest First Amendment interests. See, e.g., Zacchini v. Scripps-Howard Broad.
Contracts between the American Federation of Television and Radio Artists (AFTRA), Screen Actors Guild (SAG), Association of National Advertisers (ANA), and American Association of Advertising Agencies (AAAA). However, contractual relations in movies, television, radio programming, and screen writing are referenced when relevant.

Part I compares the historical development of the right of publicity and collective bargaining in the commercial and entertainment industries. It argues that both developed in response to the same social and technological changes. Part II compares the current collective bargaining and right-of-publicity regimes. Part III analyzes the traditional justifications for the right of publicity and suggests, with reference to the contractual relations, that unjust enrichment and natural rights theories best justify the right of publicity. Part IV considers several rules within the right of publicity. First, this part argues that the right of publicity should extend to non-celebrities. By offering evidence that non-celebrities have commercial value when appropriated, this part concludes that the right should extend to all individuals. Finally, examining the incidental use doctrine for the right of publicity, this part suggests that judges and legislators should use analogous collective bargaining provisions as a benchmark in formulating incidental use doctrine.

I. THE COMMON ORIGINS

Many scholars trace the right of publicity back to the right of privacy.\(^{23}\) However, more nuanced accounts recognize that the right of publicity originated as a response to two phenomena: a cultural shift that placed higher value on celebrity and fame, and the inadequacy of privacy rights in protecting celebrity rights.\(^{24}\) The increased value of celebrity

---


\(^{24}\) See, e.g., Madow, supra note 3, at 167 ("The right of publicity was created not so much from the right of privacy as from frustration with it. Moreover, . . . the whole matter was negotiated by courts and commentators with something less than divine ease and grace.").
occurred in three distinct periods. First, around the late nineteenth and early twentieth century, society’s icons changed from those who were known for their accomplishments—such as inventors or political leaders—to actors and athletes. Second, in the first half of the twentieth century, mass media began to treat personality as a valuable commodity as the new celebrities were employed for commercial gain. Third, in the 1950’s, with the breakdown in the studio system and the invention of television, actors’ images became particularly vulnerable to misappropriation. As will be shown, modern collective bargaining in the entertainment and advertising industries derives from these same cultural shifts, which makes it the closest set of principles to the right of publicity. Thus, it was not the law that shaped the contractual relations here. Instead, both developed independently out of the same cultural shifts.

A. The Invention of Celebrity, the Commodification of Persona, and the End of an Era

1. Cultural and Technological Shifts Lead to Commodification of Persona

The face of fame undeniably changed from the nineteenth to the twentieth century. Starting with the Revolutionary period, society’s heroes were civic leaders. At the time, merchants appropriated the founding fathers’ images with impunity. According to Professor Michael Madow, the founders “viewed their images as a kind of common republican property.” Indeed, given the treatment of public personas as common property, famous people seemingly had no right to prevent commercial appropriation of their image. Following the Revolutionary period, from 1820 to 1860, “poets, essayists, critics, historians, and preachers” wrote the national narrative,
but personas remained public property.\textsuperscript{33} Society knew the famous more by their words or actions than by their images.\textsuperscript{34} By contrast, nineteenth-century society placed a low value on actors and performers.\textsuperscript{35}

Starting in the late nineteenth century, a new face of fame emerged. As a result of changes in technology and journalism,\textsuperscript{36} society's attention focused on captains of industry and inventors.\textsuperscript{37} Specifically, the invention of photography and chromolithography revolutionized the reproduction of images.\textsuperscript{38} In the 1880s, modern newspapers “were made possible by high-speed presses, the linotype, halftone photo reproduction, and the emergence of news-gathering organizations such as the Associated Press.”\textsuperscript{39} These advancements lead to drastic changes in journalism, which Professor Madow describes as “genuinely pictorial or illustrated ‘personalities’ journalism.”\textsuperscript{40} Magazines and newspapers began to focus on prominent members of society, and often printed their pictures.\textsuperscript{41}

It is within this period that Samuel D. Warren and Louis Brandeis wrote their renowned law review article on the right of privacy.\textsuperscript{42} In response to the press's increasing interest in the private lives of prominent citizens\textsuperscript{43} and the widespread

\textsuperscript{33} Henderson, supra note 25, at 49; see also Madow, supra note 3, at 152 (“According to the social historian Neil Harris, commercial exploitation of famous persons—living and dead, political and theatrical, fictional and real—was common throughout the nineteenth century . . . .”).

\textsuperscript{34} Madow, supra note 3, at 159.

\textsuperscript{35} Id. at 226 (“A century ago actors, entertainers, and athletes were still socially marginal and politically inconsequential.”).


\textsuperscript{37} Henderson, supra note 25, at 50 (describing how society came to idolize “hero-inventors” and “captains of industry”).

\textsuperscript{38} Id. at 49.

\textsuperscript{39} Id. at 49-50; see also Madow, supra note 3, at 157 (“The closing decades of the nineteenth century also brought several related changes in popular journalism. Daily newspaper circulation jumped from 2.6 million in 1870 to 8.4 million in 1890.”).

\textsuperscript{40} Madow, supra note 3, at 158.

\textsuperscript{41} See Henderson, supra note 25, at 50 (“The new magazines such as 'McClure's' that appeared in the 1890s also played a role in enlarging the popular imagination, thereby redefining ideals of fame, success, and national heroism.”); see also Madow, supra note 3, at 159.

\textsuperscript{42} See Warren & Brandeis, supra note 23.

\textsuperscript{43} Id. at 196 (“To occupy the indolent, column upon column is filled with idle gossip, which can only be procured by intrusion upon the domestic circle.”).
use of photographs in media, the authors advocated for a new common law doctrine that guaranteed private citizens the “right to be let alone.” This new right of privacy would ostensibly prevent publishers from printing private facts about, or photographs of, ordinary citizens. However, Warren and Brandeis carefully circumscribed an exception to the rule: “The right to privacy does not prohibit any publication of matter which is of public or general interest.” This exception often resulted in courts’ denying right-to-privacy actions for misappropriating famous people’s professional identities, under the theory that famous people waived their right of privacy.

In the late nineteenth and early twentieth century, a new face of fame emerged: the celebrity. Americans first became fixated with the personal lives of “political leaders, businessmen, financiers, scientists and inventors.” However, by the 1920s, society’s attention shifted to “film actors, entertainers, athletes, and the like, people who excelled in the world of play.” According to historian Daniel Boorstin, society no longer idolized men of merit who achieved status through accomplishment; instead, Americans worshiped “celebrities,” defined as those “who [are] known for [their] well-knownness.” This shift in society’s interests resulted from social changes and technological advances. Foremost, dramatic alterations in national demographics occurred because of immigration. From 1890 through the 1920s, approximately 23 million Eastern Europeans and Italians immigrated to the United States, many

---

44 Id. at 195 (“Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life . . . .”).
45 Id. at 193.
46 Id. at 215 (“The general object in view is to protect the privacy of private life, and to whatever degree and in whatever connection a man’s life has ceased to be private, before the publication under consideration has been made, to that extent the protection is to be withdrawn.”).
47 Id. at 214.
48 See infra notes 117-21 and accompanying text.
49 See Henderson, supra note 25, at 50 (describing the public’s increasing fascination with entertainers starting in the 1880’s).
51 Madow, supra note 3, at 163.
52 BOORSTIN, supra note 36, at 57. According to Boorstin, modern celebrities are celebrated not for their achievements, but instead for constantly being in the public spotlight. See id. at 57-58. The creation of new celebrities is possible only through advancements in technology, which Boorstin labels the “Graphic Revolution.” Id. at 57.
53 Henderson, supra note 25, at 50 (discussing demographic changes and advancements in technology).
of them settling in eastern cities. This demographic change caused a cultural shift, much of which was "rooted in the
entertainment industry." Additionally, rapid technological
advancement allowed mass media to expose the public to the
new celebrity: the radio brought the celebrities' voices into the
nation's homes. The motion picture captivated audiences and
brought people closer to the actors with close-ups, creating a
more personal connection not achievable in live theater.
The television completed the creation of the modern celebrity by
bringing the picture and sound of actors and performers into the
home, which creates "the greatest degree of intimacy and
familiarity between performers and their audiences."

The rise of celebrity culture became associated with a
larger sociological shift in America from a society of production
to a society of consumption. According to some historians,
American consumption required a new way to distinguish
people within society; "[p]ersonality became a means to
distinguish our individual selves from the mass." Advertising
practices reflected consumption of personalities. With these
significant changes in culture and technology, personality
arguably became a commodity.

54 Id.
55 Id.
56 Roberta Rosenthal Kwall, Fame, 73 IND. L.J. 1, 27 (1997) ("Our society's
view of fame was most influenced at the outset by print, and then was completely
revolutionized in the wake of the birth of film and broadcasting.").
57 Id. at 30 ("Radio, while lacking the visual aspect of film, created a new
level of intimacy by bringing the performer to the listener 'live.' This intimacy allowed
people to feel there was very little separating them from the celebrity."); Henderson,
supra note 25, at 52 ("Unlike movies, radio was a household presence: in 1934 an
average radio cost about $35, and 60% of all American household had at least one set.
And, unlike records, radio was live: entertainment and information were there at the
touch of the dial.").
58 Kwall, supra note 56, at 29 ("The motion picture, like the photograph,
delivered a new level of realism, only it was superior to photographs in that it
transcended the provision of stars' images and allowed audiences to observe stars' behaviors and mannerisms.").
59 Madow, supra note 3, at 162 ("Moviegoers, in contrast [to those who
attended stage performances], got to see their favorites regularly—and, most
importantly, they got to see them in close-ups. This fostered an illusion of intimacy and
generated widespread curiosity about the stars' private lives and doings.").
60 Kwall, supra note 56, at 31.
61 Henderson, supra note 25, at 50-51.
62 Id. at 51.
63 For example, Professor Madow points to two advertising practices using
celebrities: product placements in movies and celebrity endorsements. Madow, supra
note 3, at 164-65.
64 Id. at 166.
Mass media, and in particular early Hollywood, played a substantial role in the commodification of personality. For instance, mass media notoriously “packaged” celebrities—turning nobodies into the famous through a careful media strategy to maximize commercial value.\textsuperscript{65} Prior to the 1950s, despite society’s treatment of personality as a commodity for nearly half century, property law provided no rights to protect this new value.\textsuperscript{66}

2. Media Exploits and Expands the Commodity of Persona

An explanation of the commodification of personality is incomplete without an analysis of the media structure itself. In particular, the motion picture industry, with its prominence before radio and television, played a substantial role in exploiting personas for commercial gain. Between 1912 and 1915, movie producers moved to California and opened up the first studios.\textsuperscript{67} Studios in the first half of the twentieth century approached movie making like mass production.\textsuperscript{68} In a process known as vertical integration,\textsuperscript{69} studios dominated every aspect of the industry.\textsuperscript{70} They owned the film lots, the means to produce feature length movies, and the theaters where audiences watched the final products.\textsuperscript{71} Most importantly, the studios controlled the actors of the era. Prior to 1948, studios hired actors to work exclusively for them\textsuperscript{72} and controlled nearly every aspect of actors’ careers.\textsuperscript{73} One writer summarized, “Imagine working under a seven-year contract that you cannot break and more than likely will be forced to renew, for a producer who can tell you who you can marry, what your

\textsuperscript{65} Kwall, supra note 56, at 32-34.

\textsuperscript{66} See, e.g., infra notes 116-21 and accompanying text.


\textsuperscript{68} Id.

\textsuperscript{69} See id.; see also THOMAS SCHATZ, THE GENIUS OF THE SYSTEM: HOLLYWOOD FILMMAKING IN THE STUDIO ERA 39 (First Univ. of Minn. Press 2010) (1989) (describing how studios coordinated movie production with theatre operations).

\textsuperscript{70} See The Emergence of the Hollywood Studio System, supra note 67.

\textsuperscript{71} Id. (“Key to the studio system was the Big Eight’s domination of all areas of the industry.”).


\textsuperscript{73} See SCHATZ, supra note 69, at 42 (explaining that actors had “little control over their individual careers or their pictures”).
morals must be, even what political opinions to hold." Indeed, movie studios exploited the personas of their actors by controlling the licensing of stars’ images to advertisers.

The studio system thrived on monopolistic practices. Studios refused to poach each other’s actors. Collusive norms within the movie industry and the massive amount of control that the studios exacted over their actors allowed only rare commercial misappropriation of actors.

However, in 1948, the Supreme Court handed down a landmark decision in *United States v. Paramount Pictures, Inc.*, which shocked the equilibrium within mass media. The Court held that the “Big Five” movie studios’ ownership of and dealings with theaters violated antitrust law. The courts ordered divestment from ownership of theaters and effectively ended the studio system. The studio RKO was the first to

---

74 Orsatti, *supra* note 72.
75 One notable example of the studios contracting their stars out for endorsements occurred in the tobacco industry. Given that the tobacco industry had a large national advertising budget, the studios placed their stars in “tie-ins,” where stars in tobacco ads sold both tobacco and new movies. K.L. Lum et al., *Signed, Sealed and Delivered: “Big Tobacco” in Hollywood, 1927-1951, 17 TOBACCO CONTROL 313, 314 (2008). The studios “maximize[d] marketing opportunities” through “[c]ross-promotion” advertisements that showcased tobacco products and big budget films together. *Id.* at 321. While the studios controlled when and where advertisers could use the stars’ personas, these “campaigns also paid stars substantial sums while reinforcing the stars’ notoriety, boosting their value to the studios and other national advertisers.” *Id.*
76 See Schatz, *supra* note 69, at 9 (“The Hollywood studio system was, as economists and the federal courts well understood, a ‘mature-oligopoly’—a group of companies cooperating to control a certain market.”). For example, the studios engaged in block-booking with theaters, which is “the practice of licensing . . . one feature or group of features on condition that the exhibitor will also license another feature or group of features released by the distributors during a given period.” *United States v. Paramount Pictures, Inc.*, 334 U.S. 131, 156 (1948). In order to have the rights to show quality movies, the studios forced the theaters to buy lesser quality movies. Schatz, *supra* note 69, at 39 (describing how studios used block booking and blind bidding to forced theatre owners to run second rate movies in order to access “A-class features and star vehicles”).
77 Orsatti, *supra* note 72 (“As there was a tacit agreement among studios not to raid each other for a star[,]s services at their contracts end, actors were not able to choose their roles which is crucial in building a career.”).
78 *Id.*; see K.L. Lum et al., *supra* note 75, at 318 (“[S]tudio contracts gave studios complete control over the use of their celebrity ‘brand names’. Major studios negotiated the content of testimonials, insisted that the timing of adverts and radio appearances be coordinated with movie releases, and denied permission for deals that did not serve their interest.”).
79 *Paramount*, 334 U.S. at 141-61.
80 *Id.* at 175 (remanding the case to district court to determine whether divestiture was necessary); *United States v. Paramount*, 85 F. Supp. 881, 899-900 (S.D.N.Y. 1949) (ordering studios’ divestment from theaters).
break the vertical integration by divesting from theater ownership, and the others followed suit soon afterwards. As part of the movie industry’s restructuring, studios no longer hired actors in long-term contracts; instead, the studios hired actors per film. The actors became free agents. However ironically, actors no longer received the studios’ protection, which had long defended actors’ personas through collusive practices. As a result, entertainers needed a new form of protection to stop misappropriation of their most valuable assets: their identities.

The invention of television brought additional instability. Actors feared that television stations would replay movies with impunity, resulting in fewer movie productions, less employment of screen actors, and diminution in value of personalities. One commentator explains, “By repeating episodes of favorite television programs and by airing old movies, television producers could squeeze additional revenue out of entertainment products with very little extra expenditure.”

Commodification of personality also occurred within professional athletics. Similar to the control that the studio system exacted over actors, the notion of amateurism regulated nineteenth-century athletes’ abilities to commodify their public persona. However, starting in the late nineteenth and early twentieth century, professional athletics allowed athletes to be

---

82 See id.
84 See SCHATZ, supra note 69, at 482 (“For top industry talent . . . declining studio control meant unprecedented freedom and opportunity.”).
85 See supra note 75 and accompanying text.
86 See Chi, supra note 83, at 35-36. Television, initially centered in New York, brought significant unemployment to the movie actors who were located in California. Id. at 35.
87 Id.
89 See Henry Yu, Tiger Woods at the Center of History: Looking Back at the Twentieth Century Through the Lenses of Race, Sports, and Mass Consumption, in SPORTS MATTERS: RACE, RECREATION, AND CULTURE, 320, 322 (John Bloom & Michael Nevin Willard eds., 2002) (“[A]mateur sports almost exclusively involved men of privilege whose wealth meant they did not have to exchange labor for money, and therefore their sporting activities were practices exempt from monetary transactions.”). For example, the Olympic Committee stripped Jim Thorpe of his gold medals in track after it became apparent that he had been compensated for playing in minor league baseball in the past. Covington Baker, supra note 88, at 13.
compensated for their performance.”90 Starting in the 1920s, radio broadcasting within sports, particularly within baseball, communicated sports contests throughout the nation, as athletes became icons.91 Unlike actors under the studio system, professional athletes largely had control over licensing their personas.92 For example, Babe Ruth made approximately half his salary in endorsements during the 1920s and had a substantial effect on society given “the number of personal and radio appearances . . . as well as photo images and newspaper articles about him.”93 Given professional athletes’ personal control over the public’s perceptions of them, as opposed to actors who had ceded any rights to studios that were engaged in collusive activities, there was some demand for legal protection against misappropriating their personas. This can be seen in right of privacy cases, like Hanna v. Hillerich & Bradsby Co.94 and O’Brien v. Pabst Sales Co.,95 in which athletes argued invasion of privacy when their images had been misappropriated.96 Nevertheless, like for actors, the advent of the television brought about significant increase in athletes’ exposure to the American public, and with it an increased probability that their images would be misappropriated.97 Thus, although the need to protect athletes’ personas was addressed with the advent of professional sports, the demand for greater legal protection only increased with the invention of television. Athletes’ increased demand coincided with actors’ needs, as seen from the end of the studio era.

In the late nineteenth and early twentieth centuries, the nature of fame changed, which resulted in commodifying of identity. No rights existed to protect these new commodities, but the amount of control under the studio system made these rights moot. However, the remarkable changes to movie studio structuring and the advent of television quickly required new

90 See Covington Baker, supra note 88, at 6-7 (describing the compensation for professional football teams in the 1890’s).
91 Id. at 7-11.
92 See, e.g., id. at 10-11 (describing how Red Grange, a professional football player from the 1920s and 1930s, endorsed a number of products, including ginger ale, candy bars, and meatloaf).
93 Id. at 11.
94 78 F.2d 763 (5th Cir. 1935).
95 124 F.2d 167 (5th Cir. 1942).
96 See infra notes 117-21 and accompanying text.
97 Covington Baker, supra note 88, at 14-18. For example, Baseball Weekly ranked television “as the most important change in the game of baseball during the 20th century, second only to Jackie Robinson’s breaking the color barrier.” Id. at 16.
rights. Both the right of publicity and patterns of collective bargaining in entertainment derive from a need for more protection of personalities in an age where image was valuable but no longer institutionally controlled.

B. Right of Publicity as a Response

The right of publicity can be seen as a direct legal response to the commodification of personality and technological change. Doctrinally, the right of publicity has its genesis in the right of privacy. Warren and Brandeis advocated for a new common law right of privacy, which protected the “right to be let alone.” William Prosser later famously divided privacy into four distinct torts that are generally recognized, but the only category that is relevant for the purposes of this note is the one that Prosser called “[a]ppropriation, for the defendant’s advantage, of the plaintiff’s name or likeness.” The right of privacy for misappropriation vindicates plaintiffs for mental distress, rather than commercial loss, resulting from unwillingly being placed in the public eye.

Roberson v. Rochester Folding Box Co. was the first case to test the right of privacy. The Franklin Mills Company, one of the defendants, placed a picture of Abigail Roberson, the plaintiff, on twenty-five thousand packages of flour without her consent. Roberson claimed that she suffered mental distress and that the advertisement violated her right to privacy. The court refused to recognize a common law right of privacy, but

---

See, e.g., Dymond, supra note 23, at 449; Dogan & Lemley, supra note 5, at 1167-68.  
William Prosser, Privacy, 48 CALIF. L. REV. 383, 389 (1960) (“The law of privacy comprises four distinct kinds of invasion of four different interests of the plaintiff, which are tied together by the common name, but otherwise have almost nothing in common except that each represents an interference with the right of the plaintiff, in the phrase coined by Judge Cooley, ‘to be let alone.’”); 1 McCarthy, supra note 1, § 1:19 (“The courts have almost uniformly adopted Prosser’s four-part division as the ‘gospel’ of privacy law. Anyone who refuses to talk in Prosser’s language will meet blank stares of incomprehension.”).  
Prosser, supra note 100, at 389.  
1 McCarthy, supra note 1, § 1:7.  
64 N.E. 442 (N.Y. 1902).  
Id. at 442.  
Id. (“[H]er good name has been attacked, causing her great distress and suffering, both in body and mind; that she was made sick, and suffered a severe nervous shock, was confined to her bed, and compelled to employ a physician, because of these facts . . . .”).  
Id. at 443.
suggested that the legislature could enact a statute creating such a right. In response to Roberson, the New York State legislature passed section 50 of the N.Y. Civil Rights Law, which proscribes using “name, portrait or picture of any living person without having first obtained the written consent” for advertising or trade purposes.

The Georgia Supreme Court broke with New York’s rejection of the common law right of privacy in Pavesich v. New England Life Ins. Co. Under facts similar to Roberson, the court concluded that there was a common law right of privacy and that “[t]he novelty of the complaint is no objection, when an injury cognizable by law is shown to have been inflicted on the plaintiff.” Pavesich adopts several influential rules of law from Warren and Brandeis’s article that affected the development of publicity rights. First, the court recognized that public figures waive their rights of privacy to the extent that the information disclosed is relevant to the public interest. The court recited the example that a political candidate loses a certain degree of privacy in his or her private life because private information “may throw light upon his qualifications for the office, or the advisability of imposing upon him the public trust which the office carries.” Second, the court declared that the right of privacy is a personal right that is not assignable.

With increasing acceptance of a privacy right for misappropriation, several plaintiffs unsuccessfully attempted to use privacy as a legal protection against misappropriating celebrities’ images. In essence, these claims sought damages for harm done to a professional persona rather than any actual mental distress suffered by the plaintiff. First, in Hanna

---

107 Id. (“The legislative body could very well interfere and arbitrarily provide that no one should be permitted for his own selfish purpose to use the picture or the name of another for advertising purposes without his consent.”).
108 1 McCarthy, supra note 1, § 1:16.
110 50 S.E. 68 (Ga. 1905).
111 See id. at 68-69 (suing for defendant’s use of picture in a newspaper advertisement without the plaintiff’s consent).
112 Id. at 78.
113 Id. at 69.
114 Id. at 72 (“The right of privacy, however, like every other right that rests in the individual, may be waived by him, or by any one authorized by him, or by any one whom the law empowers to act in his behalf, provided the effect of his waiver will not be such as to bring before the public those matters of a purely private nature which express law or public policy demands shall be kept private.”).
115 Id.
116 Id. at 73 (“It therefore follows from what has been said that a violation of the right of privacy is a direct invasion of a legal right of the individual.”).
Manufacturing Co. v. Hillerich & Bradsby Co., a bat manufacturer that had the exclusive rights to use the names and autographs of famous baseball players sued a competing bat manufacturer for using the same names on its bats. The court held, in part, that the plaintiffs had no right-to-privacy claim because the baseball players’ names were not property capable of assignment; therefore, the plaintiff gained no property right from the exclusive contract and had no cause of action against the infringers. Second, in O’Brien v. Pabst Sales Co., a famous college football player sued a beer manufacturer for placing his picture in their promotional calendar without consent. The court held that the right of privacy did not apply because the plaintiff waived his privacy rights regarding his football career as a public figure. Hanna and O’Brien, respectively, illustrate two major deficiencies in the right of privacy when it comes to protecting publicity rights: (1) public waiver of privacy, as seen in O’Brien, and (2) the nonassignability of personality traits, as seen in Hanna.

Despite several attempts to transform privacy into a legal protection for celebrities’ persona, the legal system recognized a right of publicity only after the immense changes in technology, such as the advent of television and the end of the studio system within the entertainment industry. In Haelan Laboratories v. Topps Chewing Gum, Inc., a case with eerily similar facts to Hanna, Judge Frank of the Second Circuit was the first to recognize a right to protect commercial value of personality. Judge Frank was well aware of the commodification of persona and the significant structural changes within the entertainment industry, as he wrote:

This right might be called a “right of publicity.” For it is common knowledge that many prominent persons (especially actors and ball-players), far from having their feelings bruised through public exposure of their likenesses, would feel sorely deprived if they no longer received

---

118 Id. at 766.
119 Id. at 766-67.
121 Id. at 170.
122 See supra Part I.A.
123 202 F.2d 866 (2d Cir. 1953).
124 Like the bat manufactures in Hanna, in Haelan, a gum manufacturer sued a rival gum manufacturer for inducing a baseball player, who had an exclusive contract with the plaintiff, to allow the defendant to use the player’s image in advertisements. Id. at 867.
125 Id. at 868.
money for authorizing advertisements, popularizing their countenances, displayed in newspapers, magazines, buses, trains and subways.\textsuperscript{126}

Academics of the era defended the newly created right of publicity and provided a more elaborate justification. In his law review article \textit{The Right of Publicity,}\textsuperscript{127} Melville B. Nimmer argued that the right of privacy “is not adequate to meet the demands of the second half of the twentieth century, particularly with respect to the advertising, motion picture, television, and radio industries.”\textsuperscript{128} His specific concern for the entertainment industry shows that the right of publicity was supposed to address not only the commodification of persona, but also the significant technological and structural changes in the entertainment industry.

\textbf{C. Collective Bargaining Agreements as a Response}

The right of publicity was not the exclusive response to the commodification of personality and rapidly improving technology. Entertainment collective bargaining agreements also sought to protect actors from misappropriation of their commercial images. Unions in the entertainment industry existed long before the 1950s.\textsuperscript{129} Under the studio system, which provided relatively stable employment,\textsuperscript{130} workers formed unions to guarantee standardized wages\textsuperscript{131} and working conditions.\textsuperscript{132} The unions’ jurisdiction divided along profession. Each position within the industry had its own craft union.\textsuperscript{133} For example, the

\begin{itemize}
\item \textsuperscript{126} Id.
\item \textsuperscript{128} Id. (emphasis added).
\item \textsuperscript{129} For example, Actors’ Equity Association, which represents theater actors, started in the early twentieth century, and was recognized by the American Federation of Labor in 1919. \textit{Historical Overview, Actors’ Equity Ass’n}, http://www.actorsequity.org/aboutequity/historicaloverview.asp (last visited Jan. 10, 2010).
\item \textsuperscript{131} See David F. Prindle, \textit{The Politics of Glamour: Ideology and Democracy in the Screen Actors Guild} 16-25 (1988) (describing how SAG was formed following a significant pay cut in 1933).
\item \textsuperscript{132} See Orsatti, supra note 72 (suggesting that SAG gained better working conditions for actors under the studio system, but “the studios still basically ‘owned’ their stars”).
\item \textsuperscript{133} See Paul & Kleingartner, supra note 130, at 666 (“DGA represents all directors, whether in film or videotape production; WGA has jurisdiction over writers, including most news writers; SAG and the American Federation of Television and Radio Artists (AFTRA) represent all performers except instrumental musicians, who are represented by the American Federation of Musicians . . . .”).
\end{itemize}
American Federation of Radio Artists, the precursor to AFTRA, started as a representative of radio performers, and SAG represented actors in movies and later television.

After the collapse of the studio system and the advent of new technology, actors faced less demand for their services and massive unemployment. Two technological advances at that time were particularly important. First, starting in the early 1940s, radio gained the ability to record and rerun radio shows. Second, starting in the 1950s, television began rerunning programming and movies on television. Both of these practices reduced the number of productions while continuing to expose the performers to the public at large, thereby diminishing the value of the actors' personas.

To address these issues, unions compromised. Following the demise of the studio system, performers' unions no longer fought for job security. Instead, the unions focused on increasing the wages and working conditions of their rank-and-file members when they actually had work. To do so, the unions negotiated for a three-tier compensation system,
which still exists today. First, all actors receive minimum wage rates. The minimum wage rates compensate actors for time spent working on a project, and compensate directors and writers for products delivered. Second, personal service-contracts provisions within the collective bargaining agreements allow individuals to negotiate outside of the collective bargaining compensation scale, provided that the wages are higher than the contract scale. These allow well-known or high-quality actors, writers, and directors to demand more for their services. Third, and most importantly, actors receive residual payments. Residual payments are supplemental payments to actors for reuse of “entertainment product in media other than the one for which it was originally created, or for its reuse within the same medium subsequent to the initial exhibition.”

Residual payments first appeared in 1941 when radio introduced recording technology. Prior to this advancement, radio performers presented their program multiple times per day—at least once on the East Coast and once on the West Coast. Since the performers were paid per performance, had the American Federation of Artists not secured residual payments, compensation would have been cut in half without supplemental payment for the programming’s reuse. Residuals emerged next in television. SAG secured residual payments for its members for rerunning television programs in 1952, for repeated use of television commercials in 1953, and for movies reformatted and aired on television in 1960. Residuals

---

142 See infra Part II.B.2 (describing how the same three-tier compensation system still operates under the current Commercials Contract).
143 Paul & Kleingartner, supra note 130, at 667 (detailing how entertainment contracts “contain[] a minimum compensation schedule”).
144 Id. (describing how entertainment contracts calculate actors’, singers’, and stunt players’ minimum rates based on either a day rate, week rate, or a rate for a specified term).
145 Id.
146 Id. at 668.
147 Id. at 669 (“Personal service contracts provide for the exchange of scarce, differentiable, and perishable talent.”).
148 Id.; see also PRINDLE, supra note 131, at 82 (describing how residual payments became a necessity to actors’ economic survival after movies could be replayed on television, as “[a]ctors discovered that they were competing with their former selves for jobs, and losing”).
149 Jackson, supra note 137.
150 Id.
151 Id.
152 Id.; Orsatti, supra note 72.
153 Orsatti, supra note 72.
expanded to nearly all collective bargaining agreements in the entertainment industry, including writers and directors.\textsuperscript{154}

Not all performers that work under entertainment contracts receive residual payments. Under the standard union contracts, only those whose professional personas are exploited by reuse are paid residuals. In film, television, and television commercials, only principal actors receive residual payments; extras are entitled only to minimum payments.\textsuperscript{155} Even writers represented by the Writers’ Guild of America (WGA) receive residual payments for screenplays only when they are given screen credit.\textsuperscript{156} Given this set of circumstances, the residual payments represent more than a deferral in compensation to offset long periods of unemployment; to fulfill that justification, they would have to extend to everyone in the industry. Instead, residuals compensate workers for dilution of their professional-persona value by exploiting reuse of contractually bound material. The emergence of residual payments, like the right of publicity, was a reaction to the commodification of identity and advancing technology, but within the contractual relationship rather than the world at large.

II. HOW THE RIGHT OF PUBLICITY AND COLLECTIVE BARGAINING PROTECT AGAINST MISUSE OF PERSONA

Thus far, this note has established that the right of publicity and patterns in collective bargaining in the entertainment industry arose from the same historical context. In the entertainment industry, the right of publicity and collective bargaining agreements also both operate to protect actors against misappropriation of their personas. This section will draw upon the AFTRA/SAG Television Commercials Contracts with the ANA/AAA (Commercials Contracts) to illustrate how collective bargaining agreements afford these protections.

A. Scope of the Rights

Both the right of publicity and the Commercials Contracts are limited in scope. However, a brief analysis of

\textsuperscript{154} See Paul & Kleingartner, supra note 130, at 671 (describing how residuals are calculated under various collective bargaining contracts).


\textsuperscript{156} Id. at 108.
where and upon whom they apply will show that both are disproportionately influential on the media industry.

1. The Scope of the Right of Publicity

The right of publicity is a state law doctrine recognized in approximately nineteen states by statute.\(^{157}\) Twenty-one states recognize a common law right of publicity,\(^{158}\) eight of which also have a statute.\(^{159}\) Therefore, thirty-one states have explicitly recognized the right of publicity. The protections afforded under the right of publicity vary from state to state. While the number of states that recognize the right of publicity is limited, the right’s effect on media is enormous because the largest and most media-concentrated states accept it. In particular, one writer points to California, New York, and Tennessee as the states “whose economies are impacted the most by the entertainment industry.”\(^{160}\)

2. Scope of the Commercials Contracts

The Commercials Contracts are collective bargaining agreements between actors’ unions and the advertising industry. The Joint Policy Committee for Broadcast Talent Relations (JPC) represents advertising management in negotiations.\(^{161}\) JPC is a multi-employer bargaining unit

\(^{157}\) See 1 Mccarthy, supra note 1, § 6:8 (identifying California, Florida, Illinois, Indiana, Kentucky, Massachusetts, Nebraska, Nevada, New York, Ohio, Oklahoma, Pennsylvania, Rhode Island, Tennessee, Texas, Utah, Virginia, Washington, and Wisconsin as recognizing right of publicity by statute); see also Hunt, supra note 23, at 1607 n.20 (listing eighteen states, but excluding Pennsylvania which passed its statute after the publication of Hunt’s article).

\(^{158}\) See 1 Mccarthy, supra note 1, § 6:3 (identifying Alabama, Arizona, California, Connecticut, Florida, Georgia, Hawaii, Illinois, Kentucky, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, Ohio, Pennsylvania, South Carolina, Texas, Utah, West Virginia, and Wisconsin as recognizing right of publicity through common law).

\(^{159}\) See id. § 6:3 nn.7-8 (identifying California, Florida, Illinois, Kentucky, Ohio, Pennsylvania, Texas, and Wisconsin as having both a common law and statutory right).

\(^{160}\) Dymond, supra note 23, at 448.

\(^{161}\) See Memorandum from Douglas J. Wood, ANA-AAAA Joint Policy Comm. on Broad. Talent Union Relations, Unions and the Prod. of Commercials for Traditional and Non-Traditional Media 1 (Jan. 28, 2009), available at http://www.adlawbyrequest.com/uploads/file/Why%20be%20an%20Authorizer.pdf (“The JPC is the multi-employer bargaining unit that represents the interests of the advertising industry in negotiations with the various unions that represent performers and musicians who perform commercials . . . . The JPC is comprised of thirty members—fifteen appointed by the Association of National Advertisers and fifteen appointed by the American Association of Advertising Agencies. The ANA also appoints the JPC Lead Negotiator and legal counsel.”).
comprising the Association of National Advertisers (ANA) and the American Association of Advertising Agencies (AAAA). JPC negotiates the collective bargaining agreements jointly with the Screen Actors Guild (SAG) and the American Federation of Television and Radio Artists (AFTRA).

The Commercials Contracts deal exclusively with advertisements on television, Internet, radio, and new media, such as video advertisements on cellular phones. The contracts do not bind non-union employers, and do not cover photography for print advertising. Like the right of publicity, the Commercials Contracts are limited in scope but influential on the advertising industry. Advertisers produce 90 percent of television commercials under the Commercials Contracts, which totals approximately $1 billion in compensation to unionized performers.

162 The ANA is a trade organization for companies that advertise. It represents more than four hundred companies and ten thousand brands, which spend more than $250 billion on advertising and marketing annually. About the ANA: Leading the Marketing Community, ASS’N NAT’L ADVERTISERS, http://www.ana.net/about (last visited Jan. 11, 2012).

163 AAAA is a national trade organization for advertising agencies. Its members produce approximately 80 percent of all advertisements within the United States annually. Join Us, AM. ASS’N ADVERTISING AGENCIES, http://www.aaaa.org/about/Pages/default.aspx (last visited Jan. 11, 2012).

164 See Wood, supra 161, at 1.


166 See SAG Commercials Contract, supra note 165, § 5 (limiting scope of contract to commercials); id. § 4 (defining “commercials” as “short advertising or commercial messages made as motion pictures, 3 minutes or less in length, and intended for showing over television” (emphasis added)); AFTRA Commercials Contract, supra note 165, §§ 1(B), 4(A) (same).


168 Wood, supra note 161, at 1.
B. **The Nature of the Right**

Both the right of publicity and the Commercials Contracts create a framework for individuals to control their personas within commercials. The significant difference between the two is that the Commercials Contracts protect against misuse by the other party to the agreement, whereas the right of publicity protects against the world at large.

1. **The Nature of the Right of Publicity**

The right of publicity is “the inherent right of every human being to control the commercial use of his or her identity.”\(^{169}\) It protects the plaintiff from unauthorized commercial use that causes damage to the commercial value of his or her persona.\(^{170}\) The right of publicity treats personality as property that the owner can exclude others from using for commercial gain.\(^{171}\) As a result, unlike the right of privacy, publicity rights are fully assignable.\(^{172}\) Furthermore, many states allow a performer’s right of publicity to pass to that performer’s heirs.\(^{173}\) Remedies for an infringement of the right of publicity include an injunction against future use, statutory damages, compensatory damages, disgorgement of profits, punitive damages, and attorney’s fees.\(^{174}\) Therefore, the right of

\(^{169}\) 1 MCCARTHY, supra note 1, § 1:3.

\(^{170}\) Zacchini v. Scripps-Howard Broad. Co., 433 U.S. 562, 573 (1977) (“[T]he State’s interest in permitting a ‘right of publicity’ is in protecting the proprietary interest of the individual . . . .”); see also 1 MCCARTHY, supra note 1, § 3:2 (“Likely damage to commercial value is a hallmark of the right of publicity, distinguishing it from the various types of ‘privacy’ rights. However, this is not to state that evidence of some quantifiable commercial damage is an essential element of proof of liability for infringement of the right of publicity.” (footnote omitted)).

\(^{171}\) See RESTATEMENT (THIRD) OF UNFAIR COMPETITION § 46, cmt. g (“[T]he commercial value of a person’s identity is . . . freely assignable to others.”).

\(^{172}\) Some states allow estates to enforce the right of publicity for a limited number of years after death. See CAL. CIV. CODE § 3344.1(g) (West Supp. 2012) (granting the estate the right of publicity for up to seventy years after death). Other states grant the estate a right of publicity for a minimum number of years followed by a right in perpetuity, which permits enforcement for as long as the estate continues to exploit the persona commercially. See TENN. CODE ANN. § 47-25-1104 (2001) (granting the estate a minimum right for ten years, and a right in perpetuity that extinguishes after two years of disuse).

\(^{173}\) See 2 MCCARTHY, supra note 22, §§ 11:21-11:38.
Right of Publicity

2012] Right of Publicity 1155

Publicity is a guarantee of control over the commercial use of one's persona against the world at large.

Since the right of publicity derives from state law, the amount of protection varies from state to state.175 Nearly every state that recognizes the right protects against at least the unauthorized use of name and likeness.176 For example, New York, which recognizes only a statutory right and not a common law right,177 protects against the unauthorized use of “name, portrait, picture or voice.”178 Other states provide for a wider array of uses that may damage the commercial value of persona.179 For instance, California’s statutory right protects against “knowingly us[ing] another’s name, voice, signature, photograph, and likeness.”180 California’s common law protects further against additional types of use, including mannerism, characterizations, and performing style.181

Additionally, the right of publicity is limited by the context in which the use takes place. States place particular constraints on the context of the usage within their statutes or common law. These constraints exist to comply with First Amendment free speech, press, and expression, as well as to prevent the doctrine from stifling cultural exchange in society.182 For example, in New York, in order for unauthorized use to be an infringement, it has to be either for advertising or trade purposes.183 Courts broadly interpret advertising purposes to include usage in solicitation to buy products or services.184 Trade purposes are uses that draw attention to the defendant’s business, but do not directly solicit.185 Furthermore, New York

175 See 1 McCarthy, supra note 1, § 6:8.
176 See, e.g., id. (listing that every statute[ilily recognized right of publicity protects against the unauthorized use of name and likeness).
177 Stephano v. News Grp. Publ’ns, Inc., 474 N.E.2d. 580, 584 (N.Y. 1985) (ruling that the right of privacy is statutory, and there is no common law cause of action in New York).
178 Dymond, supra 23, at 447.
179 See, e.g., White v. Samsung Elecs. Am., Inc. 971 F.2d 1395, 1399 (9th Cir. 1991) (holding under California law that a robot in a commercial that was made to resemble Vanna White infringed on her common law right of publicity); Motschenbacher v. Reynolds Tobacco Co., 498 F.2d 821, 825-27 (9th Cir. 1974) (holding under California law that using a car that looked similar to the plaintiff’s racing car infringed on his right of publicity).
182 See Hunt, supra note 23, at 1629-39 (discussing restrictions on media and commercial speech).
183 1 McCarthy, supra note 1, § 6:86.
184 Id.
185 Id.
exempts certain uses without the subject’s consent, including use within the news,\textsuperscript{186} and incidental uses.\textsuperscript{187}

2. The Nature of Persona Rights Under the Commercials Contracts

Like the right of publicity, the Commercials Contracts allow principal performers control over the use of their persona, but within the confines of a contractual relationship. A number of specific contractual provisions demonstrate the nature and limit of this control.

First, and most importantly, the compensation structure protects performers from damage to the commercial value of their personas by placing a supplemental price on use of the commercial. All performers are entitled to session fees of varying amounts depending on their role within the production.\textsuperscript{188} Session fees represent the minimum compensation within the three-tier structure implemented by the union.\textsuperscript{189} Principal performers, as opposed to extras,\textsuperscript{190} are central to the purpose of the commercial; their personas are used and identifiable within the commercial.\textsuperscript{191} As a result, the Contract gives additional compensation to principal performers in the form of holding fees and residual compensation.

\textsuperscript{186} Id. § 6:93 (“[T]here is no doubt that there is no statutory liability for distribution and syndication of a general interest television news program which shows a film of the plaintiff as part of a newsworthy report.”). The First Amendment places additional limits on the right of publicity. See supra note 22. However, New York’s statutory scheme prohibits newsworthy uses in advertising material. See 1 McCarthy, supra note 1, § 6:89 (“The fact that newsworthy information also appears in a context that clearly advertises a product or service does not immunize what would otherwise be a violation of the statutory right.”).

\textsuperscript{187} Id. § 6:90 (“New York recognizes an ‘incidental use’ exception from statutory liability for insignificant or fleeting usages of persona that have no real commercial significance.”).

\textsuperscript{188} See SAG Commercials Contract, supra note 165, § 20 (“Producer shall pay principal performers the following rates per 8-hour day which shall also constitute payment for the first commercial made for one designated advertiser . . . .” (emphasis added)); APTRA Commercials Contract, supra note 165, § 20 (same); SAG Commercials Contract, supra note 165, sch. D, § 6 (detailing minimum rates for extras); APTRA Commercials Contract, supra note 165, sch. C, § 2 (same).

\textsuperscript{189} See supra note 143-45 and accompanying text.

\textsuperscript{190} See infra Part IV.B (discussing the distinction between principal performers and extras, and how that distinction relates to the incidental use doctrine within the right of publicity).

\textsuperscript{191} SAG Commercials Contract, supra note 165, § 6 (listing the types of uses that constitute principal performers, including speaking lines or silent appearance of a face that can be identified with a product); APTRA Commercials Contract, supra note 165, § 6 (same).
Holding fees are payments from the advertiser to the principal performer for each thirteen-week cycle that the advertisement airs.\textsuperscript{192} The session fee constitutes the holding fee for the first thirteen-week cycle,\textsuperscript{193} but subsequent cycles require payment of a holding fee equal to the original session fee.\textsuperscript{194} An advertiser’s failure to pay holding fees voids its right to use the commercial.\textsuperscript{195} Ultimately, holding fees place a price on airing commercials over an extended period, and compensate actors for damage done to their commercial value when advertisements are aired over a long period of time.

Traditionally, residual payments paid performers for each reuse of the commercial within the thirteen-week cycle.\textsuperscript{196} But under the contract, which is the product of over fifty years of negotiation,\textsuperscript{197} the exact calculation for residuals varies depending on the television channel and the geographic reach of the commercial broadcast. For example, Class A commercials, which run on traditional networks like NBC and air in sufficient locations, receive payments every time the commercial airs.\textsuperscript{198} Cable commercials, on the other hand, compensate principal performers with residual payments for only the first two thousand airings.\textsuperscript{199} To fix anomalies, the advertising industry has proposed calculating residuals across all television programming based on gross rating points (GRP)

\textsuperscript{192} SAG Commercials Contract, supra note 165, § 31; AFTRA Commercials Contract, supra note 165, § 31.
\textsuperscript{193} SAG Commercials Contract, supra note 165, § 31(c) (“The session fee shall be deemed the holding fee payable for the first fixed cycle.”); AFTRA Commercials Contract, supra note 165, § 31(c) (same).
\textsuperscript{194} SAG Commercials Contract, supra note 165, § 31(b) (“[U]pon the commencement of each consecutive fixed cycle thereafter throughout the maximum permissible period of use or any extension thereof, a principal performer shall be paid a separate fee, herein called the holding fee, in an amount equal to a session fee . . . .”); AFTRA Commercials Contract, supra note 165, § 31(b) (same).
\textsuperscript{195} SAG Commercials Contract, supra note 165, § 31(e) (“If Producer fails to pay the holding fee on or before the date on which it is due . . . all further right of Producer to use the commercial shall cease and terminate, and the principal performer shall thereupon be automatically released from all contractual obligations with respect to the commercial.”); AFTRA Commercials Contract, supra note 165, § 31(e) (same).
\textsuperscript{196} Susan Shelley, The Screen Actors Guild and the Commercials Strike of 1978-79 (1979) (research paper), available at http://www.extremeink.com/strike.htm (last visited Jan. 13, 2012) (“Actors, in other words, would receive continuing payments as long as their work was being used to generate revenue for an advertiser.”).
\textsuperscript{197} Neff, supra note 167 (“The current system is based on a model first developed in the 1950s . . . .”); see also Shelley, supra note 196 (describing that the general terms of the SAG Commercials Contract date back to the 1950’s).
\textsuperscript{198} SAG Commercials Contract, supra note 165, § 34(a)-b; AFTRA Commercials Contract, supra note 165, § 34(a)-b.
\textsuperscript{199} SAG Commercials Contract, supra note 165, § 35(c); AFTRA Commercials Contract, supra note 165, § 35(c).
rather than a pay-per-use standard. In effect, advertisers would pay residuals based on the number of viewers that saw the commercial rather than the number of times the commercial aired. Additionally, similar to states that allow a descendible right of publicity, the heirs of principal performers are entitled to residuals postmortem. In this sense, the residuals are analogous to a property right rather than a typical contractual right.

Regardless of how the contract calculates residuals, they represent more than mere supplemental compensation. Legally, despite several attempts to treat residuals as a form of property, the law treats them as supplemental compensation. Indeed, residuals account for the largest proportion of compensation under the Commercials Contracts. Performers often survive long periods of unemployment through residuals, as they are deferred compensation. Economically, however, residuals play a larger role than typical deferred compensation. Residual payments protect principal performers against overexposure by placing a price on the advertiser for every use of

---

200 See Neff, supra note 167 (“There can be anomalies in cable buys where the media costs less than the talent . . . . We hope to switch to the same GRPs and ROI measurement we use for other media and 99% of the dollars spent, so the tail will no longer be wagging the dog.” (quoting Douglas Wood)). During the 2009 negotiations, the parties agreed to retain a consulting firm to assess the feasibility of the proposal, and to start the 2012 negotiations early to devote time to consider the proposal. SAG Commercials Contract, supra note 165, § 21(a).


202 See Estates, SCREEN ACTORS GUILD, http://www.sag.org/content/estates (last visited Jan. 13, 2012) (“As you may know, residual payments are made in perpetuity (as long as a project is being exhibited somewhere in the world). Residuals are considered property (similar to a piece of artwork) and can be passed through a last will and testament and/or a living trust.”).

203 See, e.g., Gilbert, supra note 155, at 112 n.15 (“§ XXVII of the 1955 Filmed Commercials contract, which provides that ‘The right of a player to compensation for the use and re-use of a commercial shall be a vested property right and shall not be affected by the expiration of this contract or by any act on the part of the Producer.’” (emphasis added)).

204 See id. at 104 (“From a legal standpoint, the importance of recognizing that these residuals are a type of wage payment cannot be stressed too greatly.”), Paul & Kleingartner, supra note 130, at 672 (“In the 1950’s, residuals were seen purely as a mechanism to compensate workers for lost work and over-exposure.”).


206 See Paul & Kleingartner, supra note 130, at 672 (“[Residuals] cushion the impact of unemployment, especially among the neophytes who suffer long periods of unemployment . . . .”).
the commercial. Paul and Kleingartner explain, “Residuals reconcile that divergence of interests by specifying cash compensation for the presumed devaluation of future work via present-day overexposure.” Additionally, expanding residuals into new media forms maintains an adequate level of compensation when new technology reduces the demand for performers’ services. Therefore, like the right of publicity, residuals provide compensation to actors for damage done to their commercial personas.

The Commercials Contracts provide principal performers with other means of control beyond compensation. The Contracts limit producers’ power to use principal performers’ work to only agreed-upon commercials. If a producer uses a performer’s work in another commercial, the performer is entitled to liquidated damages equal to three session payments and usage fees, as well as residual fees that would be due had the performer consented to the use. Additionally, a performer has the option to arbitrate with or

---

207 Section 1(a) of the SAG Commercials Contract states that principal performers receive residual compensation because:

[A] principal performer rendering services in a commercial performs, to a great extent, the duties of a demonstrator or salesperson of a particular product or service and as such, tends to be identified with that particular product or service . . . [T]his identification increases proportionately with the continued telecasting of a commercial . . . [A]dvertisers and their agencies seldom approve the employment of a principal performer who has become identified with another product or service, especially if the product or service is competitive. These conditions and practices tend to reduce opportunities for further employment in this field.

208 Paul & Kleingartner, supra note 130, at 672.

209 Id. at 671 (“The single most important factor in the growth of residual compensation has been the expansion of residual obligations to new entertainment markets.”).

210 SAG Commercials Contract, supra note 165, § 17(a)-(b) (“The rights granted to Producer in commercials shall be limited to the right to use, distribute, reproduce and/or exhibit such commercials over television . . . . Producer agrees that no part of the photography or sound track of a principal performer made for a commercial shall be used other than in commercials as provided hereunder without separately bargaining with the principal performer and reaching an agreement regarding such use.”); AFTRA Commercials Contract, supra note 165, § 17(a)-(b) (same). But the contracts provide exceptions that producers can use the “name and likeness of the principal performer and his/her acts, poses and appearances . . . for the purpose of publicizing the business of the Producer.” SAG Commercials Contract, supra note 165, § 17(a); AFTRA Commercials Contract, supra note 165, § 17(a). This exception allows the advertising agencies, photographers and directors to show the content of their work to prospective clients without violating the contract.

211 SAG Commercials Contract, supra note 165, § 17(b); AFTRA Commercials Contract, supra note 165, § 17(b).
sue a producer rather than accepting liquidated damages provided within the Commercials Contracts.\textsuperscript{212}

Finally, the Commercials Contracts create options for principal performers to limit the maximum length an advertiser can run a commercial to twenty-one months.\textsuperscript{213} However, principal performers waive this right if they fail to give written notification of their desire to cease running the commercial.\textsuperscript{214}

Both the right of publicity and the Commercials Contracts provide actors ways to control the usage of their images within advertising. The right of publicity prevents unauthorized usage and compensates for damage done to the commercial value of actors’ personas. The Commercials Contracts give actors some control over the usage of their images in voluntary relations with advertisers, and compensate for commercial devaluation of their personas via residuals and holding fees. Given these analogous functions, the Commercials Contracts are an apt benchmark for comparison when considering both the policy justifications and doctrinal rules within the right of publicity.

III. ANALYZING THE RIGHT OF PUBLICITY’S POLICY JUSTIFICATIONS BY REFERENCE TO THE COMMERCIALS CONTRACTS

Justifying the right of publicity has sparked significant debate among scholars, as it “is both hard to object to and hard to support.”\textsuperscript{215} Having a compelling policy justification for the right of publicity is essential for two reasons. First, it provides a general rationale for the right of publicity’s existence. Second, policy justifications affect how judges and legislators shape the rules and doctrines within the right of publicity. Adopting a flawed policy justification can lead to illogical doctrine and

\textsuperscript{212} SAG Commercials Contract, \textit{supra} note 165, § 17(b); AFTRA Commercials Contract, \textit{supra} note 165, § 17(b).

\textsuperscript{213} SAG Commercials Contract, \textit{supra} note 165, § 30(a) ("[T]he maximum period during which a commercial may be used shall be not more than [twenty-one] months after the date of commencement of the first fixed cycle. . ."); AFTRA Commercials Contract, \textit{supra} note 165, § 30(a) (same).

\textsuperscript{214} SAG Commercials Contract, \textit{supra} note 165, § 30(d); AFTRA Commercials Contract, \textit{supra} note 165, § 30(d). The performer must give the advertiser written notification no earlier than sixty days before, and no later than 120 days after the twenty-first month of use. SAG Commercials Contract, \textit{supra} note 165, § 30(d); AFTRA Commercials Contract, \textit{supra} note 165, § 30(d).

inequitable consequences.\textsuperscript{216} The Commercials Contracts—with their common origin and similar protections to the right of publicity—can be used as a tool to analyze the merits of frequently asserted policy justifications for the right of publicity. This section concludes that any analogy to copyright or trademark\textsuperscript{217} fails in light of policy analysis and actual business custom. Instead, unjust enrichment and natural rights best explain the right of publicity.

A. Economic Incentives, Labor Theory, and Copyright

A common justification for the right of publicity is that it incentivizes creative endeavors, which benefits society as a whole.\textsuperscript{218} This justification can be broken down into two steps. First, it assumes that individuals create the value of their personas through labor and effort. Professor Haemmerli likens this explanation to Lockean labor theory.\textsuperscript{219} For example, in \textit{Zachini v. Scripps-Howard Broadcasting Co.},\textsuperscript{220} the only right of publicity case the United States Supreme Court has ever heard, the Court stated that “the State’s interest is closely analogous to the goals of patent and copyright law, focusing on the right of the individual to reap the reward of his endeavors.”\textsuperscript{221} Melville B. Nimmer, in his seminal article \textit{The Right of Publicity} stated:

It is also unquestionably true that in most instances a person achieves publicity values of substantial pecuniary worth only after he has expended considerable time, effort, skill, and even money. It would seem to be a first principle of Anglo-American jurisprudence . . . that every person is entitled to the fruit of his labors unless there are important countervailing public policy considerations.\textsuperscript{222}

\textsuperscript{216} For example, adopting a trademark dilution justification leads to the conclusion that non-celebrities should not be protected by the right of publicity. See infra notes 238-44 and accompanying text.

\textsuperscript{217} See supra notes 6-8 and accompanying text.

\textsuperscript{218} Kwall, supra note 56, at 35 ("Proponents of the right of publicity often rely on a copyright law analogy and argue that publicity rights are needed to spur incentives to creation just as copyright law exists, by constitutional command, to enhance economic incentives for the betterment of society." (footnote omitted)).

\textsuperscript{219} Haemmerli, supra note 9, at 412-13.

\textsuperscript{220} 433 U.S. 562 (1977).

\textsuperscript{221} Id. at 573.

\textsuperscript{222} Nimmer, supra note 127, at 216. Other commentators frequently invoke labor theory justifications. See Goldstein & Kessler, supra note 16, at 819 ("Justification for affording legal protection to the performer rests on the theory that anyone who contributes something of value to society should be entitled to share in the fruits of his labor."); Coyne, supra note 6, at 812 ("By permitting individuals to benefit from their personal efforts, both [the right of publicity and copyright] provide incentive for creative endeavor.").
The second step of the justification contends that protecting the fruits of labor incentivizes creativity, which benefits society. Zacchini concludes: “the protection provides an economic incentive for [the performer] to make the investment required to produce a performance of interest to the public.” This justification is analogous to the justification for copyright law.

Professor Madow questions both of these propositions. First, he argues that the celebrities with the most valuable images are not necessarily the hardest working. Second, he argues that there are powerful noneconomic motivations for fame, which dilute the need for the right of publicity as an incentive for creativity.

Looking to collective bargaining at the rank-and-file level within the entertainment industry shines some light on whether the most talented actors get the best roles. Many assume, to justify the higher compensation in the Contract, that union members are higher-quality actors compared with non-unionized actors. However, Emily C. Chi questions whether entertainment union members are really better, arguing that this assumption is a myth. According to her, entertainment unions “operate de facto closed shops.” Through strong union security clauses, and economic coercion of employers, the actors’ unions require actors to be members to get roles. However, actors have difficulty becoming

---

223 Zacchini, 433 U.S. at 576.
224 See id. (“This same consideration underlies . . . copyright laws long enforced by this Court.”); Coyne, supra note 6, at 813 (“Apparently, lurking beneath the surface in both publicity right and copyright decisions is the notion that ‘protection exists primarily not to benefit the artist, but rather to benefit the public by offering artists economic incentives to create.'” (quoting Note, Human Cannonballs and the First Amendment: Zacchini v. Scripps-Howard Broadcasting Co., 30 STAN. L. REV. 1185, 1192 (1978))).
225 Madow, supra note 3, at 213 (“A handful of ‘superstars’ command huge audiences and huge incomes, while everybody else—including persons only slightly less talented than the stars, or more talented and less lucky or ruthless—is ‘pushed to the back’ and ‘unrewarded.’”).
226 Id. at 214 (“There is, first of all, the desire for fame itself: for renown, for recognition, for glory, for liberation from powerless anonymity. There is the satisfaction of realizing and exercising one’s talents, of developing and displaying proficiency at some difficult or complicated activity. There is the pleasure of winning people’s applause, inspiring their love or awe, earning their respect or gratitude.” (footnote omitted)).
227 See, e.g., Memorandum from Douglas J. Wood, Reed Smith LLP, on Unions & the Prod. of Commercials for Traditional & Non-Traditional Media 3 (Jan. 28, 2009), available at http://www.adlawbyrequest.com/uploads/file/Why%20be%20an%20Authorizer.pdf (“Without doubt, union performers are the best professional and sought after performers for commercials. They understand their craft and bring great efficiency to the workplace.”).
228 Chi, supra note 83, at 65-70.
229 Id. at 11.
230 Id. at 37-44.
members of the actors’ unions.\footnote{Id.} For example, there are only three ways to become a member of SAG: to be cast in a SAG production as a principal performer,\footnote{Id. at 40.} to work as an extra on a SAG production for three days,\footnote{Id. at 41.} or to be a dues-paying member of an equivalent acting union.\footnote{See id. (describing how SAG controls the extras in many productions and how only a “small number of lucky extras may be assigned an ‘unscripted line’ while they are on the set”).} But since the unions control the supply of jobs in the industry, a non-union actor has significant difficulty joining SAG.\footnote{Id. at 68 (“Due to the element of arbitrariness in the determination of which actors become members of the union, the lack of regular . . . SAG-provided training to ensure some basic level of skill and experience, and the uncertainty regarding the reasons why producers have not substantially resisted union control over the acting labor supply, . . . SAG [cannot] hold itself out as an expert arbiter of quality or talent.”).} Therefore, the difference between those that make it into the union and those that do not becomes, in large part, a question of luck rather than skill.\footnote{Id.} Referencing the collective bargaining relationship here provides further support for Professor Madow’s position that the value of a personality in entertainment is not proportional to the amount of the individual’s effort. This undercuts the labor theory as a justification for the right of publicity and questions the validity of any analogy to copyright.

B. Diminution in Value and Trademark

Another common argument for the right of publicity is that it protects against diminishment in the value of persona. The right of publicity allows individuals to maximize upon the value of their persona through licensing at a specified price-per-use, which ensures that advertisers to whom the persona is most valuable will purchase it.\footnote{Madow, supra note 3, 223 (summarizing Richard Posner’s argument).} However, “[u]nrestricted use of a person’s name or likeness makes that name or likeness less scarce and thus, less valuable,”\footnote{Konsky, supra note 8, at 350.} which justifies a property protection. This justification is analogous to trademark dilution under the Lanham Act.\footnote{See 15 U.S.C. § 1125(c)(1) (2006) (“Subject to the principles of equity, the owner of a famous mark that is distinctive, inherently or through acquired distinctiveness, shall be entitled to an injunction against another person who, at any time after the owner’s mark has become famous, commences use of a mark or trade name in commerce that is likely to cause dilution by blurring or dilution by tarnishment of the famous mark.”); see also Greene, supra note 2, at 532-33 (“The

\footnotesize{231 Id.}
\footnotesize{232 Id. at 40.}
\footnotesize{233 Id. at 41.}
\footnotesize{234 See id. (describing how SAG controls the extras in many productions and how only a “small number of lucky extras may be assigned an ‘unscripted line’ while they are on the set”).}
\footnotesize{235 Id. at 68 (“Due to the element of arbitrariness in the determination of which actors become members of the union, the lack of regular . . . SAG-provided training to ensure some basic level of skill and experience, and the uncertainty regarding the reasons why producers have not substantially resisted union control over the acting labor supply, . . . SAG [cannot] hold itself out as an expert arbiter of quality or talent.”).}
\footnotesize{236 Madow, supra note 3, 223 (summarizing Richard Posner’s argument).}
\footnotesize{237 Konsky, supra note 8, at 350.}
\footnotesize{238 See 15 U.S.C. § 1125(c)(1) (2006) (“Subject to the principles of equity, the owner of a famous mark that is distinctive, inherently or through acquired distinctiveness, shall be entitled to an injunction against another person who, at any time after the owner’s mark has become famous, commences use of a mark or trade name in commerce that is likely to cause dilution by blurring or dilution by tarnishment of the famous mark.”); see also Greene, supra note 2, at 532-33 (“The}
that the violator’s use creates consumer confusion.\textsuperscript{240} Trademark dilution, an alternative theory, requires a showing that the mark is famous and distinctive but requires no showing of consumer confusion.\textsuperscript{241} Congress promulgated this section to protect famous trademarks from diminishment in value by overuse, even when the use causes no consumer confusion.\textsuperscript{242} Arguably, trademark dilution is analogous to the right of publicity because both aim to protect a commercially valuable intangible property from diminishing in value. As a result, commentators often assert that the right of publicity should be restricted by the same limits placed on trademark dilution,\textsuperscript{243} which would require the mark to be distinctive and famous.\textsuperscript{244}

But rank-and-file actors in commercials also face overexposure. An advertiser’s unrestricted use of a commercial leads to the principal performer becoming associated with the product. As a result, there is a risk that other advertisers will not hire the performer for new commercials.\textsuperscript{245} The Commercials Contracts address this overexposure problem with the compensation scheme; holding fees and residuals compensate principal performers proportional to the damage to their images from overexposure.\textsuperscript{246}

However, the Commercials Contracts also prove that non-famous individuals have valuable personas. Performers that receive contract-scale wages are not famous, as well-known individuals command significantly higher wages through overscale contracts.\textsuperscript{247} Additionally, unionized

\begin{itemize}
\item \textsuperscript{240} Konsky, \textit{supra} note 8, at 354.
\item \textsuperscript{241} \textit{Id.} (“Unlike a trademark infringement action, a trademark dilution action can be brought in the absence of consumer confusion about the goods.”); see also 15 U.S.C. § 1125(c)(1) (stating that plaintiff can seek injunction “regardless of the presence or absence of actual or likely confusion, of competition, or of actual economic injury”).
\item \textsuperscript{242} Konsky, \textit{supra} note 8, at 354 (“Trademark dilution law is concerned with protecting a trademark owner against uses that ‘whittle away’ the value of the mark, diminishing its uniqueness.”).
\item \textsuperscript{243} \textit{Id.} at 359 (“The current right of publicity should be replaced with a right of publicity dilution, similar to trademark dilution law. A right of publicity dilution would prohibit the most harmful uses of a person’s name or likeness without chilling valuable commentary.”).
\item \textsuperscript{244} \textit{Id.} at 355.
\item \textsuperscript{245} See \textit{supra} note 207.
\item \textsuperscript{246} See \textit{supra} notes 203-09 and accompanying text.
\item \textsuperscript{247} See Chi, \textit{supra} note 83, at 21 (“Most of the members of these unions are neither rich nor famous; it is only in the entertainment industry that huge disparities
advertisers must pay union wages to everyone that appears as a principal, including uncast individuals in reality situations. These uncast individuals are not well-known or famous, but business practice still compensates them for damage to their images through holding fees and residual payments. While the diminution-in-value justification is an apt similarity between trademark and right-of-publicity law, personalities are not trademarks. Transplanting trademark dilution onto the right of publicity discriminates against non-celebrities, as it would protect only famous individuals despite the business practice of compensating non-famous for damage to their personas.

C. Unjust Enrichment

Courts and commentators often invoke an unjust enrichment justification for the right of publicity. This justification maintains that the law must protect individuals from misappropriators. For example, the Supreme Court has stated that “[t]he rationale for protecting the right of publicity is the straightforward one of preventing unjust enrichment by theft of good will. No social purpose is served by having the defendant get free some aspect of the plaintiff that would have market value and for which he would normally pay.” Additionally, Judge Holmes, dissenting in O’Brien, advocated for a right-of-publicity-like right because appropriation “is contrary to usage and custom among advertisers in the marts of trade. They are undoubtedly in the habit of buying the right to use one’s name or picture to create demand and good will for their merchandise.”
Residual payments provide an apt analogy to the unjust enrichment justification for the right of publicity. The unions demanded residual payments in response to the emergence of new technology that allowed reproduction of actors’ performances. According to the union, compensation for the reuse of the broadcast was fair because the technology distributed its members’ performances for free when consumers would customarily pay for them. Therefore, the residuals compensate performers for the unjust enrichment that advertisers receive from reusing advertisements, as the industry has historically compensated actors for each performance. After referencing the business custom, the unjust enrichment justification is persuasive. This conclusion affirms that the right of publicity has a theoretically grounded purpose within our jurisprudence. Additionally, it suggests that when analyzing the doctrine within the right of publicity, the right should be approached from an unjust enrichment angle rather than a quasi-copyright or trademark.

D. Natural Rights Theory

Advocates for the right of publicity justify it by arguing control over persona is an innate and natural right. Infringing on the right of publicity damages more than the commercial value of the persona; it also takes away the individual’s natural right to control the use of their persona. Professor McCarthy, for one, argues that “[p]erhaps nothing is so strongly intuited as the notion that my identity is mine—it is my property to control as I see fit. Those who are critical of this principle should have the burden to articulate some important countervailing social policy which negates this natural impulse of justice.” Professor Haemmerli draws upon the philosophy of Immanuel Kant to intellectually strengthen the natural rights justification for the right of publicity. Haemmerli concludes,

253 See supra Part I.C. For example, in radio, the unions negotiated for residuals following the radio stations’ shift from having actors perform in each time zone to replaying the same recorded broadcast in each time zone. See supra notes 148-54 and accompanying text.
254 See supra notes 148-54 and accompanying text.
255 See, e.g., Haemmerli, supra note 9, at 385-86 (“The right of publicity can also be viewed as a property right grounded in human autonomy. As such, it belongs to all, . . . and it embraces noneconomic objections to the commercial exploitation of identity.” (footnote omitted)).
256 1 McCarthy, supra note 1, § 2:1.
257 Haemmerli, supra note 9, at 416.
“The central concept of autonomy in Kantian philosophy could lend itself to a philosophical justification of a right of publicity. Autonomy implies the individual’s right to control the use of her own person, since interference with one’s person is a direct infringement of the innate right of freedom . . . .”

Analyzing a natural rights justification from the perspective of a consensual agreement seems counterintuitive. However, the Commercials Contracts provide some support for the natural rights justification. When the performer agrees to appear in a specific commercial, the Commercials Contracts allow advertisers discretion over when, where, and how much the advertisement will run, like a property rule; yet, the advertiser has to compensate the performer for its usage, which acts as a liability rule. When the producer extends the performer’s use beyond the agreed-upon terms by using the performer in a new commercial, the performer can prevent the use by going through arbitration or the judicial process. Therefore, the idea that performers have an innate right to control their personas pervades even collective bargaining agreements in entertainment. The natural rights justification reinforces the unjust enrichment justification as well: rather than efficiently allocating property—like copyright or trademark—the right of publicity promotes fairness for individuals whose identities are misappropriated. Thus, the rules within the right of publicity ought to operate as natural property rights that prevent unjust enrichment.

IV. EVALUATING DOCTRINAL RULES WITHIN THE RIGHT OF PUBLICITY

The Commercials Contracts provide a point of reference to evaluate how the right of publicity actually operates within the law. Specifically, this part will analyze the underlying assumptions of rules that operate within the right-of-publicity jurisprudence by paralleling the rules to business practice, as seen through the Commercials Contracts.

---

258 Id.
259 See supra Part II.B.2.
260 See supra Part II.B.2.
261 See supra note 212 and accompanying text.
A. **Right of Publicity to Non-Celebrities**

The question of whether the right of publicity extends to non-celebrities is the most apparent division between jurisdictions. Because the right of publicity evolved out of celebrities’ struggles to recover under the right of privacy, the question of whether non-celebrities could recover for appropriation of commercial value of their personalities has divided both courts and commentators. The bulk of jurisdictions recognize a cause of action for individuals who are not famous personalities. These jurisdictions follow Melville B. Nimmer’s principle:

It is impractical to attempt to draw a line as to which persons have achieved the status of celebrity and which have not; it should rather be held that every person has the property right of publicity, but that damages which a person may claim for infringement of the right will depend upon the value of the publicity appropriated which in turn will depend in great measure upon the degree of fame attained by the plaintiff.  

---

262 See, e.g., Bowling v. Missionary Servants of the Most Holy Trinity, No. 91-5920, 1992 WL 181427, at *5 (6th Cir. July 20, 1992) (holding that under Kentucky common law a non-celebrity has a right of publicity cause of action); Motschenbacher v. R.J. Reynolds Tobacco Co., 498 F.2d 821, 825 n.11 (9th Cir. 1974) (recognizing in dicta that under California common law non-celebrities have a cause of action for the right of publicity); Fanelle v. Lojack Corp., No. CIV.A. 99-4292, 2000 WL 1801270, at *11 (E.D. Pa. Dec. 7, 2000) (“I am convinced that the right of publicity resides in every person, not just famous and infamous individuals.”); KNB Enters. v. Matthews, 92 Cal. Rptr. 2d 713, 722 n.12 (Cal. 2d. Dist. 2000) (“In our view, determining preemption of a plaintiff’s section 3344 claim on the basis of the plaintiff’s celebrity status would be violative of California law. Under California law, the statutory right of publicity exists for celebrity and non-celebrity plaintiffs alike.”); Dora v. Frontline Video, Inc., 18 Cal. Rptr. 2d 790, 792 n.2 (Cal. 2d. Dist. 1993) (recognizing in dicta that under California common law non-celebrities have a cause of action for the right of publicity); Martin Luther King, Jr., Ctr. for Soc. Change, Inc. v. Am. Heritage Prods., Inc., 296 S.E.2d 697, 703 (Ga. 1982) (concluding that the right of publicity exists “whether the person whose name and likeness is used is a private citizen, entertainer, or as here a public figure who is not a public official”); Fergerstrom v. Hawaiian Ocean View Estates, 441 P.2d 141, 144 (Haw. 1968) (holding that there was a right of publicity cause of action for a private citizen); Ainsworth v. Century Supply Co., 693 N.E.2d 510, 514 (Ill. App. 1998) (holding that the right of publicity extends to non-celebrities); Canessa v. J.J. Kislak, Inc., 235 A.2d 62, 75 (N.J. Super. Ct. Law. Div. 1967) (“[I]t seems to us that however little or much plaintiff’s likeness and name may be worth, defendant, who has appropriated them for his commercial benefit, should be made to pay for what he has taken . . . .”); Cohen v. Herbal Concepts, Inc., 473 N.Y.S.2d 426, 431 (N.Y. App. Div.1984) (“The legislative protection is clear, extending to ‘any person’ within the general public, not merely to those with a publicly identifiable feature . . . .”); see also 1 McCarthy, supra note 1, § 4:14 (“The majority of commentators and courts hold that everyone, celebrity and non-celebrity alike, has a right of publicity.”).

These jurisdictions presume that the non-famous individual's persona has commercial value based upon the defendant's usage in a way that exploits the persona for commercial benefit. This standard comports to the unjust enrichment and natural rights justifications: the advertiser receives the plaintiff's image without his or her consent, which violates the non-famous person's natural rights. And the advertiser would normally pay for such services, which suggests unjust enrichment. The question of the value of the plaintiff's persona only becomes a factor when determining compensatory damages. Although the amount of compensatory damages may be insignificant when a non-celebrity is commercially exploited, many states permit right-of-publicity plaintiffs to disgorge profits and recover punitive damages from misappropriators. For example, in Christoff v. Nestle, a jury awarded over $15 million dollars in profits to an unknown model, who had been working as a kindergarten teacher, for using his picture on coffee packaging.

Other courts have refused to extend the right of publicity to individuals with unknown personas. Often these

---

264 1 McCarthy, supra note 1, § 4:17 (“The courts use the commonsense rule that if defendant uses plaintiff's personal identity for commercial purposes, then it will be presumed that plaintiff's identity had commercial value. This presumption is similar to a presumption well established in trademark law.”).

265 Many commentators endorse, at least in theory, a broad right of publicity that applies to everyone because of natural rights theory. See Greene, supra note 2, at 538 (“[D]enying a publicity claim to a non-celebrity discounts personality rationales of personhood . . . . Arguments against non-celebrity right of publicity claims, regardless of merit, in effect value commercial speech over rights of personhood.” (footnote omitted)); see also Goodman, supra note 4, at 249 (“[T]he right of publicity must protect all persons and must not become a special interest right for celebrities only.”); Kwall, supra note 10, at 55-56 (“This Author believes that the right of publicity has the potential for safeguarding from unauthorized use any marketable and publicly recognizable attribute of any individual, regardless of whether that person is a celebrity.”); James M. Treece, Commercial Exploitation of Names, Likenesses, and Personal Histories, 51 Tex. L. Rev. 637, 648 (1973) (“[A]non-celebrity can argue, if he chooses, that an advertising use of his personality has unlawfully invaded an economic interest.”).

266 See supra note 249 and accompanying text.

267 2 McCarthy, supra note 22, § 11:34 (“In the analogous areas of trademark and copyright infringement, recovery of the profits made by defendant from the infringing sales are a standard form of monetary recovery.”).

268 Id. § 11:36 (“Under the law of most states, punitive or exemplary damages may be obtained in privacy and publicity suits.”).

269 Christoff v. Nestle, Inc., 213 P.3d 132, 133-34 (Cal. 2009). The Supreme Court of California later overturned the verdict because the statute of limitations barred plaintiff's claim, unless the defendant could show that the defendant had hindered the plaintiff's discovery of the unauthorized use of his likeness. Id. at 134.

courts require proof that a plaintiff's personality has objective value outside of the alleged commercial use. For example, in Cheatham v. Paisano Pub., Inc., the court developed a threshold test to determine whether a plaintiff's identity has commercial value: 271 “Commercial value may be established by proof of (1) the distinctiveness of the identity and by (2) the degree of recognition of the person among those receiving the publicity.” 272 Courts that follow this reasoning reject the presumption that the commercial exploitation by a defendant proves that a plaintiff's persona has value.

Several commentators have supported denying the right of publicity to non-celebrities. 273 In particular, those who advocate for a trademark-dilution analogy assert that the right of publicity should not protect non-celebrities because trademark law does not protect unknown marks. 274 Others argue that the First Amendment bars non-celebrity right-of-publicity claims. 275 For instance, Alicia Hunt argues that protecting the right of publicity for celebrities is a state interest substantial enough to pass the First Amendment's protection of claiming the infringement of this right must show that, prior to the defendant’s use, the plaintiff’s name, likeness, or persona had commercial value.”); Cheatham v. Paisano Pub., Inc., 891 F. Supp. 381, 386 (W.D. Ky. 1995) (ruling that plaintiffs must prove commercial value to establish a right of publicity claim); Barnako v. Foto Kirsch, Ltd., Civ. A. No. 86-1700, 1987 WL 10230, at *2 (D.D.C. Apr. 16, 1987) (“[T]he plaintiff must allege, and later prove, that the defendant’s commercial benefit was derived from the identity of the plaintiff and the value or reputation which the public associates with that identity.”); Jackson v. Playboy Enters., 574 F. Supp. 10, 13 (S.D. Ohio 1983) (“[T]he complaint must allege that plaintiff’s name or likeness has some intrinsic value, which was taken by defendant for its own benefit, commercial or otherwise.”); Ali v. Playgirl, Inc., 447 F. Supp. 723, 729 (S.D.N.Y. 1978) (suggesting in dicta that the right of publicity may not extend to non-celebrities under New York’s right to privacy statute); Cox v. Hatch, 761 P.2d 556, 564 (Utah 1988) (“[T]he complaint fails because it must allege that the plaintiffs’ names or likenesses have some ‘intrinsic value’ that was used or appropriated for the defendants’ benefit.” (citing Jackson, 574 F. Supp. at 13)).

271 Cheatham, 891 F. Supp. at 386.
272 Id.
273 See, e.g., Howard I. Berkman, Note, The Right of Publicity—Protection for Public Figures and Celebrities, 42 BROOK. L. REV. 527, 533 (1976) (“In a suit grounded upon the commercial appropriation of a private individual’s name or picture, the correct measure of damages is the extent of injury to the individual’s feelings and not the value that the defendant received from the unauthorized use of his name or picture.”).
274 See Dogan & Lemley, supra note 5, at 1166 (“Doctrinally, such an approach would limit the right to circumstances in which the use of an individual’s name or likeness in connection with the sale of a product is likely either to confuse consumers or to dilute the significance of a famous name.”); Konsky, supra note 8, at 366-70 (arguing for a requirement that the plaintiff be distinct and famous in order to recover for the right of publicity).
275 See, e.g., Hunt, supra note 23, at 1609 (“The extension of the right of publicity to non-celebrities is disturbing because in many instances, it interferes with the First Amendment’s protection of commercial speech.”).
commercial speech, but that the interest in protecting the right of publicity of non-celebrities is not substantial enough. Under the Central Hudson precedent, the standard for determining if restrictions on commercial speech violate the First Amendment, “the government may restrict truthful and nonmisleading commercial speech only if it proves (1) it has a substantial state interest in regulating speech, (2) the regulation directly and materially advances the interest, and (3) the regulation is no more extensive than necessary to serve that interest.”

Hunt concludes that the non-celebrities’ rights of publicity are not substantial under the test because non-celebrities’ identities have no provable economic value. Hunt questions the legitimacy of the presumption that non-celebrities have commercial value based on the manner that the infringer uses their persona, suggesting that the commercial value of non-famous people does not change by overuse.

Business practice under the Commercials Contracts sheds some light on this issue. In particular, business practice is integral to determine the rationality of the legal presumption that personas of non-celebrities have commercial value when commercially exploited. Undoubtedly, anyone working in a commercial under the union scale could hardly be considered “famous.” Nonetheless, these individuals receive

---

276 Id. at 1639-52.
277 Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm’n of N.Y., 447 U.S. 557 (1980). Various members of the Supreme Court have at times questioned the continued application of the Central Hudson standard. See 44 Liquormart, Inc. v. Rhode Island, 517 U.S. 484, 518-19 (1996) (Thomas, J., concurring) (arguing that in cases of restrictions on truthful commercial speech, Central Hudson balancing test should be replaced by per se violation of the First Amendment); id. at 517 (Scalia, J., concurring) (expressing “discomfort with the Central Hudson test, which seems to me to have nothing more than policy intuition to support it”); see also Sorrell v. IMS Health Inc., 131 S. Ct. 2653, 2664 (2011) (holding that content based restrictions on commercial speech receive heightened scrutiny rather than applying the balancing test established by Central Hudson).
278 Hunt, supra note 23, 1619-20.
279 Id. at 1643 (“[I]t is unlikely that the government’s interest would be ‘substantial’ in cases involving private individuals with no level of fame, notoriety or goodwill attached to their identities. In these cases, the commercial value is merely ‘presumed,’ even though there is no evidence that the identity in fact has any commercial value.”).
280 Id. at 1643-44.
281 “Union scale” refers to the minimum payments allowable under the Commercials Contracts. Commercial producers generally pay famous actors above the union scale. See Prindle, supra note 139, at 14 (“Before starting on a project, stars would have had their agents negotiate a work agreement much more favorable to themselves than is the standard SAG contract.”).
compensation for loss in the commercial value of their personas through residual payments. On the other hand, there may be an argument that an actor who consistently works under the Commercials Contracts has a persona with proven commercial value, and therefore passes the “intrinsic value” threshold. In commercials made under the Commercials Contracts, which make up 90 percent of all commercials made in the United States, all performers must be paid at least the union wage. If the advertiser uses uncasted performers in its commercials, those performers must also be paid at least the union scale—advertisers will on occasion pay residuals to anonymous individuals who are not professional actors. Therefore, advocates for a non-celebrity right are supported by industry practice in paying residuals to non-celebrities in the contractual setting.

B. Incidental Use Exception for Fleeting and Insignificant Use

Another doctrinal rule within the right of publicity that could be shaped by reference to the Commercials Contracts is the incidental use exception for fleeting and insignificant use. Seemingly every state that recognizes the right of publicity makes an exception to liability when the use is fleeting or insignificant. The straightforward reason for this exception is

---

282 Neff, supra note 167.
283 See SAG Commercials Contract, supra note 165, at Sch. B, § I(A)-(C) (requiring all principal performers to join the union within 30 days after hire in order to work on a union production); AFTRA Commercials Contract, supra note 165, at Sch. B, § I(A)-(C) (same); SAG Commercials Contract, supra note 165, at Sch. D, § 6(a) (setting forth minimum compensation for all extras on unionized productions); Commercials Contract, supra note 165, at Sch. C, § 2(a) (same).
284 See, e.g., PRINDLE, supra note 131, at 93 (“[A]dvertising agencies had fallen into the habit of using ‘real people’ in their TV commercials. Once they had appeared on screen, these nonprofessionals were eligible for a SAG card, which many of them acquired for its prestige value. No one in the guild knew how many members were thus nonactors, but there was widespread agreement that they swelled the membership lists without adding to the talent pool.”).
285 See, e.g., CAL. CIV. CODE § 3344(e) (West 1997) (“[I]t shall be a question of fact whether or not the use of the person’s name, voice, signature, photograph, or likeness was so directly connected with the commercial sponsorship or with the paid advertising as to constitute a use for which consent is required.”); FLA. STAT. ANN. § 540.08(3)(c) (West 2007) (mandating that the right of publicity does not apply to “[a]ny photograph of a person solely as a member of the public and where such person is not named or otherwise identified in or in connection with the use of such photograph”); NEB. REV. ST. § 20-202(3) (2007) (mandating that the right of publicity does not apply to “[a]ny photograph of a person solely as a member of the public when such person is not named or otherwise identified in or in connection with the use of such photograph”); OKL. ST. ANN. tit. 12, § 1449(e) (West 2010) (“[I]t shall be a question of fact whether or not the use of the person’s name, voice, signature, photograph, or
that, in these cases, the individual's persona is not being exploited commercially. Additionally, courts do not want to overburden expression by exposing parties to liability when an individual incidentally appears in media.

New York has the most developed incidental use doctrine. A court first invoked the incidental use doctrine in 1915, under the right-to-privacy statute (which today also protects the right of publicity), in *Merle v. Sociological Research Film Corp.* There, the court ruled that a glimpse of the plaintiff's business sign in defendant's movie did not violate right-to-privacy law. Specifically, the court held that for there to be a violation under the right-to-privacy statute, “it must appear that the use of the plaintiff's picture or name is itself for the purpose of trade and not merely an incidental part of a photograph.” The New York Court of Appeals in *Gautier v. Pro-Football Company* established a clear precedent when a defendant's use of plaintiff's name or likeness is incidental. In *Gautier*, the plaintiff was an animal trainer who performed during the half-time show for a professional football team. The defendant broadcasted part of the plaintiff's performance without the plaintiff's express consent. The plaintiff filed a right-of-publicity-like claim under the privacy statute and claimed that the defendant appropriated his image for commercial purposes because advertisements aired during the broadcast. The court ruled that the defendant's use of the plaintiff's image was incidental because the use was not

---

287 See *Preston v. Martin Bergman Prods., Inc.*, 765 F. Supp. 116, 120 (S.D.N.Y. 1991) (“The doctrine of incidental use was developed to address concerns that penalizing every unauthorized use, no matter how insignificant or fleeting, of a person's name or likeness would impose undue burdens on expressive activity, and carry consequences which were not intended by those who enacted the statute.”).
289 *Id.* at 831-32.
290 *Id.* at 832.
292 *Id.*
293 *Id.*
directly related to the commercial purpose of the advertisement. The court reasoned that the plaintiff “was not connected with the product either by visual, oral or other reference, nor was any issue of fact created by the physical juxtaposition of the [commercial] prior to his performance.” Consequently, the dispositive question under the Gautier decision is whether defendant’s use of plaintiff’s persona directly connects with plaintiff’s advertising or trade purpose.

With varying results, New York continues to follow the Gautier test to determine incidental use today. Often this test leads to fair results, for instance when it is clear that the defendant’s usage does not exploit the plaintiff’s persona. However, courts have ruled usage incidental when a plaintiff’s name or likeness clearly helped the defendant advertise. Given the factual nature of the incidental use doctrine, courts often leave the question to juries.

California’s right-of-publicity statute requires the plaintiff to be identifiable in pictures and exempts liability when the individual is part of a group where he or she is “represented in the photograph solely as a result of being present at the time the photograph was taken and have not been singled out as individuals in any manner.” Additionally, California has a common law incidental use doctrine, which follows New York’s jurisprudence. In Aligo v. Time-Life, Books Inc., the court set out a test for determining when use is incidental:

A number of factors are relevant in this regard: (1) whether the use has a unique quality or value that would result in commercial profit to the defendant, (2) whether the use contributes something of significance, (3) the relationship between the reference to the

---

294 Id. at 488.
295 Id.
296 See, e.g., Candelaria v. Spurlock, No. 08 Civ. 1830, 2008 WL 2640471, at *4 (E.D.N.Y. July 3, 2008) (holding that the four second appearance of a fast food worker in a documentary about fast food was incidental).
297 See, e.g., D’Andrea v. Rafia-Demetrious, 972 F. Supp. 154, 157-58 (E.D.N.Y. 1997) (holding that the defendant’s usage in a hospital brochure of a picture of the plaintiff working in a hospital was incidental because it did directly convey information about the hospital).
299 CAL. CIV. CODE § 3344(b)(1) (West 1997) (“A person shall be deemed to be readily identifiable from a photograph when one who views the photograph with the naked eye can reasonably determine that the person depicted in the photograph is the same person who is complaining of its unauthorized use.”).
300 Id. § 3344(b)(3).
plaintiff and the purpose and subject of the work, and (4) the
duration, prominence or repetition of the name, or likeness relative
to the rest of the publication.\textsuperscript{302}

The Commercials Contracts, and collective bargaining in
the entertainment industry more generally, provide an apt basis
for comparison to evaluate the rationality of the incidental use
doctrine. Specifically, the divide between principal performers
and extras in the Commercials Contracts mirror the incidental
use doctrine. Principal performers receive residual compensation
and holding fees for the usage of their personas;\textsuperscript{303} advertisers
pay extras only session fees and no residuals or holding fees.\textsuperscript{304}
Therefore, an advertiser's use of extras is analogous to
unauthorized incidental usage: neither requires compensation
for commercial exploitation of an individual's persona.

Historically, the entertainment industry treated acting
and extra work as two completely different trades. SAG and
AFTRA represented actors and performer; the Screen Extras
Guild (SEG) represented extras.\textsuperscript{305} For example, the National
Labor Relations Board excluded from an extras union's
jurisdiction individuals who performed more than extra work,
including stunts, singing, or performances involving lines.\textsuperscript{306}
The decision explains, "all work before the motion picture
camera falls primarily in two main classes, the one being
known as acting work . . . and the other being known as extra
work . . . customarily described in the industry as atmospheric
or background work."\textsuperscript{307} Although SEG disbanded in 1992 and
SAG acquired SEG's former jurisdiction,\textsuperscript{308} there is a historical
norm of separating principal performers from extras.

This norm is readily apparent in the Commercials
Contracts. The contracts provide a general definition of a
principal performer in a television commercial: "Anyone who is
seen and who speaks a line or lines of dialogue . . . ."\textsuperscript{309}

\textsuperscript{302} See id. at \textsuperscript{\#3} (citations omitted).
\textsuperscript{303} See supra notes 190-95 and accompanying text.
\textsuperscript{304} See SAG Commercials Contract, supra note 165, at sch. D, \S 6; APTRA
Commercials Contract, supra note 165, sch. C, \S 2.
\textsuperscript{305} Television and Movie Agreement—Collective Bargaining Agreement; Screen
Actors Guild, American Federation of Television and Radio Artists, Alliance of Motion
p/articles/mi_m1153/is_n8_v115/ai_12624085/.
\textsuperscript{307} Id. at 113.
\textsuperscript{308} Television and Movie Agreement, supra note 305.
\textsuperscript{309} SAG Commercials Contract, supra note 165, \S 6(A); APTRA Commercials
Contract, supra note 165, \S 6(A)(1).
“[a]nyone whose face appears silent, alone in a stationary camera shot, and is identified with the product . . . .”\textsuperscript{310}; or “[a]nyone whose face appears silent and is identifiable and whose foreground performance demonstrates or illustrates a product or service, or illustrates or reacts to the on- or off-camera narration or commercial message.”\textsuperscript{311} The contracts treat performers not within the general definition of the principal performer as extras,\textsuperscript{312} with several exceptions for, among other things, close-ups,\textsuperscript{313} stunt performers,\textsuperscript{314} dancers,\textsuperscript{315} and off-camera voice usage.\textsuperscript{316} The detailed distinction between a principal performer and an extra provides a framework to determine when a television commercial uses an individual’s persona.

The division between principal performers and extras leads to two observations about the right of publicity. First, it demonstrates that the connection between the advertising purpose and the individual’s part within the commercial affect whether the individual will receive compensation for reuse of persona. This can be seen in the definition of a principal performer within the Commercials Contracts for “foreground performance [that] demonstrates or illustrates a product or service or illustrates or reacts to the on or off-camera narration or commercial message.”\textsuperscript{317} This standard parallels Gautier’s directly-related-to-the-advertising-or-trade-purpose standard for incidental use in a right-of-publicity case.

Second, the Commercials Contracts present a reference point for courts to use when determining if an unauthorized appropriation of identity is incidental in a television commercial. Courts should find unauthorized use—and not incidental use—in a commercial that qualifies as principal

\textsuperscript{310} SAG Commercials Contract, supra note 165, § 6(B); AFTRA Commercials Contract, supra note 165, § 6(A)(2).

\textsuperscript{311} SAG Commercials Contract, supra note 165, § 6(C); AFTRA Commercials Contract, supra note 165, § 6(A)(3).

\textsuperscript{312} SAG Commercials Contract, supra note 165, § 6(C) (“Persons appearing in the foreground solely as atmosphere and not otherwise covered by the foregoing shall be deemed extra performers.”); AFTRA Commercials Contract, supra note 165, § 6(A)(4) (same).

\textsuperscript{313} SAG Commercials Contract, supra note 165, § 6(D); AFTRA Commercials Contract, supra note 165, § 6(A)(4).

\textsuperscript{314} SAG Commercials Contract, supra note 165, § 6(F) (“Stunt performers need not be identifiable per se; only the stunt performed need be identifiable.”); APTRA Commercials Contract, supra note 165, § 6(A)(6) (same).

\textsuperscript{315} SAG Commercials Contract, supra note 165, § 6(G); AFTRA Commercials Contract, supra note 165, § 6(A)(7).

\textsuperscript{316} SAG Commercials Contract, supra note 165, § 6(H); AFTRA Commercials Contract, supra note 165, § 6(A)(8).

\textsuperscript{317} SAG Commercials Contract, supra note 165, § 6(C); AFTRA Commercials Contract, supra note 165, § 6(A)(3).
performance under the Commercials Contracts. By referencing the Commercials Contracts, courts gain an elaborate benchmark when considering if the use is directly related to the advertising or trade purpose.

Use of the Commercials Contracts in this context is easy to apply. For example, in Pooley v. National Hole-in-One Ass’n, the plaintiff won a million dollars from the defendant for hitting a hole-in-one in a competition. The defendant used six seconds of footage of the plaintiff winning the competition without the plaintiff’s consent during an eight-minute infomercial advertising the defendant’s services. Although the plaintiff appeared only briefly, his appearance was integral to the infomercial, as he was the only winner of the defendant’s competitions. The court found, applying the Aligo standard, that the incidental use doctrine did not apply. Under the Commercials Contracts, the plaintiff’s part within the infomercial would qualify him as a principal performer. The plaintiff’s “face appear[ed]” and his “foreground performance demonstrate[d] or illustrate[d]” the defendant’s service, as he actually won the contest.

CONCLUSION

Courts and commentators should no longer automatically look to other intellectual property rights or academic frameworks when analyzing the right of publicity. Instead, with common origins and analogous rights to the right of publicity, collective bargaining agreements in entertainment provide an appropriate tool to scrutinize both the substantive rules and policy justifications for the right of publicity. Unjust enrichment and natural rights justifications for the right of publicity are supported by reference to the Commercials Contracts. Furthermore, the Commercials Contracts favor extending the right of publicity to non-celebrities. Finally, courts and lawmakers can import the Commercials Contracts when analyzing whether an unauthorized use of an individual’s image or persona is incidental.

The comparison to the Commercials Contracts is not necessarily limited to these points. The right of publicity and collective bargaining agreements both face the challenge of

319 Id. at 1110-11.
320 Id. at 1113 (“There is nothing, however, insignificant about the use of Plaintiff’s name and footage in the videotape. His name, while only briefly mentioned, prominently stands out as the highlight of Defendant’s advertisement.”).
321 See supra note 317 and accompanying text.
adapting to rapidly changing technology. When faced with these new technologies, looking to how collective bargaining agreements have dealt with these new obstacles may be the best way to shape the right of publicity in a rational manner going forward.

Andrew T. Coyle

---


1 B.S., Industrial and Labor Relations, Cornell University, 2008; J.D. Candidate, Brooklyn Law School, 2012. I wish to thank the Brooklyn Law Review staff for their assistance in the editing process, especially Elina Shindelman, Ari T. Tran, Shawna C. MacLeod, T. Daris Isbell, Philip F. Weiss, and Edward W. Thrasher. I also want to thank my family and friends for their constant support, including, but not limited to, Kaitlyn, Mom, Dad, Mike, and Barry.
Disability or Identity?

STUTTERING, EMPLOYMENT DISCRIMINATION, AND THE RIGHT TO SPEAK DIFFERENTLY AT WORK

INTRODUCTION

More than three million Americans stutter,¹ and stuttering affects about 1 percent of the worldwide population.² Stuttering refers to involuntary interruptions in a person’s speech, where speech “is broken by repetitions (li-li-like this), prolongations (llll-like this), or abnormal stoppages (no sound) of sounds and syllables.”³ The cause of stuttering is unknown,⁴ and no cure for the condition has been found.⁵ Stuttering ranges in degree from mild to severe, and it often leads to “physical tension and struggle” in the speech muscles.⁶ Significantly, most people who stutter experience feelings of embarrassment, anxiety, and fear.⁷

² Id.
⁷ In this note, the terms stutterer and person who stutters will be used interchangeably. For an insightful piece that discusses these terms as labels, see John C. Harrison, Are You or Are You Not a Stutterer?, STUTTERING HOMEPAGE (May 1, 1996), http://www.mnsu.edu/comdis/kuster/Infostuttering/stuttererornot.html; see also MARTY JEZER, STUTTERING: A LIFE BOUND UP IN WORDS 16-20 (1997).
Employment discrimination is a major concern for stutterers.9 Several studies indicate that people who stutter are disadvantaged in the employment context.10 According to 85 percent of employers in one study, “stuttering decreases a person’s employability and opportunities for promotion.”11 According to vocational rehabilitation counselors, who train individuals to enter the workforce, stuttering is “handicapping.”12 Because stutterers are concerned with discrimination they might face in the workplace, this note considers the extent to which stuttering is covered under federal antidiscrimination statutes. One way to proscribe discrimination based on stuttering is to consider stuttering a disability under the Americans with Disabilities Act (ADA) as amended in 2008;13 another way is through Title VII of the Civil Rights Act of 1964.14 Congress would need to amend the Civil Rights Act to accomplish this. This note examines both of these possibilities.

This note argues that the federal government should ban discrimination based on stuttering. This note also argues that the law must carefully contemplate the nature of stuttering; in crafting stuttering antidiscrimination law, policymakers must acknowledge the population of stutterers who do not stutter often but who are still greatly limited by their stuttering, and they must determine how to provide legal protection for these individuals.15 Either of the two alternatives mentioned above can solve this problem, and this note will demonstrate how both of these solutions might play out. Ultimately, while coverage under Title VII would be more extensive, coverage under the amended ADA is more practical.

Part I of this note explains what stuttering is, including the physical and emotional components of stuttering. Part I also describes the common misconceptions of stuttering and documents why workplace discrimination based on stuttering is a problem that needs to be addressed. Part II discusses the

10 Id.
11 Id. (citing M.I. Hurst & E.B. Cooper, Employer Attitudes Toward Stuttering, 8 J. FLUENCY DISORDERS 1 (1983)).
12 Id. (citing M.I. Hurst & E.B. Cooper, Vocational Rehabilitation Counselors' Attitudes Toward Stuttering, 8 J. FLUENCY DISORDERS 13 (1983)).
15 See infra Part I.
original ADA, explaining why it was passed and the statutory scheme through which to bring a discrimination claim. Part II also documents cases involving stuttering discrimination claims that have been brought under the ADA as well as Supreme Court cases in which the ADA's definition of disability was interpreted narrowly. Part III analyzes the ADA Amendments Act (ADAAA), the purposes of the amendments, and how the amendments affect stuttering discrimination claims. Part IV describes how it would be possible for an employee to bring a stuttering discrimination claim under the amended ADA. In Part V, this note proposes an alternative way to view stuttering—as an element of one's personhood and identity, rather than as a disability. Under this identity model, stuttering discrimination would be covered under Title VII of the Civil Rights Act. Part V goes on to compare the two characterizations of stuttering and concludes that, although the identity model is more effective in obtaining stuttering discrimination coverage under the law, viewing stuttering as a disability is the more realistic path to coverage.

I. OVERVIEW OF STUTTERING: SCIENTIFIC BACKGROUND, STEREOTYPES, AND EVIDENCE OF EMPLOYMENT DISCRIMINATION

Stuttering is complex and variable (in kind and degree), and this contributes to misunderstanding and prejudice against stutterers. This part describes what stuttering is and highlights both the physical and emotional aspects of the condition. In documenting the common stereotypes of stutterers, this part demonstrates how these misconceptions limit employment opportunities for people who stutter. Ultimately, this part shows that employment discrimination based on stuttering is a widespread problem. It also describes the emotional element of stuttering, which courts and lawmakers should consider in order to adequately combat stuttering discrimination in the workplace.

A. Stuttering as a Physical and an Emotional Condition

There are two major components of stuttering that must both be analyzed to determine a person's severity of
stuttering. First, there is the physical or behavioral part—what an outside observer can perceive. Second, there is the emotional part, which is characterized by a stutterer’s feelings and attitudes about stuttering. Severity—or how much someone is limited by stuttering—is the sum of both of these physical and emotional factors for any one person and is therefore highly individualized.

The physical or behavioral characteristics of stuttering are important because they describe the actual disfluency that the outside listener can hear. Factors included in this analysis are how often moments of stuttering occur, how long they last, how much struggle is involved with them, and the types of disfluencies that are involved. Many people think of a severe stutterer as someone who has frequent moments of stuttering that tend to last a long period of time (numerous seconds).

The emotional aspects of stuttering, though, also play a large role in assessing a stutterer’s severity—the actual extent to which a person is affected by stuttering. Emotional issues do not cause stuttering; rather, they are often an effect of stuttering. People who stutter often have feelings associated with their stuttering, such as nervousness, anxiety, fear, frustration, shame, and guilt. But people who stutter also have certain attitudes associated with stuttering—that stuttering is bad or wrong, or that stuttering is a sign of weakness and

17 Id.
18 Id.
19 Id. Regarding word choice, this note will not often use the terms mild, moderate, and severe to refer to degree of stuttering because the terms are misleading. This note will instead discuss how often someone physically gets into a stuttering block by using terms that refer to the frequency of stuttering. For example, a way to discuss degrees of stuttering is to say that some people stutter frequently, while for other individuals it is barely noticeable. When this note does refer to mild or severe stuttering, though, assume that this is a reference to frequency of overt stuttering unless otherwise noted.
20 Id.
21 Id.; see also Barry Yeoman, Wrestling with Words, PSYCHOL. TODAY, Nov./Dec. 1998, at 42, 44 (describing how for some stutterers the condition “means an intense and visible struggle to force individual syllables through their lips, a phenomenon that is physically exhausting for the speaker and mentally awkward for the listener”).
22 Hood & Roach, supra note 16.
23 Id.
25 Hood & Roach, supra note 16.
failure. People who have infrequent moments of stuttering as well as these feelings and attitudes often consciously try not to have a physical block whenever possible. If they are successful at avoiding physical blocks, the emotional aspects of stuttering may increasingly affect them; in this way, the severity of their stuttering conditions would be higher than observable.

Some people are so successful at hiding the physical aspects of stuttering that they become covert stutterers; they are able to hide their stuttering to such a great extent—through various tricks and crutches—that they are able to pass as fluent speakers. Covert stutterers pay an insufferable price, though, because they fear the constant risk of being exposed.

One covert stutterer described the cost of hiding: “Constant terror! Fear, panic and anxiety lived with me every waking minute and even into sleep. Thoughts of discovery paralyzed me.” This demonstrates the inaccuracy of describing a covert stutterer as having a mild problem. Crucial to this understanding is that there is not only an emotional aspect to stuttering but that it can actually serve to reduce the frequency of actual physical stuttering blocks. In other words, the emotional aspect can lead an individual to use extremely emotionally taxing behaviors to disguise the physical aspect.

Every person stutters with different levels of physical and emotional severity. An accurate analysis of how much a
person is affected by stuttering would include an assessment of both the physical and emotional aspects.35 This is a widely held view in the field of speech pathology.36 Moreover, this background knowledge is crucial to fully understanding the next section’s discussion of stuttering in the workplace.

B. Stuttering Stereotypes and Their Relation to Stuttering in the Workplace

People who stutter are subject to society’s widespread negative stereotypes about stuttering.37 There is a widespread belief that stutterers as a group “exhibit certain negative personality traits such as being shy, quiet, nervous, tense, afraid, self-conscious, etc.”38 Moreover, there are common myths that people who stutter are not as intelligent as those who are fluent39 and that underlying nervousness causes stuttering.40

Stuttering in the workplace is a significant issue for people who stutter.41 A large study of employers’ attitudes toward stuttering conducted in 1983 demonstrates employers’ widespread negative attitudes toward people who stutter: 30 percent of employers thought that stuttering interferes with job performance, 40 percent thought it negatively affects promotion possibilities, 44 percent thought that stutterers should seek employment that does not require a lot of speaking, and 85 percent thought that stuttering decreases employability to at least some degree.42 In a 1994 survey of people who stutter, 16 percent of the stutterers had been told that they would not be hired because of their stuttering, more than half thought that their supervisor had misjudged their capabilities because of stuttering, and more than one-third received negative

35 Hood & Roach, supra note 16.
36 See J. Scott Yaruss, Assessing Quality of Life in Stuttering Treatment Outcomes Research, 35 J. FLUENCY DISORDERS 190, 190 (2010) (noting how “stuttering can involve far more than just observable speech disfluencies” and citing a plethora of sources).
38 Id.
42 Id. (citing Hurst and Cooper, supra note 11, at 1).
performance evaluations because of stuttering. A 2004 survey also documented that people who stutter believe that their stuttering is a “major handicap” in their working lives. More than 70 percent of the stutterers surveyed thought that they had a decreased opportunity to be hired and promoted than nonstutterers, and 69 percent believed that their past job performance was hindered because of stuttering.

II. THE ORIGINAL AMERICANS WITH DISABILITIES ACT: BACKGROUND AND CASES INVOLVING STUTTERING

One possible way to deal with employment discrimination based on stuttering is through the federal statute that protects employees who have disabilities, the Americans with Disabilities Act. This part begins by discussing Congress’s purpose in passing the Act, and it continues by describing how an employee may bring a discrimination claim under the ADA. Several employees who stutter have brought ADA claims against their employers, but these claims have been largely unsuccessful. One significant reason for this lack of success is that the U.S. Supreme Court had narrowly interpreted certain key provisions of the ADA regarding the definition of disability. Ultimately, this part demonstrates that while there was once some likelihood that the ADA would cover stuttering, by the early 2000s the Supreme Court had significantly limited the probability that the ADA would adequately cover claims of discrimination based on stuttering.

A. Purpose of the ADA

According to Congress, the ADA is meant to counter discrimination faced by people with disabilities who “have often had no legal recourse to redress such discrimination.” Congress found that discrimination against people with

---

44 Id. at 266.
45 Id.
46 42 U.S.C. §§ 12101-12213 (2006 & Supp. II 2009). This note focuses on Title I, which deals with employment discrimination. Id. § 12112. There are two other main parts to the ADA. Title II prohibits discrimination in public services. 42 U.S.C. § 12132 (2006). Title III prescribes discrimination in public accommodations. Id. § 12182.
47 Id. § 12101(a)(4).
disabilities was not only a historical problem but a problem that continues to affect American society in a “persist[ent]” and “pervasive” way, too often according people with disabilities “an inferior status in our society.” President George H. W. Bush stated that the ADA would “signal[] the end to the unjustified segregation and exclusion of persons with disabilities from the mainstream of American life.” While the existing Rehabilitation Act of 1973 had been effective, its area of coverage was limited and it “left broad areas of American life untouched or inadequately addressed.”

B. Framework of the ADA

There are several steps to determining discrimination based on disability under the ADA. Courts apply the McDonell Douglas test, which is the standard for most discrimination litigation. To be successful, the employee must first satisfy three requirements to establish a prima facie case of discrimination. First, the claimant must establish disability. Under the ADA, there are three different ways an individual can show disability, any one of which is enough to establish that a disability exists. A plaintiff can establish the presence of a disability if the condition fits within the statutory

48 Id. §§ 12101(a)(2)-(3), (6).
50 Id. at 1165. The Rehabilitation Act of 1973 was the first federal handicap discrimination statute. 29 U.S.C. §§ 791-94a (2006). It had only a limited effect, though, because it applied just to the federal government, federal contractors, and federal grant recipients. Larry M. Schumaker, The Americans with Disabilities Act of 1990, 47 J. MO. B. 542, 542 (1991). Conversely, the ADA has a significantly wider impact on employment: it covers all public and private employers with fifteen or more employees. 42 U.S.C. § 12111(5)(A). Congress defined disability under the ADA in the same way that handicap was defined in the Rehabilitation Act, and Congress expected the ADA definition to be applied consistently with the definition in the Rehabilitation Act. Schumaker, supra, at 543. While this note does not focus on the Rehabilitation Act because of the ADA’s wider scope, the Rehabilitation Act is relevant, and this note will refer to it and its case law where appropriate.
54 Andresen, 2004 WL 2931346, at *4.
definition of a disability—“a physical or mental impairment that substantially limits one or more [of an individual’s] major life activities.”56 An employee can also establish the presence of a disability if there is “a record of such an impairment” as denoted in the above statutory definition.57 The third way an individual can show disability is if he or she is “regarded as having such an impairment.”58 Here, regarded as means that the employer (the person who took the adverse—and potentially discriminatory—employment action) is the one who is regarding (perceiving) the individual as having a disability (even if the person does not actually have a disability). Assuming that a claimant can establish that a disability exists under one of these three possibilities, the next part of the prima facie case requirement is to determine whether the person is “qualified to perform the essential functions of the job, with or without reasonable accommodation.”59 The third and final element necessary to establish a prima facie case is that the employee suffers an “adverse employment action” because of the disability.60 If the claimant can establish a prima facie case, a burden-shifting analysis ensues. The employer can rebut this presumption of discrimination “by articulating a legitimate, non-discriminatory reason for the adverse employment action.”61 If the employer can do this, the burden then shifts back to the plaintiff “to demonstrate that the employer’s non-discriminatory reason is pretextual.”62

C. Stuttering Cases Under the ADA

Several cases decided prior to the ADAAA held that stuttering is not a disability. The principal reason for this interpretation was that the plaintiffs failed to claim that stuttering was a significant obstacle in their lives. For example, in Zhong v. Tallahatchie General Hospital and

---

56 Id. § 12102(1)(A). The Equal Employment Opportunity Commission’s (EEOC) regulations refer to this prong as “actual disability” in order to distinguish it from the other two prongs (not to suggest that this prong brings with it greater rights). Regulations to Implement the Equal Employment Provisions of the Americans with Disabilities Act, As Amended, 76 Fed. Reg. 16,978, 16,980 (Mar. 25, 2011). This note will use this term where appropriate to avoid confusion.
58 Id. § 12102(1)(C).
60 Id.
61 Id. (quoting Kiel v. Select Artificials, Inc., 169 F.3d 1131, 1135 (8th Cir. 1999)).
62 Id. (quoting Kiel, 169 F.3d at 1135).
Extended Care Facility, a hospital said that it fired a medical technologist because he was incompetent. The fired employee filed several discrimination claims, one of which was an allegation that he was improperly terminated under the ADA because of his stutter. The District Court for the Northern District of Mississippi dismissed the claim because the plaintiff, in his deposition, described his stuttering as “very mild, very, very mild” and occurring “[o]nly occasionally.” Plaintiff also testified that his stuttering did not affect his work and that he was able to perform his required tasks “without an accommodation.” In granting summary judgment to the defendant, the court stated that, based on plaintiff’s own testimony, plaintiff’s stuttering clearly did not substantially limit his ability to speak or work and, further, that he had not shown that defendant regarded him as having a disability.

The court in Preacely v. Schulte Roth & Zabel similarly held that stuttering is not a disability. There, a legal word processor was fired because the law-firm employer said that he “ma[de] his co-workers feel uncomfortable and unsafe” due to his inappropriate comments and drawings. The plaintiff claimed stuttering discrimination in violation of the ADA. In affirming the district court’s grant of summary judgment to the law firm, the Second Circuit Court of Appeals highlighted that the former employee “admitted that his stutter was neither a physical nor a mental disability . . . and that it did not interfere with his ability to work or talk.”

The court’s decision in Detko v. Blimpies Restaurant also involved denial of a stuttering discrimination claim—albeit under Title III of the ADA, which involves claims regarding discrimination by private entities in public accommodations, rather than a discrimination claim under Title I. The plaintiff, a customer of defendant Blimpies

---

64 Id.
65 Id. at *3.
66 Id.
67 Id.
68 Preacely v. Schulte Roth & Zabel, 17 F. App’x 57, 58 (2d Cir. 2001).
69 Id.
70 Id. at 58-59.
71 Detko v. Blimpies Rest., 924 F. Supp. 555, 557 (S.D.N.Y. 1996). Detko is still relevant because the same definition of disability in Title I applies to Titles II and III of the ADA, and whether the plaintiff was disabled was the issue in the case. See 42 U.S.C. § 12102(1) (Supp. II 2009).
Restaurant, filed a claim of discrimination on the basis of his stuttering after an incident that occurred when he attempted to order a sandwich in the restaurant. The facts, as alleged by the plaintiff, are particularly interesting. He claimed that he tried to order a sandwich with extra mayonnaise and stuttered on the word “mayonnaise.” The employee serving him, who turned out to be the store’s manager, yelled at him to “[h]urry it up,” and the customer “became embarrassed and distressed.” After the manager stopped preparing the customer’s sandwich, the customer asked to speak with the manager to file a complaint. Meanwhile, another employee served the customer the sandwich. The manager refused to identify himself, threw the sandwich in the trash, grabbed the plaintiff “by the neck[,]” and dragged him out of the restaurant. The court granted Blimpie’s motion to dismiss because the customer did “not allege[] that his impediment substantially limit[ed] his speaking or that he is regarded as having such an impairment. . . . [H]e merely allege[d] that he stutters, and has particular difficulty with the letter ‘M.’”

In Zhong, Preacely, and Detko, the courts hearing the cases did not find stuttering to be a disability under the ADA. But, these results may be attributable to the fact that each of the three plaintiffs failed to allege that stuttering was a substantial limitation—or admitted outright that it was not. Thus, these plaintiffs do not appear to be exemplars by which to determine stuttering’s status under the ADA.

The court in Andresen v. Fuddruckers, Inc. sent a strong signal that stuttering could be covered under the ADA if a plaintiff could demonstrate that her stuttering was severe. Fuddruckers restaurant terminated the plaintiff’s employment after sixteen years of service. While the restaurant claimed that the former employee was fired for poor performance and for drooling and spitting into food, plaintiff “allege[d] that she was fired because she stutters.” The court denied Fuddruckers's

---

72 Detko, 924 F. Supp. at 556.
73 Id.
74 Id.
75 Id.
76 Id.
77 Id.
78 Id.
79 Id. at 557 (second emphasis added).
81 Id. at *1.
82 Id.
motion for summary judgment, finding a genuine issue of fact as to whether plaintiff’s “stuttering constitute[d] a ‘disability’ under the ADA.” Plaintiff claimed that her stuttering was severe, and that she avoided saying certain words and sounds and entering into speaking situations; she also alleged that she had difficulty speaking in general, especially when communicating with strangers and on the telephone, and that people had difficulty understanding her. A speech pathologist confirmed that her stuttering was severe. Also, the former employee sometimes had excess saliva that, according to her, only happened when she stuttered. Notwithstanding her stuttering and saliva issues, Andresen enjoyed a lengthy, successful term of employment with Fuddruckers until new managers took over the restaurant. Plaintiff’s “evidence [was] sufficient to create a genuine issue of fact that her stuttering [was] severe and that it substantially limit[ed] her ability to speak.” In its analysis, the court examined Zhong and Preacely and distinguished them on their facts. Those precedents did “not persuade the Court that [plaintiff’s] stuttering [could not], as a matter of law, constitute a disability.” The court also determined “that a triable issue of fact exist[ed] as to whether Andresen was qualified to do her job . . . [and] whether she was terminated because of her stuttering.”

The Equal Employment Opportunity Commission’s (Commission or EEOC) decision in Manning v. United States Postal Service also demonstrates that stuttering can be considered a disability. The complainant, Robert Manning, described his stuttering as severe. He said that he could not speak in public and was embarrassed to take classes. Manning claimed that he noticed derogatory graffiti on the stalls in two men’s restrooms: the writing mocked his stuttering. Although Manning noticed the graffiti in late winter or early spring 2000

---

82 Id. at *6, *8.
83 Id. at *1.
84 Id.
85 Id.
86 Id. at *1-3.
87 Id. at *6.
88 Id. at *6 & n.10 (noting that the plaintiffs in Zhong and Preacely both admitted that they do not stutter frequently).
89 Id. at *6-7.
90 Manning, E.E.O.C. Dec. 01A42153, 2004 WL 1810386, at *1 (Aug. 5, 2004). This case was before the Commission under the Rehabilitation Act, as Manning worked for a federal agency. Id.
91 Id.
92 Id.
93 Id.
and discussed the matter with his supervisor, the graffiti was not cleaned off until October 27, 2000.94 It then reappeared and remained on the walls until the complainant left the Postal Service in March 2001.95 A psychiatric evaluation stated that Manning only stuttered mildly,96 but the Commission reversed the administrative judge’s ruling of a decision without a hearing.97 The Commission found a “genuine issue of material fact” as to whether the complainant was “substantially limited in the major life activity of speaking.”98 According to the Commission, a hearing was not only necessary to resolve the issue of fact, but it was crucial: the hearing would give the administrative law judge an opportunity to hear how Manning stuttered.99

This finding connotes the individualized inquiry that would be done were stuttering considered solely under a disability theory; stuttering would not be a disability for every person who stutters, but only if it substantially affects that person’s speaking.100 So, while not yet finding that Manning’s stuttering substantially limited his speaking, the Commission noted that his stuttering could be deemed substantially limiting if the facts at the hearing bore that out.101 The Commission further ruled that Manning could possibly be covered under the regarded-as prong of the definition of disability—which protects against discrimination based on stereotypes—if further facts demonstrated this.102 The Commission noted that stuttering is a condition that is characterized by stigmatizing stereotypes and “attitudinal barriers” that can affect a stutterer’s employment opportunities.103 These stereotypes include the beliefs that people who stutter are “nervous, shy, quiet, self-conscious, withdrawn, tense, anxious, fearful and guarded.”104 In sum, then, considering these pre-ADAAA cases, courts only found stuttering to be a disability when the physical component of stuttering severely

94 Id.
95 Id.
96 Id. at *2.
97 Id. at *3.
98 Id.
99 Id.
100 See id.
101 Id.
102 Id. at *3-4. Note that under the ADAAA, as explained infra, a claimant can meet the requirements for the regarded-as prong without showing that he or she was regarded as being substantially limited in a major life activity. See infra Part III.B.3.
104 Id.
affected an individual. In part, the Supreme Court’s interpretation of the ADA required this narrow definition of disability, as discussed in the next section.

D. Supreme Court Cases Limiting the Definition of Disability Under the ADA


In *Sutton*, the Supreme Court addressed the question of whether corrective or mitigating measures must be considered when determining whether an impairment substantially limits a major life activity, the first prong of the ADA test for determining whether a disability exists.\(^{105}\) The case arose after two sisters applied for employment with United Air Lines as commercial airline pilots.\(^{106}\) Both sisters had severe myopia; their eyesight was poor enough that without corrective lenses, they could not participate in numerous daily activities.\(^{107}\) However, with glasses or contact lenses, they could see as well as people who did not have impaired eyesight.\(^{108}\) After United Air Lines initially invited them for an interview, it realized that they did not meet the company’s minimum uncorrected vision requirement.\(^{109}\) United Air Lines subsequently canceled the interview, and the sisters filed suit, alleging disability discrimination under the ADA.\(^{110}\) They specifically claimed that “they actually [had] a substantially limiting impairment or [were] regarded as having such an impairment.”\(^{111}\) The district court and the Court of Appeals for the Tenth Circuit found that they had not stated a claim of disability within the meaning of the ADA.\(^{112}\)

The Supreme Court affirmed with respect to the first claim and held that if a person is taking measures to correct or mitigate an impairment, the effects of those measures “must be taken into account when judging whether that person is ‘substantially limited’ in a major life activity and thus ‘disabled’ under the Act.”\(^{113}\) To reach this result, the Court

---


\(^{106}\) Id. at 475.

\(^{107}\) Id.

\(^{108}\) Id.

\(^{109}\) Id. at 476.

\(^{110}\) Id.

\(^{111}\) Id.

\(^{112}\) Id. at 476-77.

\(^{113}\) Id. at 477, 482 (quoting 42 U.S.C. § 12102(2)(A) (2006)).
looked to three separate provisions of the ADA.\(^\text{114}\) The Court first reasoned that the phrase *substantially limits* applies to the present, so that a “person whose physical or mental impairment is corrected by medication or other measures does not have an impairment that presently ‘substantially limits’ a major life activity.”\(^\text{115}\) Next, the Court reasoned that because a determination of disability under the ADA is an “individualized inquiry,” judges should examine an individual based on that person’s actual condition, not general information on how a group of people with the same impairment is usually affected.\(^\text{116}\) For the individualized analysis to be accurate, the Court noted that judges must consider a person’s use of mitigating measures.\(^\text{117}\) Finally, the Court looked at the number of people with disabilities that Congress cited in the ADA—forty-three million—to conclude that the legislature intended to take a “functional approach to determining disability” rather than a nonfunctional approach.\(^\text{118}\) Therefore, because the plaintiffs wore corrective lenses, they could not successfully make a claim that they were substantially limited in any major life activity.\(^\text{119}\) The Court in *Sutton* also provided a strict standard for the regarded-as prong of the disability definition. In order to be regarded as disabled, the Court held, the employer must regard the claimant substantially limited in a major life activity; thus, even under the regarded-as prong, substantial limitation is the standard for disability.\(^\text{120}\)

2. *Toyota Motor Manufacturing, Kentucky, Inc. v. Williams*

*Toyota* further limits the definition of disability under the ADA. In *Toyota*, the Supreme Court needed to interpret the terms *substantially* and *major* to determine whether a person

\(^{114}\) *Id.* at 482.

\(^{115}\) *Id.* at 482-83.

\(^{116}\) *Id.* at 483.

\(^{117}\) *Id.* at 483-84.

\(^{118}\) *Id.* at 484-87. The functional approach does not include people who successfully use mitigating measures to overcome their limitations. *Id.* at 485. The nonfunctional approach, also known as the “health conditions approach,” “looks at all conditions that impair the health or normal functional abilities of an individual.” *Id.* at 485, 487. Using this approach, over 160 million Americans would be considered disabled. *Id.* at 487.

\(^{119}\) *Id.* at 488-89.

\(^{120}\) *Id.* at 491, 493.
was disabled under the ADA.\textsuperscript{121} The case involved Ella Williams, who had been employed in an automobile manufacturing plant by Toyota Motor Manufacturing (Toyota).\textsuperscript{122} Over the course of several years, Williams developed pain and was diagnosed with carpal tunnel syndrome and several other conditions.\textsuperscript{123} This made it difficult for her to continue working at Toyota.\textsuperscript{124} After Williams's employment was terminated, she filed suit against her former employer, claiming that Toyota violated the ADA.\textsuperscript{125} Williams claimed that she was disabled because her physical impairments substantially limited her in six ways,\textsuperscript{126} each of which she claimed was a major life activity.\textsuperscript{127} She also alleged that she was disabled under the ADA because she had a record of impairment and because she was regarded as having an impairment.\textsuperscript{128} While the district court ruled in favor of Toyota, the Court of Appeals for the Sixth Circuit reversed and found that Williams was disabled because she had been substantially limited in performing manual tasks.\textsuperscript{129} The Supreme Court granted certiorari "to consider the proper standard for assessing whether an individual is substantially limited in performing manual tasks."\textsuperscript{130}

In deciding to strictly interpret both \textit{substantially} and \textit{major}, the Supreme Court reversed.\textsuperscript{131} The Court reasoned that these terms must be interpreted strictly, in part because of Congress's intent, as discussed in \textit{Sutton}.\textsuperscript{132} In order to be substantially limited in performing manual tasks, the Court held that "an individual must have an impairment that prevents or severely restricts the individual from doing activities that are of central importance to most people's daily lives [and that the impairment's impact must also be
Taking into account the individualized inquiry required by the statute, and also the fact that symptoms can “vary widely from person to person” with certain impairments, the Court reasoned that mere evidence of a medical diagnosis would be “insufficient . . . to prove disability status;” rather, whether a person has a disability under the ADA must be based upon the extent of that individual’s impairment. When analyzing the major life activity of performing manual tasks, the Court stated that the main question must be whether the individual “is unable to perform the variety of tasks central to most people’s daily lives,” not just the tasks associated with that person’s specific job. This was a crucial point because “the manual tasks unique to any particular job are not necessarily important parts of most people’s lives.” In this case, there were some manual tasks that Williams was able to do at work, and outside of work she was able to perform many of the manual tasks that are central to most people’s daily lives. Therefore, the court of appeals was incorrect to find that Williams was disabled under the ADA.

III. THE ADA AMENDMENTS ACT

In response to how the ADA had been interpreted by the Supreme Court in *Sutton* and *Toyota*, Congress passed the ADA Amendments Act in 2008. The Amendments Act significantly broadened the scope of disability under the ADA. After expanding on Congress’s purpose in passing the ADAAA, this part details the specific changes in the ADAAA and discusses how each of the changes affects discrimination claims brought on the basis of stuttering. Taken as a whole, the ADAAA decreases plaintiffs’ burdens to show that their stuttering is a disability.

---

133 *Id.* at 198.
134 *Id.* at 198-99.
135 *Id.* at 200-01.
136 *Id.* at 201.
137 *Id.* at 202.
138 *Id.* at 203.
A. Congressional Purpose

One of the major purposes of the ADAAA was to place the emphasis of ADA claims on whether a qualified person has been discriminated against on the basis of disability, rather than on the preliminary question of whether a plaintiff is disabled.\textsuperscript{141} Congress deemed the definition of disability less important than the determination of whether covered entities complied with their obligations not to discriminate.\textsuperscript{142} The amended statute embraces a wider-encompassing meaning of disability, defining the term “in favor of broad coverage of individuals.”\textsuperscript{143} Congress removed two original findings from the ADA because they provided a justification for the Supreme Court to narrowly construe the definition of disability:\textsuperscript{144} (1) that there are forty-three million Americans with disabilities,\textsuperscript{145} and (2) that individuals with disabilities constitute “a discrete and insular minority.”\textsuperscript{146} Removing the findings enlarges the class of individuals that the statute is intended to protect and, by extension, allows for an increasing number of impairments to be considered disabilities.\textsuperscript{147} Moreover, the old ADA prohibited discrimination against a qualified individual “with a disability because of the disability of such individual,”\textsuperscript{148} while the ADAAA prohibits discrimination against a qualified individual “on the basis of disability.”\textsuperscript{149} Therefore, the major purpose of the ADAAA is to make it easier for individuals to be

\textsuperscript{141} Id. at 3554; see also 29 C.F.R. § 1630.1(c)(4) (2011) (“The primary object of attention in cases brought under the ADA should be whether covered entities have complied with their obligations and whether discrimination has occurred, not whether the individual meets the definition of disability. The question of whether an individual meets the definition of disability under this part should not demand extensive analysis.”).
\textsuperscript{142} ADA Amendments Act of 2008 § 2, 122 Stat. at 3554.
\textsuperscript{144} ADA Amendments Act of 2008 § 3, 122 Stat. at 3554-55.
\textsuperscript{145} 42 U.S.C. § 12101(a)(1).
\textsuperscript{146} Id. § 12101(a)(7).
\textsuperscript{147} See ADA Amendments Act of 2008 § 2, 122 Stat. at 3553 (“[L]ower courts have incorrectly found in individual cases that people with a range of substantially limiting impairments are not people with disabilities . . . .”).
\textsuperscript{149} Id. § 12112(a) (Supp. II 2009). This (subtle) change in language—specifically, the removal of “with a disability”—further demonstrates the way Congress wanted courts to more easily dispense with the question of whether an individual is disabled. UNDERSTANDING THE NEW DISABILITY AND GENETIC DISCRIMINATION LAWS 2008, at 23 (Joyce Gentry et al. eds., 2008) [hereinafter UNDERSTANDING]. In fact, this language was structured on Title VII of the Civil Rights Act. Id.
considered disabled under the statute. In and of itself, this legislative purpose makes it more likely that individuals who stutter will be protected by the ADAAA.

B. Specific Changes Under the ADAAA and How They Relate to Stuttering

The amendments made numerous changes to the ADA. The changes affect several areas: the definition of substantially limits, episodic impairments and impairments in remission, which activities are considered major life activities, effects of mitigating measures, and the requirements of the regarded-as prong. Each of these changes has substantial implications regarding the extent to which stuttering is considered a disability under the ADA.

1. Speaking as a Major Life Activity

No major life activities were listed in the old ADA. The amended statute, though, provides a nonexhaustive list of activities. Under the ADAAA, speaking is explicitly listed as a major life activity. In our society, a limitation on one’s ability to speak can interfere with life activities in which people without such limitations regularly engage. Speaking plays a vital role in communication. If there was any doubt as to the significance of speaking in people’s daily lives, the statute now removes the ambiguity. Further, because of the “substantially limits” requirement, it would not make sense for the statute to only cover a total inability to speak—i.e., muteness. For many people who stutter, their speech impediment substantially limits—but does not preclude—their ability to speak. Therefore, the inclusion of speaking as a major life activity would make it easier for stuttering to be considered a disability under the ADA.

---

150 See ADA Amendments Act of 2008 § 2, 122 Stat. at 3554 (stating that the purpose of the amendments is “to carry out the ADA’s objectives . . . by reinstating a broad scope of protection to be available under the ADA”).
151 42 U.S.C. § 12102(2) (2006) (disability is defined, but no list is given); UNDERSTANDING, supra note 149, at 21. The EEOC did promulgate a regulatory definition, though, and speaking was on that list. 29 C.F.R. § 1630.2(i) (2010) (amended 2011).
153 Id.
2. Substantially Limits: A Less Demanding Standard

The ADAAA establishes that the substantially limits requirement was to be construed significantly more broadly than courts had been interpreting the term.\(^\text{154}\) Moreover, there are three specific ways in which \textit{substantially limits} has become a more inclusive standard.

First, only one major life activity needs to be substantially limited for an individual to have a cognizable disability under the ADAAA.\(^\text{155}\) Therefore, it would suffice if stuttering substantially limited speaking without substantially limiting any other major life activity.

Second, the amended statute provides coverage for impairments that are episodic or in remission, so long as they fit the statutory definition of disability when they are active.\(^\text{156}\) A useful test for determining if an individual's impairment is substantially limiting is whether an individual's activities are limited in “condition, duration and manner.”\(^\text{157}\) Stuttering could be considered episodic because of how the physical stuttering block does not occur constantly. Sometimes, a person does not physically stutter for long periods of time; there may be days or more between stutters.\(^\text{158}\) This person who stutters only intermittently would likely be covered under this amendment. When a person is stuttering, the involuntary interruptions can substantially limit one's ability to speak.

Third, the ADAAA states that the use or lack of use of mitigating or corrective measures cannot be taken into account

\(^{154}\) Id. § 12102(4)(B) (Supp. II 2009) (“The term ‘substantially limits’ shall be interpreted consistently with the findings and purposes of the ADA Amendments Act of 2008.”); ADA Amendments Act of 2008 § 2, 122 Stat. at 3554. Specifically, one purpose of the amendments is to convey congressional intent that the standard for “substantially limits” as articulated in \textit{Toyota} “has created an inappropriately high level of limitation necessary to obtain coverage under the ADA.” Id. Additionally, Congress found that the EEOC regulations that defined “substantially limits” as ‘significantly restricted’ are inconsistent with congressional intent, by expressing too high a standard.” Id.

\(^{155}\) 42 U.S.C. § 12102(4)(C) (Supp. II 2009) (“An impairment that substantially limits one major life activity need not limit other major life activities in order to be considered a disability.”); 29 C.F.R. § 1630.2(j)(1)(viii) (2011). This is the same as under the original ADA. 42 U.S.C. § 12102(2)(A) (2006).

\(^{156}\) 42 U.S.C. § 12102(4)(D) (“An impairment that is episodic or in remission is a disability if it would substantially limit a major life activity when active.”); 29 C.F.R. § 1630.2(j)(1)(vii).

\(^{157}\) UNDERSTANDING, supra note 149, at 18.

\(^{158}\) See Hood & Roach, supra note 16.
when determining whether or not a person has a disability.159 The statute lists various mitigating measures that cannot be considered: (1) medication and medical supplies and equipment, (2) assistive technology, (3) reasonable accommodations or auxiliary aids or services, or (4) learned behavioral or adaptive neurological modifications.160

There are several ways—including speech therapy161 and use of assistive devices162—people can attempt to mitigate their stuttering. One of the reasons this provision is so important in relation to stuttering is that each of these measures has varying levels of effectiveness for each person who uses them. Just as the cause of stuttering is not understood, what makes these methods effective is also not understood. In this way, it is not a person’s fault if these measures do not work to reduce a person’s stuttering.163 In turn, it would be unfair if the availability of mitigating factors weighed against considering stuttering a disability; this would harm those people on whom these techniques were not effective. Furthermore, some people who stutter do not believe in using these methods, and it would be unfortunate to create a situation where—because stuttering could not be recognized as a disability—there is more pressure on people who stutter to use these measures because there would be no other recourse in the workplace.

3. Regarded-As Prong Changed

The ADA amendments both broaden and narrow the scope of coverage under the regarded-as prong of disability. A person can now be regarded as having a disability if this

---

159 42 U.S.C. § 12102(4)(E)(i) (“The determination of whether an impairment substantially limits a major life activity shall be made without regard to the ameliorative effects of [numerous enumerated] mitigating measures . . . .”); 29 C.F.R. § 1630.2(j)(1)(vi); see also ADA Amendments Act of 2008 § 2, 122 Stat. at 3554 (explicitly noting a purpose of the amendments to reject Sutton, which had held that impairments need to be determined with regard to mitigating measures).
160 42 U.S.C. § 12102(4)(E)(i)(I)-(IV). Ordinary eyeglasses and contact lenses, though, can be considered in determining whether an impairment substantially limits a major life activity. Id. § 12102(4)(EX)(ii).
161 See JEZER, supra note 7, at 68, 76 (noting that there are two different major approaches to speech therapy, but explaining that “what works in the clinic doesn’t easily carry over into the real world”).
individual can show discrimination based on an actual or perceived impairment, even if the impairment does not limit—or is not perceived to limit—a major life activity. In this way, Congress reinstated the reasoning of School Board of Nassau County, Florida v. Arline.

In Arline, an elementary-school teacher was fired after she suffered a third relapse of tuberculosis in the span of two years. She brought suit, claiming a violation of the Rehabilitation Act. The Supreme Court found that she was a person with a handicap under the regarded-as prong. Under that part of the definition, the “negative reactions of others” to an impairment can limit a person’s ability to work. In explaining the regarded-as prong, the Court looked to congressional intent to reason that “society’s accumulated myths and fears about disability and disease are as handicapping as are the physical limitations that flow from actual impairment.”

Coverage under the regarded-as prong was narrowed under the amendments, though, in two ways. First, impairments that are transitory and minor are not covered under the regarded-as prong. Stuttering is permanent, so it would not be restricted by this provision. Second, and more importantly in this context, the ADAAA does not require reasonable accommodations to be made for people who only fit the definition of disability under the regarded-as prong. This part of the statute resolved a circuit split over whether the third prong required reasonable accommodations.

---

164 42 U.S.C. § 12102(3)(A) (“An individual meets the requirement of ‘being regarded as having such an impairment’ if the individual establishes that he or she has been subjected to an action prohibited under this chapter because of an actual or perceived physical or mental impairment whether or not the impairment limits or is perceived to limit a major life activity.” (emphasis added)); 29 C.F.R. § 1630.2(j)(2).
166 Arline, 480 U.S. at 276.
167 Id.
168 Id. at 284-86.
169 Id. at 283.
170 Id. at 284.
171 42 U.S.C. § 12102(3)(B) (Supp. II 2009) (defining transitory as being six months or less); 29 C.F.R. § 1630.15(f) (2011). But, a transitory impairment can fit the definition of disability under the actual-disability prong or the record-of prong. Id. § 1630.2(j)(1)(ix).
172 42 U.S.C. § 12201(h); 29 C.F.R. § 1630.9(e). Because of this, the EEOC noted that it is unnecessary to proceed under the actual-disability or record-of prong when an individual is not seeking a reasonable accommodation; the analysis could then be made solely under the regarded-as prong. Id. § 1630.2(g)(3).
173 UNDERSTANDING, supra note 149, at 24-25.
This is quite significant in relation to stuttering because it addresses the reality that there may be less protection for people who do not stutter frequently under the actual-disability prong of disability.\footnote{The reason for this is that infrequent stuttering may not be viewed as a substantial limit on speaking. See supra Part I.} Under this provision, if a person who stutters is discriminated against because of stuttering, this individual can be regarded as having a disability—even if courts would not consider stuttering as limiting the person’s speaking. This provision seems to allow people who do not stutter frequently to nevertheless gain protection against discrimination. Furthermore, and quite significantly, this may be incentive for people who choose to hide their stuttering—by avoiding speaking situations—to speak up, with the knowledge that (even based on just a few physical blocks) any discrimination can have a legal remedy.

C. ADAAA Case Law

In \textit{Medvic v. Compass Sign Co., LLC}, a Title I case analyzed under the ADAAA, plaintiff’s claim of discrimination based on stuttering survived defendant’s motion for summary judgment.\footnote{\textit{Medvic v. Compass Sign Co., LLC}, Civ. A. No. 10-5222, 2011 WL 3513499, at *1, *10 (E.D. Pa. Aug. 10, 2011).} The plaintiff, Donald Medvic, was employed as a sheet-metal mechanic for the defendant, Compass Signs.\footnote{\textit{Id.} at *1.} Medvic stuttered, and, although he never asked for an accommodation, his supervisors were aware of his speech impediment because they could hear it.\footnote{\textit{Id.} at *1-2. The court described Medvic’s stutter as making it difficult for him to communicate orally and sometimes causing him to be unable to say what he wants to say for several minutes. \textit{Id.} at *1. Medvic’s coworkers, at least once, needed to help him order food while out for dinner. \textit{Id.} at *2. Medvic accused his supervisors of making fun of his stuttering. Specifically, he claimed they asked him to sing for them and would tell him to just spit it out. \textit{Id.} at *3. But Medvic said that his stutter did not affect his ability to do his job, and the defendant agreed; once, Medvic’s supervisors considered warning a customer about his stuttering, but they ultimately chose not to do so because he “always found his way.” \textit{Id.}} Medvic was laid off, and he brought two claims against Compass Signs based on the ADA—that his termination was due to his stuttering disability, and that he was subjected to a hostile work environment.\footnote{\textit{Id.} at *1.} In evaluating Compass Signs’s summary-judgment motion, the court first analyzed whether Medvic was disabled under the
terms of the ADAAA. The court found that there was a genuine issue of fact as to whether Medvic was substantially limited in communicating: stuttering sometimes delayed his speech for minutes at a time, stuttering impeded his social life, he stuttered during his deposition, and his coworkers testified to his stuttering hindering his ability to communicate. The company argued that, because Medvic sat for his deposition, he could not have been substantially limited in his ability to communicate, but the court rejected this argument. The court maintained that Medvic could still be substantially limited in communicating even if he could communicate effectively sometimes. The court cited the ADAAA for further support. Then, the court found that Medvic also survived summary judgment regarding whether he was otherwise qualified for the job and whether the company's action was taken because of his stuttering.

IV. BRINGING A WORKPLACE STUTTERING DISCRIMINATION CLAIM UNDER THE AMENDED ADA

The ADAAA was intended to make it easier for claimants to show that they have a disability under the statutory definition of that term. The amendments do make it easier for stutterers to show that they have a disability. This part methodically goes through the statute to document how a successful stuttering discrimination claim can be brought under the ADAAA. The analysis will show how applying the complex nature of stuttering to the definition of disability under the ADAAA can result in broad antidiscrimination coverage for people who stutter. Significantly, one way to address this problem of the emotional aspect of stuttering is to demonstrate that even an infrequent overt stutterer can be considered substantially limited in speaking. In fact, this part will focus on this type of stutterer because of the law's heretofore lack of attention to this area. Still, there are

---

179 Id. at *5-7.
180 Id. at *6-7. The court seems to stress that Medvic stuttered frequently and with great struggle, "at times rendering him incapable of verbally communicating for himself." Id. at *7. The court referred to the ADAAA: it noted that Medvic took medicine to help him stutter less but that mitigating measures must not be considered, and it explained that his stutter is substantially limiting even though it is episodic. Id. at *7.
181 Id. at *7.
182 Id.
183 Id. The court pointed to the ADAAA here to demonstrate that Congress intended to broaden protections. Id.
184 Id. at *7-10.
shortcomings of coverage in approaching stuttering as a disability.

A. Establishing Stuttering as a Disability Under the ADAAA

1. Stuttering Substantially Limits a Major Life Activity

Under the ADAAA, a person who stutters may be able to show that stuttering substantially limits speaking, a major life activity under the Act. This would be an individualized inquiry. But for the inquiry to be proper, it would need to take into account both the physical and the emotional aspects of stuttering. With respect to the physical aspect, the law already accounts for a certain sect of the stuttering population.\textsuperscript{185} People who stutter frequently should be able to show that their disfluencies are substantially limiting, while people who stutter infrequently would not be able show that their speech impediments substantially limit their speaking.

The law, however, has not accounted for the emotional aspect of stuttering. Nevertheless, stutterers who are greatly affected by the emotional aspect of stuttering should also be able to show that they are substantially limited in a major life activity. Stuttering can be quite debilitating for an individual, even if an outside listener does not hear many physical stuttered words.\textsuperscript{186} This mental struggle, while less overt, can nevertheless be substantially limiting. Significantly, the ADA does cover mental disabilities in addition to physical disabilities.\textsuperscript{187} Like certain forms of mental illness, stuttering is sometimes not readily apparent to the outside observer.\textsuperscript{188} Coverage for these types of stutterers is not only necessary to seriously deal with stuttering, but it is also practical, as it is possible to gauge the extent to which a stutterer is emotionally affected by stuttering. Significantly, and rather than relying upon the plaintiff’s testimony, there is a way for speech pathologists to assess stuttering’s full impact on a person’s quality of life—an assessment instrument known as the Overall Assessment of the Speaker’s Experience of Stuttering

\textsuperscript{185} See supra Parts II & III.
\textsuperscript{186} See supra Part I.
\textsuperscript{187} See supra Part II.
\textsuperscript{188} While this note is not attempting to equate mental illness with stuttering, the ADA does cover mental disabilities. 42 U.S.C. § 12102(1) (Supp. II 2009).
Using the OASES would add a measure of objectivity and uniformity to the task of assessing the emotional impact of stuttering on one’s life. Although this type of stutterer may be less likely to be discriminated against because this person’s physical stuttering is infrequent, qualifying as disabled is nevertheless important because it would allow for reasonable accommodations.

2. Regarded as Disabled

The third way to establish a disability—the regarded-as prong—was included in the ADA to protect disabilities not noticeable to the naked eye. Therefore, it is especially relevant to stuttering. This provision was first included in the Rehabilitation Act to protect employees who were discriminated against whether or not they were recognized as handicapped under the definition of the statute. In Arline, the Supreme Court interpreted this provision to include an expansive definition of perceived handicaps, finding that Congress was concerned with “protecting individuals from discrimination based on outdated and stereotypic laws and attitudes.”

The third prong and its interpretation in Arline are crucial to protecting claimants who suffer from a disability that is perceived based on untrue stereotypes. So many of the problems facing people who stutter are based on false stereotypes and myths. Therefore, even if an employee is not substantially limited by stuttering, his or her employer might perceive the employee as being substantially limited because of the prevalence of these preconceived notions of stuttering. In this way, the regarded-as prong adds a significant layer of protection to those who face discrimination on the basis of stuttering. If, for example, a person is an infrequent stutterer, the speech impediment may not be too bothersome, and this

---

189 See generally Yaruss, supra note 36 (explaining this kind of evaluation).
190 Returning to the plaintiffs in Zhong, Preacely, and Detko, perhaps they would have fit into this category of individuals. We do not know, but they could have been given a quality-of-life assessment. See supra note 189 and accompanying text. It is also strange that these plaintiffs would downplay the limiting nature of the very impairment on which they were bringing their disability lawsuit. Perhaps they were covering—downplaying their speech difficulties in order to seem more normal. See infra notes 226, 228 and accompanying text.
192 Id. at 258-59.
person may not be substantially limited in speaking. But, if an employer hears this individual stutter and discriminates on that basis because of a stereotype associated with stuttering, this person would have a claim under the regarded-as prong. This employee would not be entitled to a reasonable accommodation, but if this individual stutters infrequently and is not bothered by stuttering, it is unlikely that any accommodation would even be desired.

Another instance of the regarded-as prong applying to stuttering is if a person stutters frequently but is not bothered by his or her stuttering. Such an individual would not be substantially limited in speaking because this person would not view stuttering as limiting. But, this person may still face discrimination based on stuttering. Although this employee may fail to qualify as a person with a disability under prong one, prong three should provide coverage: this individual is regarded as being disabled even though the individual does not view stuttering like this. No accommodation would be requested in this situation, as the employee would not feel limited in speaking. This should not be a catchall for truly severe stutterers, though. While it would be a fallback option, a severe stutterer should attempt to show disability under prong one so that he or she is entitled to reasonable accommodations.

B. Determining a Qualified Individual

Once a claimant can establish disability under the terms of the ADAAA, this person must then be able to show the requisite qualifications for the position at issue. The term ‘qualified individual’ means an individual who, with or without reasonable accommodation, can perform the essential functions of the employment position . . . .” Deference is given to the employer in determining which job functions are essential.

People who stutter may find themselves in a “catch-22” situation when speaking or possessing excellent communication skills is an essential job requirement: “If they prove they are ‘substantially impaired’ in speaking, they will not be ‘qualified’ for the job.” “On the other hand, if they prove that they are

---

194 Id. § 12111(8).
195 Id.
196 Parry, supra note 9.
197 Id.
‘qualified’ to hold a speaking job, they will not be ‘substantially limited’...”

Either way, as the reasoning goes, such a claimant would likely fail to make a prima facie showing of discrimination if he or she is subject to an adverse employment action. While this is a serious concern, it need not always be true. Note that it may be easier for infrequent stutterers who demonstrate that they are regarded as being disabled to show that they are qualified for the job because there is no possible catch-22 in that situation. At the same time, though, these employees would not be entitled to a reasonable accommodation.

C. What Kinds of Accommodations Are Reasonable?

It is next necessary to examine what would be considered a reasonable accommodation that an employer could make for an employee who stutters. Under its definition in the statute, a reasonable accommodation can include “job restructuring” and “reassignment to a vacant position,” but the accommodation cannot pose an “undue hardship” to the employer.

There are several possible reasonable accommodations for people who stutter. Presumably, if a person who stutters is uncomfortable because the job involves a lot of speaking (e.g., if this person is often on the telephone, needs to make presentations, or is required to meet with clients), it may be a reasonable accommodation for an employer to assign the employee to another position that involves less speaking, or perhaps change the current position to require less speaking. This may make both the employer and the employee more comfortable. Another example of a reasonable accommodation would be moving an employee’s desk to a less crowded part of the office so that it is easier for the employee to speak on the phone.

A major issue regarding accommodations is disclosure. Employees with disabilities that are not visibly apparent need to disclose these disabilities to employers to be eligible for reasonable accommodations. If an employer does not know about a disability, the employer cannot possibly make any accommodations.

---

198 Id.
199 Id.
200 42 U.S.C. § 12111(9).
201 Id. § 12111(9)(B).
202 Id. § 12111(10).
Stuttering is one such disability that may not be readily apparent to supervisors. If a person is a covert stutterer or an infrequent overt stutterer who is uncomfortable enough with stuttering that this person would like the employer to make a reasonable accommodation, it is also likely that the person is uncomfortable with the idea of disclosing the stuttering. A stutterer may be reluctant to tell an employer about his stuttering because of the myths and stereotypes that pervade the public's understanding about stuttering. Certainly, if someone is a covert stutterer, it is the fear and shame associated with stuttering that is keeping this person in the shell.

Here again is a link with mental illness. Stigma and fear of mental disorders make disclosure to employers risky, and “[p]oor self-awareness or self-denial may also make disclosure difficult.” Disclosure is “deliberate” and often “wrenching” for people with psychiatric disorders. Perhaps ironically, considering the purpose of the ADA, people with psychiatric disorders fear disclosing their impairments due to stigma and discrimination; by telling their employer about their condition, such an individual “risks discrimination, teasing or harassment, isolation, [and] stigmatizing assumptions about her ability.”

Conversely, disclosure may be a positive step for people with psychiatric disabilities, as it “may enhance self-esteem, diminish shame, permit coworkers and others to offer support, and even empower another individual’s revelation.” This assessment of the positives and negatives of disclosure by people with mental disabilities echoes the dilemma faced by many covert or mild stutterers. If these individuals choose to disclose, they may be entitled to accommodations, but they must also come out of their stuttering shells. Ultimately, many mild stutterers probably choose to suffer in silence. If they do not discuss their stuttering with their family and friends, it is probably unlikely that they would choose to do so with their

205 Laura Lee Hall, Making the ADA Work for People with Psychiatric Disabilities, in MENTAL DISORDER, WORK DISABILITY, AND THE LAW, supra note 204, at 241, 258.
206 Id. at 259.
207 Id. at 260.
208 Such a comparison is by no means precise because stuttering is not a psychiatric disability. It is relevant only inasmuch as mental disabilities and sometimes stuttering would not be readily apparent to an employer. See supra Part I.
bosses. And, if they do not choose to disclose, they should not shirk any of their duties. Failure to disclose and discuss reasonable accommodations with an employer combined with failure to fulfill the duties of the job means that an employee can be fired and left without any recourse under the ADA.

V. AN ALTERNATIVE VIEW OF STUTTERING DISCRIMINATION BASED ON IDENTITY RATHER THAN DISABILITY

Another way to proscribe discrimination on the basis of stuttering is to view stuttering itself in a different way—not as a disability, but as an identity characteristic. This part discusses the ways stuttering is different from what is usually considered a disability. It then introduces several theories that demonstrate how stuttering might not be considered a disability. This different way of looking at stuttering—as an identity trait, not as a disability—can be covered under antidiscrimination law under Title VII of the 1964 Civil Rights Act.

There are both advantages and disadvantages to coverage of stuttering under Title VII. Among the advantages is a second way to account for the emotional aspect of stuttering; if all stutterers are covered under law, the problem of the infrequent physical stutterer is abrogated. Another advantage is that explicit statutory coverage of stuttering discrimination would eliminate the need under the ADA to show that stuttering is a disability and that the individual is qualified for the job. Quite significantly, a further advantage is that such a personhood characterization of stuttering may help to empower people who stutter. Among the disadvantages are that, under Title VII, there would be no accommodations for stutterers, and the statute may be overinclusive and unrealistic.

A. Viewing Stuttering as Something Other than a Disability

While enforcement against discrimination based on stuttering can be analyzed through disability jurisprudence, there is another way to approach this type of discrimination.

209 Note that someone who stutters only infrequently would only be entitled to a reasonable accommodation if it was established that the emotional aspect of stuttering is considered a substantial limitation on speaking. See supra Part IV.A.1.
210 Parry, supra note 9.
There is something strange about viewing stuttering as a disability. It is not as though stutterers cannot speak or cannot express themselves. They are physically able to say whatever they want to say; what makes stutterers different from nonstutterers is that it takes them longer to say things. This is unlike what one usually considers a disability, when a disabled person is completely unable to do a certain activity. While it is true that stutterers cannot speak quickly, there is very rarely any need to speak so quickly. Rather, it is society that has determined that it is normal to speak without involuntary interruptions and that it is therefore inferior to speak with these interruptions. The theories of acceptance, transfluency, and covering do much to inform this discussion.

Although stutterers commonly feel ashamed of their stuttering and view it as a terrible burden, a growing number of people who stutter are growing to accept it. This self-acceptance can begin with the realization that this is how one talks and that one cannot ever fully change it. With this in mind, acceptance can offer the option of a more self-fulfilling and life-affirming mindset for people who stutter because it eliminates the “shame, guilt and embarrassment that makes speaking difficult.” Note that, under this philosophy, it is the negative feelings that make speaking hard, not the actual physical production of sounds and words.

Transfluency is one scholar’s extension of this idea of self-acceptance. Under the concept of transfluency, stuttering

---

212 See JERzer, supra note 7, at 18 (discussing how stuttering differs from other disabilities). But see Douglas C. Baynton, Bodies and Environments: The Cultural Construction of Disability, in EMPLOYMENT, DISABILITY, AND THE AMERICANS WITH DISABILITIES ACT: ISSUES IN LAW, PUBLIC POLICY, AND RESEARCH 387, 388 (Peter David Blanck ed., 2000) (discussing society’s role in constructing even such an “obvious” disability as mobility impairment by providing the example of a person who cannot walk but who nonetheless can move about freely in a wheelchair to the extent that the “built environment” allows for wheelchair use, noting that “[a]n impairment-centered definition of disability [which the ADA is modeled on] selects walking as a major life activity and rolling on wheels as an inferior substitute necessitated by the inability to engage in a normal life activity”). Still, though, stuttering is different: a person who cannot walk is able to move about in a wheelchair, but a person who stutters can speak intelligibly without any device or aid.

213 JERzer, supra note 7, at 13 (discussing the importance of time and listener reactions to stuttering).


215 See Yeoman, supra note 21, at 43 (describing the philosophy of the National Stuttering Project, the forerunner to the National Stuttering Association).

216 Stuttering Info, Frequently Asked Questions, supra note 214.

is viewed as a “distinctive feature or a manifestation of human diversity, but never as a pathological symptom.”

Transfluency considers stuttering a “manifestation of diversity in speech pattern, as being black, homosexual and left-handed are expressions of diversity in race, sexual orientation and hemispheric dominance.”

According to transfluency, stuttering is a “dramatically different speech pattern,” but it is “as human—or as natural—as the fluent one.” The conception is partially based upon the view that interruptions in stutterers’ speech are not the cause of the problem; rather, the problem is the social stigma that often accompanies those interruptions. People who stutter are often banished to the “closet,” the result of being faced with stereotypes and cruelty from society—manifested in “disapproving gestures, looks, mockery”—that results in “a personal identity associated with pain and suffering.”

Transfluency dignifies people who stutter, calling them out of the closet to express themselves and live freely. Therefore, transfluency theory holds that stuttered speech is not worse than fluent speech; it is just different. Under this conception, discrimination protection through the ADAAA

---

218 Id. at 131. Perhaps, though, such a characterization can more broadly be applied to the concept of disability, blurring this note’s line of disability and identity. See Baynton, supra note 212, at 387 (stating that “activists in the disability rights movement and scholars in the new disability studies increasingly argue that . . . the concept of disability is fraught with ambiguity and based on highly variable cultural rules and values concerning the body, personal competence, social interaction, individual responsibility, dependence and independence”).

219 Loriente, supra note 217, at 131. Employment discrimination on the basis of sexual orientation is not banned under a federal statute. See Matthew Barker, Note, Employment Law⎯Antidiscrimination—Heading Toward Federal Protection for Sexual Orientation Discrimination, 32 U. Ark. Little Rock L. Rev. 111, 129 (2009) (noting that “Congress has made repeated unsuccessful attempts to pass the Employment Non-Discrimination Act,” a bill that would proscribe employment discrimination on the basis of sexual orientation). But, numerous states and municipalities have taken this action, including both New York State and New York City. See, e.g., N.Y. Exec. Law § 296 (McKinney 2010); N.Y.C. Code § 8-107 (2011). A state or municipality could include stuttering discrimination protections in its antidiscrimination statutes before this is done on the federal level in order to judge on a smaller scale whether such a change is effective and meaningful.

220 Id. at 137.

221 Loriente, supra note 217, at 131.

222 Id. at 137.

Id. uses this term: he argues that “medicalization” of stuttering “conveys a lonely and marginalized way of living (symbolized by the metaphor of the closet).” Id. at 136. Loriente explains, “The way of life of those living in the closet is directed by lies, secrecy, and silence.” Id. at 137 n.6. While closet certainly conjures connections with sexual orientation, the concept is not out of place in the context of stuttering. Thus, this note will use closet when appropriate, while acknowledging that the term may not bring with it the exact same meaning that it has in the context of sexual orientation.

223 Id. at 137.

224 Id. at 140.
makes less sense because stuttering no longer exists as a disability; it is rather an element of an individual's identity.

Another scholar proposes a legal theory that is relevant in continuing the analysis of stuttering as an identity trait. Described as covering, “a subtler form of discrimination has risen,” where discrimination does not aim at groups as a whole but rather at the subset of the group that refuses to assimilate (i.e., cover). Kenji Yoshino identifies a judicial bias towards covering that he views as dangerous because of its perpetuation of inequality, “what reassures one group of its superiority to another.” Groups that society requests to cover are asked “to be small in the world,” to accept inequality and a “second-class citizenship.” According to Yoshino, everybody covers, there is no mainstream, and “[i]t is not normal to be completely

---


226 Id. Through the lens of sexual orientation, Yoshino describes conversion and passing as concepts that precede covering. Conversion refers to “attempts to convert homosexuals into heterosexuals.” KENJI YOSHINO, COVERING: THE HIDDEN ASSAULT ON OUR CIVIL RIGHTS 32 (2006). Passing describes gay individuals living in the closet. Id. at 69. Covering, then, refers to muting one’s identity on the axis of appearance, affiliation, activism, and association in order to gain mainstream acceptance. Id. at 79-80 (referring to gays “acting straight” as an example of covering). While this is not equivalent to stuttering, it is possible to analogize. Conversion would refer to the still-widespread attempts to speak fluently through speech therapy. Passing, therefore, would describe the attempts to hide oneself as a stutterer—using tricks to try not to stutter, the most extreme form of which is to become covert. See supra Part I. And, covering, then, would refer to the way people who are open about their stuttering still try not to stutter in certain situations where they believe it is less acceptable, or how they continue to downplay the large role that stuttering plays in their lives. Covering would also refer to the people who stutter so frequently that they try not to talk; people know they stutter, but they try to be more “normal” by speaking less. Interestingly, stuttering (as a condition in society) seems to be simultaneously going through each of these three phases.

227 Yoshino, supra note 225. Theoretically, the accommodation model of disability-discrimination law should protect disabled individuals from needing to cover. YOSHINO, supra note 226, at 173. But courts have limited this accommodation principle, instead continuing to prefer assimilation. Id. at 174-76. Interestingly, Yoshino explains that courts have done this by interpreting the definition of disability strictly, which the ADAAA is designed to change. Id. at 175.

228 Yoshino, supra note 225. Yoshino provides some examples of disabled individuals covering—a visually impaired person who dresses well, does not use a cane, and memorizes what she must read aloud, as well as people with mobility impairments who “use able-bodied people as ‘fronts’” to travel with. YOSHINO, supra note 226, at 172. Remember, though, that individuals with sight and motion impairments would have difficulty covering because their impairment would tend to be obvious. This is not the case for many stutterers, who can choose to not speak or be covert. Stuttering, then, remains more in the passing phase, which is why explicitly enshrining stuttering antidiscrimination provisions in a statute would likely help stutterers leave their closets. See infra Part V.B. And note again how stuttering is different from other disabilities, as it is easier to pass as a nonstutterer than to pass as lacking many other physical impairments.
normal.” Yoshino believes that the “free[dom] to develop our human capacities without the impediment of witless conformity . . . extends beyond traditional civil rights groups.” While stutterers do not constitute a traditional civil rights group, Yoshino believes that this freedom of individual personhood should extend beyond the traditional groups to confront coerced conformity everywhere, blurring the view of what it means to be normal and in the mainstream.

B. Coverage for Stuttering in the Civil Rights Act

The alternative view of stuttering would lead to coverage of stuttering discrimination under Title VII of the Civil Rights Act of 1964. This statute prohibits employers’ discrimination on the basis of race, color, religion, sex, or national origin. These categories are all traits or aspects of personhood with which people identify but have historically been imputed with inferiority by society. The statute would need to be amended to cover stuttering: this would not work under the existing framework. While a statute cannot directly change a person’s outlook, explicit coverage of stuttering in Title VII would not only affect the actions of employers, but it may also empower people who stutter to discover more of their true potential.

There are several advantages to covering stuttering under the Civil Rights Act rather than under the ADA. Were stuttering included in this category, the processes under the ADAAA would be dispensed with (i.e., whether stuttering is substantially limiting, whether a stutterer was regarded as having an impairment, and whether a stutterer is qualified for the job). This recognition would eliminate the catch-22 problem.

---

229 Yoshino, supra note 225.
230 Id.
231 Id.
233 Alternatively, an entirely new statute could be created, like the Age Discrimination in Employment Act, which was enacted for age discrimination. The goal is not Title VII coverage precisely, but rather a statutory scheme where stuttering need not undergo the preliminary determination of whether it is substantially limiting for it to be covered (as it does in the disability context).
234 See Yeoman, supra note 21, at 47 (describing one stutterer’s outlook that stutterers need to hear other people stutter so that they can have role models). An organization called Our Time Theatre Company strives to give young people who stutter a safe space to express themselves and pursue their artistic abilities. Taro Alexander, Our Time Provides Kids Their Time, 1 J. STUTTERING, ADVOC. & RES. 33, 33-34 (2006) (describing how “an environment of unconditional acceptance transforms . . . fear and shame into confidence and self-esteem”). At Our Time, stuttering on stage during live shows is allowed and encouraged. Id. at 35.
of ADAAA enforcement. Plaintiffs would not need to worry that they would be considered substantially limited in speaking and thus not qualified\textsuperscript{235} for the job. Conversely, they would not need to be concerned that their qualifications would prevent them from being considered substantially limited. By changing the way stuttering discrimination is conceived, employees would be able to stutter as much or as little as they needed to—or wanted to—fully aware that they are protected against discrimination. And, furthermore, covering all stutterers would solve the problem of taking into consideration stutterers who are more affected by the emotional aspect of stuttering than the physical aspect. In this way, the enforcement of nondiscrimination in the workplace would be more easily accomplished for employees who stutter.\textsuperscript{236}

Another reason why the identity model of stuttering antidiscrimination enforcement would be positive is that it would encourage stutterers to come out of their closets.\textsuperscript{237} Congress’s automatic recognition of stuttering in a statute as a trait protected against discrimination would more easily allow for people who stutter to choose not to hide their speech. To the extent that society wants to encourage openness, the personhood model would likely go far in encouraging (and perhaps accomplishing) it. In fact, research indicates that stutterers face problems in the workplace not only because of discrimination from supervisors, but also because of their own attitudes about their stuttering.\textsuperscript{238} One study indicates that some stutterers did not choose the career they wanted because of stuttering and avoided jobs that required use of the telephone or making oral presentations.\textsuperscript{239} Another study shows that 50 percent of stutterers looked for jobs requiring little speaking and 21 percent have declined a new job or promotion because of fears associated with stuttering.\textsuperscript{240} Some people who stutter feel trapped in an unwanted job because of their stuttering.\textsuperscript{241}

\textsuperscript{235} Qualified in this context means that an employer could not say that a job requirement of excellent oral communication skills requires nonstuttered speech. To be qualified, the stutterer would, of course, still need to otherwise satisfy that requirement and meet all other job requirements.

\textsuperscript{236} The relative complexity of the ADAAA model that this note proposes would be disposed of. See supra Part IV.

\textsuperscript{237} See supra note 222.

\textsuperscript{238} Klein & Hood, supra note 41, at 256.

\textsuperscript{239} Id. at 257 (citing R. Hayhow et al., \textit{Stammering and Therapy Views of People Who Stammer}, 27 J. FLUENCY DISORDERS 1 (2002)).

\textsuperscript{240} Id. at 266.

\textsuperscript{241} Id. at 267.
indicates the extent to which stutterers limit themselves in regards to their own employment opportunities. The changing of attitudes like these will ultimately be necessary in order for stutterers to reach their full employment capabilities.

There are also drawbacks to this identity model of stuttering. Because stuttering is a variable trait, the ADAAA may work better; under the substantially limits conception, courts would understandably find people who stutter severely to be disabled, and people who stutter mildly not to be disabled.\(^\text{242}\) This personhood theory would be overinclusive in that it would cover stutterers who stutter rarely and at the same time are not emotionally affected by stuttering.\(^\text{243}\) Also, perhaps some stutterers have such severe impediments or conditions that they are not realistically employable; companies may have business reasons not to hire a severe stutterer. Under the ADAAA, such an individual would not be covered because he or she would not be qualified for the job. Under Title VII, coverage would depend upon the way the statutory amendment is laid out;\(^\text{244}\) it’s more likely here, though, than under the ADAAA, that this person would be covered, as the identity theory rests on a modicum of acceptance of stuttering in the society at large that is probably not yet present in the culture. But Title VII does not cover everything.\(^\text{245}\) Accordingly, a public-safety exception for stuttering may be advisable; stutterers may be ill-suited for certain jobs that depend upon rate of speed in talking (e.g., air-traffic controller).\(^\text{246}\) Furthermore, under Title VII, stutterers

---

\(^{242}\) This individualized inquiry would make sense if the level of severity took into account both the physical and emotional aspects of stuttering. But if courts only looked at frequency of stutters, rather than at how much stuttering affects a person holistically, employees would too frequently be mischaracterized. See supra Part I.

\(^{243}\) Some kind of medical diagnosis of stuttering would probably be a wise requirement here, so that individuals who do not stutter do not take advantage of this new statutory provision.

\(^{244}\) The McDonnell-Douglas test would still apply. McDonnell Douglas Corp. v. Green, 411 U.S. 792 (1973); RUTHERGLEN, supra note 53, at 37-42.

\(^{245}\) Title VII permits classifications on the basis of “religion, sex, or national origin in those certain instances where religion, sex, or national origin is a bona fide occupational qualification reasonably necessary to the normal operation of that particular business or enterprise.” RUTHERGLEN, supra note 53, at 129 (quoting 42 U.S.C. § 2000e-2(e)(1) (2006)). This provision, known as the bona fide occupational qualification (BFOQ), does not include race. Id. The BFOQ, though, has been interpreted very narrowly. Id. at 129, 143.

\(^{246}\) And, if the statute were written in a less progressive way, there could be other exceptions besides a safety exception. For example, perhaps employers of broadcasters would still be able to discriminate, if the public would not accept the transfluency notion of people stuttering across the airwaves. For an example of stuttering being accepted on the stage, see supra note 234.
would likely not be entitled to reasonable accommodation.\textsuperscript{247} While this makes sense in the context of the personhood model,\textsuperscript{248} it would be impractical for a stutterer for whom an accommodation would be very helpful. This lack of entitlement to accommodation is a significant drawback because it is likely that most stutterers would not suddenly come to believe in acceptance and transfluency. In other words, many stutterers at this point in time would likely want access to accommodations because they will not be comfortable with their stuttering.

Ultimately, while the identity model would be the more effective way to counter discrimination, the disability model is the more practical method of dealing with this problem at this point in time. Stuttering discrimination conceived as a disability can be covered under the existing ADA, while protection against stuttering discrimination as an identity characteristic would require Congressional action.\textsuperscript{249}

\section*{Conclusion}

Employment discrimination on the basis of stuttering is an important issue to examine because it is widespread and limiting for stutterers, and because there should be legal protection for people who stutter. Coverage for stuttering should include both the physical and emotional aspects of the condition. One way for stuttering to be covered is as a disability under the ADA. The recent amendments to the ADA increase the likelihood that it would be covered. Another way for stuttering to be covered is as an identity characteristic under the Civil Rights Act. Both methods of coverage are compelling. While the Civil

\textsuperscript{247} This assumes that a Title VII amendment for stuttering would track race discrimination, as there is no right to accommodation in that area. There is, though, a certain accommodation right in Title VII regarding religious discrimination, so it’s possible that a Title VII amendment for stuttering could include such an accommodation. \textit{See Rutherford}, \textit{supra} note 53, at 146-47.

\textsuperscript{248} No accommodation would be necessary if stuttered speech were just as valid as nonstuttered speech—there would be nothing for which to accommodate.

\textsuperscript{249} Is there a third way to conceive of stuttering? Ruth Colker describes the term “hybrid” as “people who lie between bipolar legal categories—bisexuals, transsexuals, multiracial, and the somewhat disabled.” \textsc{Ruth Colker, Hybrid: Bisexuals, Multiracial, and Other Misfits Under American Law} xi (1996). Regarding disability, she reports that the phrase “temporarily able-bodied” has been used to describe the “transient nature of . . . disability status.” \textit{Id.} at xiii. While Colker focuses on categories like body size, perhaps this can describe stuttering. \textit{Id.} at 165. The variable nature of stuttering can cause it to greatly affect a person’s life at times, yet not be any issue at other times. Sometimes this shift can occur over a period of months or years, and other times it can occur within hours or minutes.
Rights Act would provide broader protection, stuttering will more realistically be covered under the amended ADA because it does not require any further congressional action. In whichever way, discrimination based on stuttering should be proscribed so that people who stutter can more easily enjoy the full measure of their civil rights.

*Jared Gilman*†

† J.D. Candidate, Brooklyn Law School, 2012; B.A., Binghamton University, State University of New York, 2008. I want to recognize the crucial role that the National Stuttering Association plays in providing support for the stuttering community, and to acknowledge the organization’s significant effect on my own life, from which this note topic developed. I wish to thank the editors and staff of the *Brooklyn Law Review* for their passion and commitment. I would also like to thank Professor Nelson Tebbe for his useful guidance and suggestions, and Eric Jackson for contributing illuminating source material. Finally, I am eternally grateful for the love, support, and encouragement of my parents and Katherine McCrink.
A Standardless Standard

HOW A MISAPPLICATION OF KELO ENABLED COLUMBIA UNIVERSITY TO BENEFIT FROM EMINENT DOMAIN ABUSE

INTRODUCTION

In 2010, the New York Court of Appeals in *Kaur v. New York State Urban Development Corp.* granted Empire State Development Corporation (ESDC), a public authority, permission to take land in Manhattanville by eminent domain for sale to Columbia University, a private institution. The taking indirectly displaced thousands of vulnerable residents and failed to create meaningful public benefits. Though ESDC justified the taking as a means to eliminate urban blight, substantial evidence strongly indicated that its primary motivation was Columbia’s private benefit. By deferring to ESDC’s findings, the court misapplied important judicial principles and failed to prevent an unconstitutional exercise of eminent domain.

The government’s power to take property for public use is both created and limited by the Public Use Clause of the Fifth Amendment, which dictates that “private property [shall not] be taken for public use, without just compensation.” Since the founding of the United States, courts have interpreted the requirement that property only be taken to serve public purposes as a necessary restriction on the power of legislatures to seize land. By restricting the government’s authority to take private property only for public purposes, the clause imposes a safeguard against governmental favoritism and the abusive

---

3 *Kaur*, 933 N.E.2d at 724.
4 See infra Part V.
5 *U.S. CONST.* amend. V (emphasis added).
dispossession of property owners.7 Because takings often disproportionately harm vulnerable populations—such as the elderly and ethnic minorities—courts must ensure that such power is in fact used only to further the public good.8 Failing to do so in Kaur, the court approved an unconstitutional taking9 with devastating effects on many Manhattanville residents.10

The Supreme Court established the standard for public-use review in Kelo v. City of New London. In Kelo, the Court upheld a taking where the City intended to transfer property to a private developer with the public purpose of encouraging economic development.11 In so holding, the Court explicitly established the government’s ability to take private property and subsequently convey such property to another private party so long as a predominantly public purpose is served.12 Nevertheless, the Kelo Court maintained the judiciary’s traditional responsibility to review whether a taking is actually intended to serve a public purpose rather than solely to provide a private benefit.13 The Court suggested the existence of a taking where the evidence of hidden impermissible favoritism is so substantial as to warrant a presumption of constitutional invalidity.14 Such a case requires heightened judicial scrutiny into whether the taking is actually intended to accomplish a public purpose rather than the traditional deference applied in Kelo.15

This note argues that where affected landowners present sufficient evidence that the purported public purpose of a taking is merely pretextual to bestowing a private advantage, courts must consider all such evidence and deny absolute deference to the condemning authority.16 This way, courts can prevent governmental agencies from abusing their power of eminent domain to transfer property from vulnerable populations to private parties who enjoy governmental favoritism.17

---

8 See infra Part I.
9 See infra Part V.
10 See infra Part I.
12 See id.
13 Id. at 478.
14 See infra Part II.
15 See infra Part II.
16 See infra Part II.
17 Kelo, 545 U.S. at 522 (2005) (Thomas, J., dissenting) (citing id. at 505 (O’Connor, J., dissenting)).
Since *Kelo*, courts have disagreed over whether the requirement of serious judicial inquiry into substantiated allegations of pretext extends to takings justified by blight remediation. While some courts correctly extend the *Kelo* analysis to any takings challenged with substantial evidence of impermissible favoritism, the Second Circuit explicitly rejected the application of heightened scrutiny in *Goldstein v. Pataki*, where a taking was intended to remediate blight because the court found that *Kelo*'s pretext analysis only applied to economic development takings. However, though findings of blight were traditionally limited to unsafe and unsanitary conditions, the modern definition of blight removal applied in New York is so broad that it encompasses the spirit of economic development addressed by the Court in *Kelo*. Therefore, the Second Circuit's absolute deference to the condemning authority's findings of blight in *Goldstein* failed to properly apply *Kelo* and invited future abuses of the eminent domain power.

This failing led the New York Court of Appeals in *Kaur* to review the exercise of eminent domain with absolute deference where ESDC claimed its actions were intended to eliminate blight. Substantial evidence indicated the taking was primarily intended to benefit Columbia: Columbia created blighting factors, ESDC assisted in manufacturing a blight study at Columbia's behest, and ESDC sought to withhold important documents from the challengers during litigation, clearly indicating a conspiratorial relationship between ESDC and Columbia. Furthermore, the procedural protections for property owners seeking to bring public use challenges in New York are prone to abuse because the statutes governing eminent domain procedure do not allow trial-level review of such claims. Nevertheless, the New York Court of Appeals upheld the taking.

---

23 See infra Part III.
24 *Franco v. Nat'l Capital Revitalization Corp.*, 930 A.2d 160, 170-71 (D.C. 2007) (finding impermissible favoritism where an agency gave taken land to a private party for the purpose of developing a purportedly blighted area); *In re Condemnation Proceeding by the Redevelopment Auth. of Phila.*, 938 A.2d 341, 345 (Pa. 2007) (citing Kennedy's concurrence for the requirement to seriously review the record for impermissible favoritism).
25 *Goldstein v. Pataki*, 516 F.3d 50, 64 n.10 (2d Cir. 2008).
27 See infra Part IV.
28 See infra Part III.
30 See infra Part IV.
31 See N.Y. EM. DOM. LAW § 207 (Consol. 2011).
without even mentioning the heightened standard required by *Kelo*. By failing to apply heightened scrutiny, the court misapplied *Kelo* and enabled ESDC to abuse the power of eminent domain. Thus, the Supreme Court should have reversed the New York Court of Appeals and remanded *Kaur* for review using the heightened scrutiny required by *Kelo*. In denying the challengers' petition for certiorari, the Court failed to defend the vulnerable populations who will be harmed by eminent domain abuse in Manhattanville and also missed an opportunity to clarify its misunderstood holding in *Kelo*.

This note will argue that the New York Court of Appeals applied an overly deferential standard of review to the taking at issue in *Kaur* and, in doing so, disobeyed the constitutional requirements of the Fifth Amendment under *Kelo* to the detriment of Manhattanville's economically disadvantaged citizens. Part I will describe the harms imposed on vulnerable populations when courts permit eminent domain abuse. Part II will explain that *Kelo* requires a heightened standard of judicial review where challengers to a taking must present substantial evidence that a condemning authority's stated public purpose is mere pretext for bestowing a private benefit. Part III will discuss the divergent standards of review applied by courts to public use challenges where takings are not solely justified by the need for economic development. Part IV will argue that courts in New York should apply heightened scrutiny to takings for blight remediation where challengers allege an unconstitutional private purpose because the factors in New York for determining whether an area is blighted and those for determining the need for economic development are indistinguishable from one another. Part V will argue that the New York Court of Appeals failed to prevent eminent domain abuse in *Kaur*. This part will argue that the court should have applied heightened scrutiny to ESDC's motives because substantial factual evidence supported a finding of impermissible favoritism and because the procedures for challenging a taking under the Public Use Clause in New York are particularly prone to abuse. It will further suggest that, due to the failures of the New York judiciary to prevent eminent domain abuse, the state legislature should take action

27 See *Kaur*, 933 N.E.2d at 737.
28 *Tuck-It-Away, Inc.*, 131 S. Ct. at 822-23.
29 See infra Part I.
30 See infra Part III.
to protect vulnerable parties from private takings with merely pretextual public benefits. Finally, this note concludes by suggesting further discussion.

I. THE NEED FOR HEIGHTENED SCRUTINY OF SUSPICIOUS TAKINGS TO PREVENT HARMFUL ABUSE

The government’s power to take private property has the dangerous potential to harm vulnerable populations and thus must be restricted to its constitutional limitations. The Public Use Clause forbids the exercise of eminent domain for purely private transfers, which could otherwise unconstitutionally harm property owners who do not enjoy governmental favoritism.31 Without proper restraints on legislative power to take and transfer property, “[t]he specter of condemnation hangs over all property. Nothing is to prevent the State from replacing any Motel 6 with a Ritz-Carlton, any home with a shopping mall, or any farm with a factory.”32 Thus, eminent domain abuse results in legislatures favoring rich and powerful citizens over those with less means to promote their economic and political interests,33 just as the residents of Harlem were harmed in favor of Columbia University.

Takings for economic development “disproportionately harm racial and ethnic minorities, the elderly, and the economically underprivileged.”34 This proposition, championed by Justice Thomas’s dissent in Kelo,35 is supported by substantial anecdotal evidence.36 As of 2007, eminent domain project areas nationally were composed of, on average, 58 percent of minority residents while the surrounding

31 Kelo v. City of New London, 545 U.S. 469, 522 (2005) (Thomas, J., dissenting) (quoting id. at 505 (O’Connor, J., dissenting)).
32 Id. at 503 (O’Connor, J., dissenting).
33 See id. at 505.
35 See Kelo, 545 U.S. at 522 (Thomas, J., dissenting).
36 See NAACP Amicus Brief, supra note 34, at 10 (“In San Jose, California, ninety-five percent of the properties targeted for economic redevelopment are Hispanic or Asian-owned, even though only thirty percent of businesses are owned by minorities . . . . In Ventnor, New Jersey, forty percent of the city’s Latino community lives in a zone targeted for economic redevelopment . . . . In Mt. Holly Township, New Jersey, officials have targeted for economic redevelopment a neighborhood in which the percentage of African-American residents (44%) is twice that of the entire Township and nearly triple that of Burlington County, and in which the percentage of Hispanic residents (22%) is more than double that of all of Mt. Holly Township, and more than five times that of the county.”).
communities contained an average of only 45 percent of minorities. Similarly, the median incomes of persons living within eminent domain project areas was $18,935.71, while the median income of persons in the surrounding communities was $23,113.46. Generally, properties are often selected for eminent domain partially due to their low market values, which dictates the amount of compensation the government is required to pay upon condemnation. Thus, displaced citizens typically face difficulties finding “adequate replacement housing.” This is particularly burdensome on the elderly, many of whom do not own their homes and are more likely to spend the end of their lives in nursing homes if displaced. Clearly, those harmed by takings for economic development and blight remediation are groups with relatively little political and economic power who are in the greatest need of protection by the courts.

The exercise of eminent domain in Manhattanville will primarily harm economically disadvantaged residents. In an amicus brief in support of the challengers in Kaur, New York State Senator Bill Perkins urged against the taking because the proposed development would indirectly displace between three thousand and five thousand Harlem residents. In particular, as the affected area was composed of 29.4 percent African-Americans and 52.3 percent Latinos, the taking would disproportionately burden minority citizens. Like the takings before it, the use of eminent domain in Manhattanville will invariably impose hardships on economically disadvantaged and politically impotent residents while further enriching a wealthy and powerful private actor, here Columbia University.

Standing up to legislative abuse on behalf of powerless citizens is an essential function of the courts and should be embraced under the Public Use Clause. Justice Stone’s famous Footnote Four in United States v. Carolene Products Co.

38 Id.
39 NAACP Amicus Brief, supra note 34, at 13.
40 Id.
41 Id. at 14.
43 Perkins Amicus Brief, supra note 2, at 1.
44 Id. at 19.
45 Id. at 19-20.
46 CARPENTER II & ROSS, supra note 37, at 6.
47 See Kelo, 545 U.S. at 521-22 (Thomas, J., dissenting).
established the need for heightened judicial scrutiny of legislative decisions when they harm “discrete and insular minorities.” Eminent domain abuse tends to cause such harm because it “eliminates (or severely undermines) established community support mechanisms and has a deleterious effect on those groups’ ability to exercise what little political power they may have established as a community.” Thus, takings that benefit private parties should not receive deferential treatment when challenged under the Public Use Clause because they “curtail the operations of those political processes ordinarily to be relied upon to protect minorities . . . .” For these reasons, courts should apply *Kelo*’s test for impermissible favoritism to takings where evidence suggests that a stated public purpose is mere pretext for the unconstitutional transfer of property to a private party. The New York Court of Appeals failed to do so in *Kaur* and thus unconstitutionally harmed the vulnerable residents of Manhattanville.

II. *Kelo v. City of New London*: Setting the Standard of Review

The Supreme Court established the standard for reviewing challenges to eminent domain takings under the Public Use Clause in *Kelo*.

Where substantial evidence shows a taking “is intended to favor a particular private party, with only incidental or pretextual public benefits,” the court must not defer to the condemning authority. Rather, the court must review the evidence to determine if the taking will result in actual benefits to the public. If such review reveals impermissible favoritism rather than a valid public use, then the taking is unconstitutional. In *Kaur*, the court misapplied this standard and, in doing so, ratified a flagrant violation of the Public Use Clause.

In *Kelo*, the New London Development Corporation (NLDC), a nonprofit entity authorized to assist the City of New

---

49 NAACP Amicus Brief, *supra* note 34, at 15.
50 *Carolene Products*, 304 U.S. at 152 n.4.
51 See infra Part II.
52 See infra Part V.
54 See id. at 491-93 (Kennedy, J., concurring).
55 See id.
56 See id. at 491.
57 See infra Part V.
London, sought to take private land for the purpose of promoting economic development. NLDC planned to give the land to the pharmaceutical giant Pfizer Inc. so they could build a $300 million research facility. The City projected the project would create over one thousand jobs, increase tax revenues, and generally improve the area’s economy. Susette Kelo, a homeowner whose property was at risk from the proposed project, filed a suit against the City for violating the Public Use Clause. Justice Stevens, writing for the five-justice majority, rejected the petitioner’s claim and held that private transfers of land taken through eminent domain are constitutional under the Public Use Clause where the purpose is to promote economic development.

Though the Court gave deference to the legislature’s determination that the public purpose of economic development would be served by the taking, Justice Stevens noted that the government may not “take property under the mere pretext of a public purpose, when its actual purpose was to bestow a private benefit.” Such was not the case in Kelo, Stevens noted, because the taking was “executed pursuant to a ‘carefully considered’ development plan.” Thus, the Court left the door open for a hypothetical taking that might be found unconstitutional due to a stated public purpose that is “mere pretext” for enriching a private party. However, in applying Kelo, subsequent courts have experienced difficulty applying this restriction because “the Kelo majority did not define the term ‘mere pretext.’”

Justice Kennedy, in a concurrence that qualified his agreement with the majority in Kelo, took steps to outline a hypothetical taking that would be unconstitutional under the Public Use Clause for “mere pretext.” He defined as unconstitutional “transfers intended to confer benefits on particular, favored private entities, and with only incidental or

---

58 Kelo, 545 U.S. at 473.
59 Id.
61 Kelo, 545 U.S. at 475.
62 Id. at 489-90.
63 Id. at 483.
64 Id. at 478 (emphasis added).
65 Id.
66 Id. at 486-87.
68 See Kelo, 545 U.S. at 490-93 (Kennedy, J., concurring).
pretexutal public benefits.” Though he agreed with Justice Stevens that generally courts should afford legislatures deference in their determinations to take land for public purposes, he noted that “[a] court confronted with a plausible accusation of impermissible favoritism to private parties should treat the objection as a serious one and review the record to see if it has merit.” While the majority opinion approached the stated public purpose with a “presumption of validity,” Justice Kennedy put forth a hypothetical where the risk of hidden impermissible favoritism in a transfer of taken land to private parties is so severe that courts should instead apply a presumption of constitutional invalidity. Such a case would exist where “the transfers are . . . suspicious, or the procedures employed . . . [are] prone to abuse, or the purported benefits [to the public] are . . . trivial or implausible . . . .” Thus, Kennedy’s concurrence sets forth guidelines by which a successful claim of pretext for impermissible favoritism can be brought to prevent a taking that benefits a private party.

In joining the majority to uphold the taking in Kelo, Justice Kennedy identified several reasons why the taking at issue did not exhibit signs of impermissible favoritism. Among these factors were the City’s formulation of a development plan and commitment of public funds for the project before the private beneficiaries were identified. In addition, the City’s compliance with “elaborate procedural requirements that facilitate review of the record and inquiry into the city’s purposes” led Kennedy to join the majority in upholding the taking as constitutional. Notably, such procedural safeguards included a seven-day trial before the superior court with regard to the public use challenge. As indicated in his concurrence, the absence of such factors would have prompted Justice Kennedy to apply a heightened level of scrutiny, which could have resulted in a finding that the taking was an unconstitutional violation of the Public Use Clause.

---

69 Id. at 490.
70 Id. at 491.
71 Id. at 493.
72 Id.
73 Id.
74 Id. at 491-93.
75 Id.
76 Id. at 493.
77 Id. at 475 (majority opinion).
78 Id. at 475 (Kennedy, J., concurring).
Justice Kennedy’s concurrence is essential to clarifying the *Kelo* decision with regard to “mere pretext” because, as the deciding vote on a split court, Kennedy explicitly conditioned his agreement with the majority on specific actions taken by the City’s planning process to convince him of the taking’s constitutionality. Though the concurrence expresses the opinion of only a single justice, it makes clear, in conjunction with the majority’s warning about pretextual takings, that courts must not apply an absolutely deferential standard when property owners raise legitimate and well-founded public use challenges. Rather, “deference to the government’s public purpose determination may be overcome . . . if the party challenging the taking makes a ‘clear showing’ that the government’s stated public purpose is ‘irrational,’ with ‘only incidental or pretextual public benefits.” Thus, courts should give close review to takings that clearly exhibit the possibility of unconstitutional transfers as described in Kennedy’s concurring opinion. As such, *Kelo* establishes a “federal baseline” under which courts must apply a heightened standard of review to challenges under the Public Use Clause that are supported by substantial evidence of impermissible favoritism. Takings that benefit private parties while only creating incidental public benefits are unconstitutional under the Public Use Clause. Thus, where substantial evidence indicates that eminent domain is exercised with impermissible favoritism, courts must apply a heightened standard of review to prevent violations of the Fifth Amendment. By failing to apply this scrutiny in *Kaur*, the New York Court of Appeals did not meet its constitutional obligation to prevent a taking justified by merely pretextual benefits to the public.

---

80 *Kelo*, 545 U.S. at 478 (majority opinion).
82 *Id.* (citations omitted).
83 *See Tuck-It-Away Cert. Petition, supra note 79, at 20.
84 *Cormack v. Settle-Beshers*, 474 F.3d 528, 531 (8th Cir. 2007).
85 W. Seafood Co. v. United States, 202 F. App’x 670, 675 (5th Cir. 2006).
87 *Id.* at 490-93.
88 *Id.*
89 *See infra* Part V.
III. MISAPPLYING Kelo—A SPLIT IN THE COURTS

While courts uniformly apply Kelo to takings for which the stated purpose is economic development, jurisdictions differ as to whether the heightened standard for reviewing pretext claims should be used for other categories of takings, such as those purportedly executed to eliminate blight. Those courts that afford a deferential presumption of validity to takings where substantial evidence suggests the stated public purpose of blight remediation is mere pretext in order to bestow a private benefit operate against the dictates of Kelo. The New York Court of Appeals suffered from this failing when it improperly upheld the taking in Kaur.

In cases where the legitimacy of a taking for economic development is challenged under the Public Use Clause due to evidence of favoritism, courts uniformly apply Kelo’s standard for reviewing pretext claims. For example, in Western Seafood Co. v. United States, the Fifth Circuit considered a public use challenge to a taking that sought to promote economic development through transfer of Western Seafood’s waterfront property to Hiram Walker Royall, a private developer. There, the court considered factual evidence of impermissible favoritism to determine whether the stated public purpose was merely a pretext for conferring a private benefit. This evidence supported allegations that the private developer had himself proposed the development project and that the City had granted him complete operational control over the project. However, the court upheld the taking as constitutional because the evidence “[did] not support the inference that the City exhibited favoritism or [had] a purpose other than to promote economic development.” In reaching this conclusion, the court considered all of the evidence presented by challengers to determine
whether a heightened standard of review was necessary and thus correctly applied the analysis required by \textit{Kelo}.

On the other hand, a conflict exists between jurisdictions as to whether the heightened standard of review envisioned by \textit{Kelo} for suspicious takings applies to cases where the stated public purpose is blight remediation rather than economic development.\footnote{\textit{Id.}; see also \textit{City of Norwood v. Horney}, 853 N.E.2d 1115, 1137 (Ohio 2006) (applying Kennedy’s concurrence in \textit{Kelo} to a taking for economic development purposes to support the proposition that the court should “strike down a taking that, by a clear showing, is intended to favor a particular private party, with only incidental or pretextual benefits”).} While the Second Circuit has expressly limited heightened scrutiny to economic development takings,\footnote{\textit{Compare} \textit{Goldstein v. Pataki}, 516 F.3d 50, 64 (2d Cir. 2008), \textit{with} \textit{Franco v. Nat’l Capital Revitalization Corp.}, 930 A.2d 160, 171-72 (D.C. 2007).} the D.C. Circuit Court of Appeals has correctly applied \textit{Kelo}’s pretext analysis to a taking for blight remediation.\footnote{\textit{Goldstein}, 516 F.3d at 64.}

The Second Circuit failed to apply \textit{Kelo}’s heightened standard of review to a blight remediation taking in \textit{Goldstein}.\footnote{\textit{Franco}, 930 A.2d at 171-72.} There, ESDC took petitioner’s property with the intention of transferring it to Forest City Ratner Company (FCRC), a private developer, to develop a new sports stadium for the New Jersey Nets in the Atlantic Yards Project Area.\footnote{\textit{Goldstein}, 516 F.3d at 64.} A blight study commissioned by ESDC found that the neighborhood was characterized by “unsanitary and substandard conditions,” such as vacant and underutilized buildings, irregularly shaped lots, and a long-abandoned and deteriorating rail line.\footnote{\textit{Id.} at 59.} The challenging property owners, however, claimed the finding of blight was merely pretext for the private benefit to FCRC.\footnote{\textit{Id.} at 52-53.} The allegations of impermissible favoritism were supported by evidence that the private developer first conceived of the project and proposed the geographic boundaries thereof, that the blight study occurred after the project had been announced, and that the required public review was a “sham.”\footnote{\textit{Id.} at 55-56.} Regardless, the court rejected the public use challenge on its face, finding that the \textit{Kelo} pretext analysis did not apply because “private economic development is neither the sole, nor the primary asserted justification” for
The court also refused to apply Justice Kennedy’s concurrence, finding that, “Kennedy may well have intended [his opinion] to apply exclusively to cases where the sole ground asserted for the taking was economic development.”

Conversely, the D.C. Circuit Court of Appeals correctly applied _Kelo_ in a public use challenge to a taking for blight remediation. In _Franco v. National Capital Revitalization Corp._, the court reversed the trial court’s decision to dismiss a claim for impermissible favoritism where a municipal agency took and transferred land to a private party for the purpose of developing a purportedly blighted area. There, the city alleged that the challenger’s shopping center was a blighting factor and thus subject to taking by eminent domain. The city based its finding on evidence that the area was “characterized by underused, neglected, and poorly maintained properties,” and that fragmented ownership encouraged an increase in crime, trash, and “other blighting factors.” Further, the city claimed development by the private recipient would result in crime reduction, increased sanitation, local job creation, expansion of the tax base for the city, and the general “revitalization of an economically distressed community.” On the other hand, the landowner made “specific factual allegations” in support of his pretext claim, including allegations that the municipal agency had entered into an agreement with the private developer two years before the development program was introduced to the city council, that the agency had “refused to discuss redevelopment plans with any present owners,” and that the site was not actually blighted.

The D.C. Circuit remanded the case for the court below to try the pretext claim on its merits, finding that “_Kelo_ makes clear that there is room for a landowner to claim that the legislature’s declaration of a public purpose is a pretext

---

108 _Id._ at 64.
109 _Id._ at 64 n.10. Furthermore, as Part IV of this note will explain, the criteria necessary in New York to take property for either blight remediation or economic development are indistinguishable from one another. See _infra_ Part IV. Thus, the distinction is an unsound basis upon which to alter the application of the Public Use Clause. See _infra_ Part IV.
111 _Id._ at 162-63.
112 _Id._ at 163.
113 _Id._
114 _Id._
115 _Id._ at 170-71.
designed to make a taking for private purposes.”116 Unlike the Second Circuit in Goldstein,117 other courts have correctly applied the heightened scrutiny required by Kelo.118

Due to the devastating consequences of eminent domain abuse to vulnerable communities,119 courts should not apply a uniform standard of absolute deference when deciding whether the stated purpose of blight remediation is merely pretextual to bestowing a private benefit.120 The Second Circuit failed to apply heightened scrutiny in Goldstein and thus violated the Kelo standard.121 Similarly, the New York Court of Appeals failed to even mention Kelo when it upheld a taking justified by blight remediation despite well-founded allegations of impermissible favoritism in Kaur.122 Thus, the court deferred to ESDC’s finding of blight and, in doing so, upheld a harmful and unconstitutional taking.

IV. BLURRING THE LINE: THE EQUIVALENCE OF ECONOMIC DEVELOPMENT AND BLIGHT REMEDIATION

To justify a taking under the Public Use Clause, the factors for finding blight in New York are indistinguishable from those used to determine the need for economic development.123 Thus, the Second Circuit’s theory that takings for blight remediation require less judicial scrutiny than those justified by economic development was illogical.124 The New York Court of Appeals presumably followed Goldstein as precedent in Kaur when it improperly deferred to ESDC’s finding of blight in Manhattanville despite substantial evidence of impermissible favoritism.125 Therefore, the deferential standard applied by the court failed to meet the constitutional

116 Id. at 171-72.
117 Goldstein v. Pataki, 516 F.3d 50, 64 n.10 (2d Cir. 2008).
118 See Hawaii v. C & J Coupe Ltd. P’ship, 198 P.3d 615, 650 (Haw. 2008) (directing the trial court in a public use challenge of a taking allegedly accomplished to improve a public highway to “consider any and all evidence . . . indicating that the private benefit . . . predominated”); In re Condemnation Proceeding by the Redevelopment Auth. of Phila., 938 A.2d 341, 345 (Pa. 2007) (considering a pretext claim pursuant to Justice Kennedy’s concurrence in Kelo against a taking for blight remediation, and finding that the record does not support a bad faith claim).
119 See supra Part I.
120 See supra Part II.
121 See supra Part III.
122 See infra Part V.
124 Goldstein v. Pataki, 516 F.3d 50, 64 (2d Cir. 2008).
125 See infra Part V.
requirement of judicial review under the Public Use Clause as interpreted by the Supreme Court in *Kelo*.

Courts considered blight removal to be a valid public purpose for condemnation long before the Supreme Court ruled that economic development was, additionally, a valid public purpose.\textsuperscript{126} The traditional definition of blight, however, was narrow: it was limited to “slums . . . whose eradication was itself found to constitute a public purpose for . . . condemnation” because they created conditions that threatened the health and welfare of the surrounding community.\textsuperscript{127} Today, New York’s criteria for condemnation due to blight include a wide array of factors, including a simple lack of economic development.\textsuperscript{128} The New York State Urban Development Corporation Act (UDC Act) prescribes that, for a finding of blight, the condemning authority must determine that “the area in which the project is to be located is a substandard or insanitary area, or is in danger of becoming a substandard or insanitary area and tends to impair or arrest the sound growth and development of the municipality.”\textsuperscript{129} Though “insanitary” is a historical criterion for blight, “substandard” describes a broad range of conditions that a legislature might find in any area it determines to be in need of economic development without finding traditional blighting factors.\textsuperscript{130} The New York Court of Appeals embraced this broad definition of blight in *Yonkers Community Development Agency v. Morris*, holding that, “areas eligible for . . . renewal are not limited to ‘slums’ as that term was formerly applied, and that, among other things, economic underdevelopment and stagnation are also threats to the public sufficient to make their removal cognizable as a public purpose.”\textsuperscript{131} This decision not only established economic development as a legitimate public purpose in New York; it also expanded the definition of blight to encompass the need for economic development. Thus, the line between economic development and blight remediation was blurred in New York well before *Kelo* established the federal baseline for economic development,\textsuperscript{132} and certainly before the Second Circuit limited the application of *Kelo’s* heightened standard of review to allegations of pretext.

\textsuperscript{126} See, e.g., Berman v. Parker, 348 U.S. 26, 32 (1954).
\textsuperscript{128} Id.
\textsuperscript{129} Id.
\textsuperscript{130} N.Y. UNCONSOL. LAW § 6260(c)(1) (McKinney 2000).
\textsuperscript{131} 
\textsuperscript{132} Id. (emphasis added).
\textsuperscript{132} Cormack v. Settle-Beshears, 474 F.3d 528, 531 (8th Cir. 2007).
concerning takings accomplished solely for economic development purposes.\footnote{Goldstein v. Pataki, 516 F.3d 50, 64 (2d Cir. 2008).}

The blurred distinction in New York between blight remediation and economic development is exemplified by the cases surrounding the use of eminent domain in Times Square during the mid-1980s.\footnote{See Natural Res. Def. Council, Inc. v. New York, 672 F.2d 292, 294 (2d Cir. 1982); see also In re G. & A. Books, Inc., 770 F.2d 288, 291 (2d Cir. 1985).} There, courts relied on factors such as underutilization and suboptimal tax revenues to justify the elimination of urban blight through eminent domain.\footnote{Id.} In \textit{Natural Resources Defense Council, Inc. v. New York}, the Second Circuit upheld a determination of blight based on, among other factors, the underutilization of property, high vacancy rates, rundown storefronts, and the presence of pornographic businesses.\footnote{Id.} These conditions, the court held, led to an “unproductive use of potentially valuable land,” and thus justified the use of eminent domain.\footnote{\textit{Id.}; see also \textit{In re G. & A. Books, Inc.}, 770 F.2d at 292 (“Only 4,000 people, an extraordinarily low figure for a five-block area located adjacent to one of the world’s most densely developed business districts, work in the area. As a result of an absence of development for more than half a century, the existing buildings are old and rundown; most are substandard for their intended commercial uses and many are vacant above the first floor. While the area is zoned for the highest density allowed in the City, 16\% of the land area is used only for parking, 72\% of the development rights have not been used, and 18\% of the developed parts is vacant. The tax yield from the Project area is commensurately low: the FEIS estimated that while the existing properties in the Project area were expected to pay approximately $5.4 million in taxes in 1984-85, a single building a block away was expected to pay $6.2 million in taxes.”.).}

In \textit{In re G. & A. Books, Inc.}, the findings of blight were clearly dependent on determinations that Times Square was less economically productive than it could have been with proper redevelopment.\footnote{\textit{Id.}} This is precisely the justification the City of New London provided in determining the need to exercise its power of eminent domain as described in \textit{Kelo}.\footnote{Kelo v. City of New London, 545 U.S. 469, 472 (2005).} Thus, courts should apply the standard developed in \textit{Kelo} to takings in New York that claim blight remediation as their public purpose.

As in blight cases, underutilization has also been used to justify takings solely for economic development in New York, underscoring the convergence of the two public purposes.\footnote{Sunrise Props., Inc. v. Jamestown Urban Renewal Agency, 614 N.Y.S.2d 841, 842 (N.Y. App. Div. 1994).} In \textit{Sunrise Properties v. Jamestown Urban Renewal Agency}, the
taking of private land was authorized for job creation, infrastructure development, and general economic improvement of the project area.\footnote{Id.} In upholding the taking as constitutional, the court stated, “The finding . . . that the property is \textit{underutilized} is equivalent to a determination that condemnation of the property and subsequent development will serve a public purpose.”\footnote{Id. (emphasis added).} Thus, underutilization was used to justify a taking for economic development just as it was used to justify takings for blight remediation in Times Square.\footnote{See \textit{In re G. & A. Books, Inc.}, 770 F.2d at 292.}

Furthermore, just as creating jobs and increasing tax revenue were used to justify blight remediation in Times Square these exact factors are the basis for takings in New York aimed at promoting economic development. In \textit{In re Fisher}, New York City took land in Lower Manhattan for a private transfer to the New York Stock Exchange without a finding of blight.\footnote{Id. at 517.} There, the court upheld the city’s actions because it found that the taking would spur economic development through increased job opportunities and tax revenues.\footnote{\textit{Kelo} v. City of New London, 545 U.S. 469, 478 (2005).} Clearly, the line between blight remediation and economic development in New York is blurred beyond any substantial distinction. Therefore, challenges to a taking under the Public Use Clause in New York should receive the same standard of review regardless of whether the taking is justified as a means to create economic development or to eliminate conditions that cause blight.

Though \textit{Kelo} requires an inquiry into takings where the stated public purpose is mere pretext for bestowing a private benefit,\footnote{See \textit{In re Fisher}, 730 N.Y.S.2d 516, 516-17 (N.Y. App. Div. 2001).} New York courts apply an absolutely deferential standard to a legislative or administrative decision to take land for blight remediation:

\begin{quote}
It is only where there is no room for reasonable difference of opinion as to whether an area is blighted, that judges may substitute their views as to the adequacy with which the public purpose of blight removal has been made out for those of the legislatively designated agencies; where . . . “those bodies have made their finding, not corruptly or irrationally or baselessly, there is nothing for the courts
\end{quote}

\footnote{Id.}
This deferential treatment is not appropriate for cases in which challengers allege impermissible favoritism in violation of the Public Use Clause and support such allegations with substantial evidence.\textsuperscript{148} Thus, the standard applied by New York courts is below the federal baseline established in \textit{Kelo} and enables legislatures to abuse their power of eminent domain.\textsuperscript{149}

New York is one of only eleven states that permit takings for solely economic development purposes.\textsuperscript{150} Thus, a


\textsuperscript{148}See supra Part II.

\textsuperscript{149}See supra Part II.

\textsuperscript{150}The following research was conducted by Justin Kamen, Whitney Philips, and Ellie Merle, under the direction of Norman Siegel, Esq., from June-August, 2010. The following eleven states allow eminent domain takings for the sole purpose of economic development:

\begin{itemize}
\item[Hawaii] HAW. REV. STAT. § 101-2 (2010).
\item[Illinois] 735 ILL. COMP. STAT. 30/5-5-5(c), 620/9 (2011); see Friends of the Parks v. Chi. Park Dist., 786 N.E.2d 161, 167 (Ill. 2003); Sw. Ill. Dev. Auth. v. Nat'l City Envtl., L.L.C., 768 N.E.2d 1, 13-14 (Ill. 2002).
\item[Rhode Island] See R.I. GEN. LAWS § 42-64.12-8 (2011).
\item[South Dakota] S.D. CODIFIED LAWS § 11-8-50 (2010).
\item[Utah] See UTAH CODE ANN. § 10-8-2(e)(iii) (LexisNexis 2011).
\end{itemize}

The following thirty-nine states do not allow eminent domain takings for the sole purpose of economic development:

\begin{itemize}
\item[Alabama] See ALA. CODE § 18-1B-2 (2010).
\item[Alaska] ALASKA STAT. § 09.55.240(d) (2010).
\item[Arizona] A.Z. CONST. art. II, § 17; see ARIZ. REV. STAT. ANN § 12-1111 (2010).
\item[Colorado] COLO. REV. STAT. § 38-8-101(b)(1) (2010).
\item[Connecticut] CONN. GEN. STAT. § 8-127a(1) (2010).
\item[Florida] FLA. STAT. § 73.014 (2010).
\item[Georgia] GA. CODE ANN. § 22-1-19(B) (2011).
\item[Idaho] IDAHO CODE ANN. § 7-701A(2)(b) (2010).
\item[Indiana] IND. CODE §§ 32-24-4.5-1(a), 36-7-14-20, 36-7-14-43(7) (2010).
\item[ Iowa] IOWA CODE § 6A.22/2) (2010).
\item[Kentucky] KY. REV. STAT. ANN. § 416.675 (West 2010).
\item[Louisiana] LA. CONSTIT. art. I, § 4(B)(3).
\item[Maine] ME. REV. STAT. ANN tit. 1, § 816 (2010).
\item[Minnesota] MINN. STAT. § 117.025(11)(b) (2010).
\item[Missouri] MO. ANN. STAT. § 523.271 (West 2010).
\item[Montana] See MONT. CODE ANN. § 70-30-102 (2010).
\item[Nebraska] NEB. REV. STAT. § 76-710.04(1) (2010).
\item[Nevada] See REV. STAT. ANN. § 37.010 (2010).
\item[New Mexico] N.M. STAT. ANN. § 3-60A-10(L)(3) (2010).
\item[Ohio] OHIO REV. CODE ANN.
condemning authority can claim blight remediation, rather than economic development, as its purpose for taking property so as to avoid close judicial review of its motives.\textsuperscript{151} In states that do not permit eminent domain solely for economic development purposes, courts review evidence to ensure that condemnors do not manufacture blight findings in attempts to conceal a true purpose of strictly economic development.\textsuperscript{152} For example, in \textit{City of Norwood v. Horney}, the Supreme Court of Ohio struck down the use of eminent domain where the condemning authority sought to take property in what it seemed to be a “deteriorating area.”\textsuperscript{153} There, the court found the taking illegal because the factors used to determine whether an area was deteriorating, including increased traffic, numerous curb cuts, and small front yards, created “a standardless standard,” by which a condemning agency could describe practically any city.\textsuperscript{154} Therefore, the taking was founded solely on the promotion of economic development and, thus, did not have a sufficient public purpose under the Ohio Constitution.\textsuperscript{155} However, in New York, no such scrutiny is applied to blight takings because economic development is a sufficient public purpose in and of itself for the use of eminent domain.\textsuperscript{156} Also, the criteria used in New York to justify takings for both blight elimination and for economic development are nearly identical, rendering the line between these two public purposes blurred beyond substantive recognition.\textsuperscript{157} Thus, it is vitally important that New York courts apply the federal baseline established in \textit{Kelo} to blight-remediating takings.\textsuperscript{158} Otherwise, legislatures can circumvent

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{151}See Goldstein v. Pataki, 516 F.3d 50, 64 (2d Cir. 2008).
\item \textsuperscript{152}See, e.g., City of Norwood v. Horney, 853 N.E.2d 1115, 1123 (Ohio 2006).
\item \textsuperscript{153}\textit{Id.} at 1145.
\item \textsuperscript{154}\textit{Id.} at 1144-45.
\item \textsuperscript{155}\textit{Id.} at 1142.
\item \textsuperscript{158}See supra Part II.
\end{itemize}
\end{footnotesize}
Kelo’s ban on takings for impermissible private purposes by claiming a purpose of blight remediation rather than economic development, as was the case in Kaur.\textsuperscript{159}

With the Second Circuit’s refusal to extend Kelo’s pretext analysis to takings for blight remediation in Goldstein,\textsuperscript{160} condemning authorities in New York can avoid a heightened standard of review for takings that exhibit evidence of impermissible favoritism\textsuperscript{161} by simply claiming blight remediation as the purpose for a taking rather than economic development. Thus, such an authority can bypass judicial scrutiny and bestow a private benefit through eminent domain by simply manufacturing a blight study.\textsuperscript{162} This was the case in Kaur, where the finding of blight in Manhattanville was supported by the desire to create jobs, increase tax revenue, and prevent underutilization of property.\textsuperscript{163} In this way, the public purpose of blight remediation stated in Kaur is very similar to the public purpose of economic development used elsewhere in the State.\textsuperscript{164} As such, judicial review of purpose in eminent domain takings should be equivalent in these instances. Because the New York Court of Appeals applied a deferential standard in Kaur and declined to apply the heightened scrutiny demanded by Kelo, the Supreme Court should have remanded the case back to the New York Court of Appeals for argument on the merits as to the claims of impermissible favoritism.\textsuperscript{165} Instead, the Supreme Court denied certiorari and tacitly condoned the erroneous deference applied by the New York Court of Appeals to ESDC’s exercise of eminent domain for the sole benefit of Columbia University.\textsuperscript{166}

V. Kaur v. New York State Urban Development Corporation: An Unchecked Abuse of Power

In Kaur, the New York Court of Appeals declined to address the mere pretext analysis demanded by Kelo despite

\begin{itemize}
\item \textsuperscript{159} See Kaur, 933 N.E.2d at 724.
\item \textsuperscript{160} Goldstein v. Pataki, 516 F.3d 50, 64 (2d Cir. 2008).
\item \textsuperscript{161} See supra Part II.
\item \textsuperscript{162} See supra Part II.
\item \textsuperscript{163} Kaur, 933 N.E.2d at 729.
\item \textsuperscript{164} Id. at 726.
\item \textsuperscript{166} See supra Part II.
\item \textsuperscript{167} See infra Part V.
\item \textsuperscript{168} Tuck-It-Away, Inc. v. N.Y. State Urban Dev. Corp., 131 S. Ct. 822, 822-23 (2010).
\end{itemize}
well-founded allegations that the taking was motivated by a private purpose.\textsuperscript{169} There, ESDC purportedly sought to remediate blight in Manhattanville, a neighborhood on the Upper West Side of Manhattan, by transferring property to Columbia University.\textsuperscript{170} Because the challengers in \textit{Kaur} presented a “plausible accusation of impermissible favoritism to private parties,”\textsuperscript{171} the court should have applied heightened scrutiny rather than the standard deference afforded to legislatures in typical public use challenges. Moreover, the evidence of impermissible favoritism in \textit{Kaur} was nearly identical to the hypothetical factors outlined in Justice Kennedy’s \textit{Kelo} concurrence, which would be subject to heightened scrutiny.\textsuperscript{172} Finally, the procedures in New York for effecting a condemnation of property are particularly prone to abuse and thus challenges to such takings should receive proper consideration by the judiciary.\textsuperscript{173} Thus, the New York Court of Appeals should have applied a heightened standard of review in \textit{Kaur} due to evidence strongly suggesting that the taking, purportedly justified by blight remediation, was actually intended to bestow a purely private benefit.\textsuperscript{174} Because the courts have failed to prevent eminent domain abuse in New York, the legislature should take steps to reduce the instances and inequalities of takings that have private beneficiaries.

\textbf{A. Evidence of Impermissible Favoritism in Kaur}

The facts of \textit{Kaur} were strikingly similar to those described as highly suspect by Justice Kennedy’s hypothetical example of an impermissible taking in his \textit{Kelo} concurrence, and thus the case demanded heightened judicial scrutiny.\textsuperscript{175} Unlike in \textit{Kelo}, where the private beneficiary was unknown at the time the redevelopment plan originated,\textsuperscript{176} Columbia’s attorneys, consultants, and architects drafted every document concerning the Manhattanville redevelopment plan.\textsuperscript{177} Additionally, numerous actions taken by both ESDC and Columbia throughout the

\textsuperscript{169} See \textit{Kaur}, 933 N.E.2d at 732.
\textsuperscript{172} \textit{Id.} at 493.
\textsuperscript{173} See infra Part V.B.
\textsuperscript{174} See \textit{In re Kaur}, 892 N.Y.S.2d at 28.
\textsuperscript{175} See supra Part II.
\textsuperscript{176} \textit{Kelo}, 545 U.S. at 493 (Kennedy, J., concurring).
\textsuperscript{177} \textit{In re Kaur}, 892 N.Y.S.2d at 20.
condemnation process suggest the kind of conspiratorial relationship envisioned by the Supreme Court in *Kelo* to constitute impermissible favoritism.\(^{178}\) Such evidence includes Columbia’s misdeeds as the dominant property owner in Manhattanville,\(^{179}\) ESDC’s reliance on blight studies performed by Columbia’s advisor,\(^{180}\) and ESDC’s attempts to obfuscate the record by withholding important documents during the litigation process.\(^{181}\)

Perhaps the most striking evidence of eminent domain abuse in *Kaur* was Columbia’s own role in creating the factors that ultimately led to ESDC’s determination of blight in the area, such as underutilization of property and the existence of building code violations.\(^{182}\) When ESDC considered developing the area in 2002, its Master Plan described no blight or blighted conditions in Manhattanville.\(^{183}\) No blight studies were conducted thereafter until 2006, when Columbia had already taken control of “the very properties that would form the basis for a subsequent blight study.”\(^{184}\) As owner of these properties, Columbia vacated much of the real estate by forcing more than 50 percent of the tenants out of seventeen buildings.\(^{185}\) Additionally, Columbia facilitated the degeneration of the neighborhood by failing to address water infiltration and building code violations, allowing tenants to violate local codes and ordinances, and maintaining garbage and debris in its properties for several years.\(^{186}\) Further, the Appellate Division found that Manhattanville “was not in a depressed economic condition when . . . ESDC embarked on their Columbia-prepared-and-financed quest.”\(^{187}\) As such, the taking in *Kaur* provided substantial evidence of impermissible favoritism.

Columbia’s creation of blighting factors distinguishes *Kaur* from *Western Seafood*, where a private developer received taken land in an economically stagnant community to spur development.\(^{188}\) In that case, the Fifth Circuit applied *Kelo*’s test for mere pretext because the private developer had a suspicious

\(^{178}\) Id. at 21.
\(^{179}\) Id.
\(^{180}\) Id. at 12-13.
\(^{181}\) Id. at 29 (Richter, J., concurring).
\(^{182}\) Id. at 21.
\(^{183}\) Id. at 19.
\(^{184}\) Id. at 21.
\(^{185}\) Id.
\(^{186}\) Id.
\(^{187}\) Id. at 19.
\(^{188}\) W. Seafood Co. v. United States, 202 F. App’x 670, 675 (5th Cir. 2006).
amount of control over the project. However, the court allowed the taking: “because the [private developer] own[ed] acres of property along the river where the marina [was] to be built, the City’s interest in their collaboration [was] logical.”

While this situation is similar to the facts of *Kaur*, *Western Seafood* shows no evidence that the private developer caused the very blighting factors that supplied the need for a taking. In *Kaur*, on the other hand, the city rewarded Columbia for causing blight by giving the irresponsible property owner even more land. Therefore, heightened scrutiny was appropriate in *Kaur* but not in *Western Seafood*.

ESDC’s complicity with Columbia’s efforts to justify the use of eminent domain in Manhattanville provides further evidence of impermissible favoritism. In 2006, ESDC hired private consultant Allee King Rosen and Flemming, Inc. (AKRF) to conduct a blight study of Manhattanville. AKRF’s finding of blight was tainted, however, by its previous role in assisting Columbia to develop and execute an expansion plan in 2004. When the challenging property owners issued a Freedom of Information Law (FOIL) request for documents concerning the relationship between AKRF, ESDC, and Columbia, the court forced ESDC to disclose the documents because “the difficulty of offering perfectly objective advice while serving two masters elevat[e]d the FOIL appeal beyond the average agency-consultant relationship that the FOIL exemptions are designed to foster and protect.” By employing Columbia’s consultant for its initial blight study, ESDC clearly favored Columbia in the process of determining the need for exercising eminent domain in Manhattanville. Though ESDC subsequently replaced AKRF with Earth Tech, a consultant without suspicious ties to Columbia, ESDC requested that Earth Tech “‘replicate’ the AKRF study using the same flawed

---

189 Id.
190 Id.
191 See id.
193 Id. at 20.
195 Id. at 60.
196 Id.
197 See *In re Kaur*, 892 N.Y.S.2d at 20.
methodology” previously employed by AKRF. Thus, ESDC’s final blight study was just as tarnished by favoritism to Columbia as the original AKRF study had been. Additionally, Earth Tech’s study was not completed until 2008, at which point “the ESDC/Columbia steamroller had virtually run its course to the fullest.” This is exactly the kind of public-private conspiracy Justice Kennedy envisioned in Kelo to demand heightened scrutiny under the Public Use Clause.

One of the documents unveiled by the aforementioned FOIL litigation reveals the dramatic extent of ESDC’s role in creating a pretextual justification for the exercise of eminent domain on Columbia’s behalf. In an e-mail sent before hiring AKRF in 2006, ESDC Senior Counsel Joseph Petillo questioned the wisdom of conducting a blight study: “I am uncomfortable with [ESDC] shining a spotlight on the process used to manufacture support for condemnation . . . . [M]aybe we want to craft the support for our blight findings in a less public way . . . .” This e-mail was clearly evidence of unconstitutional collusion, as ESDC intended to “manufacture” support for the taking to benefit Columbia. Furthermore, ESDC not only withheld this and similar documents from challenging property owners despite numerous FOIL requests and litigations, but it also refused to keep the record open until the FOIL litigation initiated by the landowners was completed. As such, ESDC exhibited impermissible favoritism toward Columbia from the planning phase through the entire litigation process.

The Appellate Division, which first heard Kaur, concluded from this evidence that ESDC used its blight finding as pretext to bestowing a benefit on Columbia and, as per the Supreme Court’s instructions in Kelo, held that ESDC did not take the private property for a legitimate public purpose. When the New York Court of Appeals reversed the Appellate Division’s decision, the court failed to even mention Kelo, let alone apply Kennedy’s test for heightened scrutiny. Thus, the

198 Id. at 22.
199 See id.
200 Id. at 19.
201 See supra Part II.
202 Tuck-It-Away Cert. Petition, supra note 79, at 8.
203 Id.
204 In re Kaur, 892 N.Y.S.2d at 29 (Richter, J., concurring).
205 Id. at 29-30.
206 Id. at 28 (majority opinion).
Court of Appeals misapplied the *Kelo* standard where it was clearly applicable and therefore decided *Kaur* incorrectly.

While the Second Circuit and the New York Court of Appeals also declined to apply a heightened standard of review in the *Goldstein* cases, the facts of the Atlantic Yards taking were more closely aligned with the facts of *Kelo* than with Kennedy's hypothetical, and thus, unlike in *Kaur*, the courts were justified when they dismissed the public use challenges. When ESDC decided to improve the Atlantic Yards Project Area through private development by the FCRC, the finding of blight was supported by over forty years of previous studies that had reached similar conclusions. Particularly, the blight studies determined the need to eliminate a large abandoned railway that was the main factor in the area's economic deterioration. This supported ESDC's decision to condemn property for transfer to FCRC on the basis that the area was, in fact, blighted. As blight remediation is a legitimate public purpose for the exercise of eminent domain, the courts were correct to uphold the taking of land in Atlantic Yards despite their failure to properly apply *Kelo*. Such evidence was not present in *Kaur*, however, and thus that case was erroneously decided under an overly deferential standard of review.

### B. Procedural Impediments to Public Use Challenges in New York

The New York Court of Appeals should have applied heightened scrutiny in *Kaur* not only because factual evidence strongly suggested impermissible favoritism, but also because, under New York law, the challengers were unable to seek review at a trial level court. Instead, they were required to begin the litigation process at the appellate level after establishing the record at a public hearing conducted by

---

208 Goldstein v. Pataki, 516 F.3d 50, 64 (2d Cir. 2008); see *In re Goldstein v. N.Y. State Urban Dev. Corp.*, 921 N.E.2d 164, 172 (N.Y. 2009).

209 *See supra* Part II.


211 *In re Goldstein*, 921 N.E.2d at 175.


213 *See supra* Part III.


215 *See N.Y. EM. DOM. LAW § 207* (Consol. 2011).
ESDC. In *Kelo*, Justice Kennedy’s call for heightened scrutiny in some public use challenges included cases where “the procedures employed [are] prone to abuse . . . .” Kennedy found a seven-day bench trial before the Superior Court to be an adequate procedural safeguard because property owners could challenge the legitimacy of the taking in a fair, adversarial proceeding. On the other hand, New York is currently the only state in which challenges to the legitimacy of an alleged public purpose for the exercise of eminent domain do not receive judicial review at the trial court level. Thus, the

---

216 *Id.*


218 *Id.*

219 The following research was conducted by Justin Kamen, Whitney Philips, and Ellie Merle, under the direction of Norman Siegel, Esq., from June to August 2010. The following forty-nine States provide trial court level judicial review to determine the legitimacy of a stated public purpose for the exercise of the eminent domain power:

- **Alabama:** See Ala. Code §§ 18-1A-91(b), 18-1A-94(b), 18-1A-130 (2010).
- **Alaska:** Alaska Stat. §§ 09.55.290, 09.55.300 (2010).
- **Colorado:** See Dunham v. City of Golden, 504 P.2d 360, 361-62 (Colo. App. 1972) (affirming the district court’s decision to allow condemnation of private property where the city sought to widen a public street).
- **Connecticut:** Hall v. Weston, 355 A.2d 79, 85 (Conn. 1974).
- **Florida:** Rukab v. City of Jacksonville Beach, 811 So. 2d 727, 733 (Fla. Dist. Ct. App. 2002).
- **Idaho:** Idaho Code Ann. §§ 7-706, 7-721(2) (2010).
- **Iowa:** Iowa Code § 6A.24 (2010).
- **Kentucky:** See Commonwealth v. Cooksey, 948 S.W.2d 122, 123 (Ky. Ct. App. 1997).
- **Maryland:** See Ance Arundel Cnty. v. Burnopp, 478 A.2d 315, 316, 318 (Md. 1984).
- **Massachusetts:** See Poremba v. Springfield, 238 N.E.2d 43, 48 (Mass. 1968) (affirming the superior court’s decision to uphold the taking of private land for the construction of a public highway).
- **Minnesota:** See Minn. Stat. § 117.075 (2010).
- **Mississippi:** See Miss. Code Ann. § 11-27-15 (2010); Mayor of Vicksburg v. Thomas, 645 So. 2d 940, 941 (Miss. 1994).
- **Montana:** See Groundwater v. Wright, 588 P.2d 1003, 1004 (Mont. 1979).
- **Nebraska:** See City of Omaha v. Tract No. 1, 778 N.W.2d 122, 125 (Neb. Ct. App. 2010).
- **North Dakota:** City of Medora v. Golberg, 569 N.W.2d 257, 258 (N.D. 1997).
- **Ohio:** Ohio Rev. Code Ann. §§ 163.08, 163.63 (2011).
- **Oklahoma:** See City of Midwest City v. House of Realty, Inc., 100 P.3d 678, 682 (Okla. 2004).
- **Oregon:** Or. Rev. Stat. § 35.015(c) (2009); see Eugene v. Johnson, 192 P.2d 251, 255 (Or. 1948).
- **Pennsylvania:** See In re Condemnation Proceeding by the Redevelopment Auth. of Phila. (1839 N. Eighth St.), 938 A.2d 341, 344 (Pa. 2007).
New York Court of Appeals should have applied more exacting scrutiny in *Kaur* due to suboptimal procedural protections against the exercise of eminent domain for an illegitimate private purpose.  

Under New York's Eminent Domain Procedural Law (EDPL), a condemning authority is required to make a “determination and findings concerning the proposed public project,” in which it must specify “the public use, benefit or purpose to be served by the proposed public project.” Prior to its decision to take property, the condemnor must hold a public hearing where “any person . . . shall be given a reasonable opportunity to present [a] . . . statement . . . concerning the proposed public project. A record of the hearing shall be kept . . . .” Any party wishing to challenge a decision to condemn property “may seek judicial review thereof by the appellate division of the supreme court . . . . The court shall either confirm or reject the condemnor's determination and findings. The scope of review shall be limited to whether . . . a public use, benefit or purpose will be served by the proposed acquisition.” In *Jackson v. New York State Urban Development Corp.*, the New York Court of Appeals found that the EDPL does not “[require] a trial-type hearing to challenge a tentative decision to condemn.” Therefore, a property owner who wishes to challenge the legitimacy of a taking under the Public Use Clause does not have the opportunity to present evidence to a neutral decision maker in a trial court for preservation of a record. Rather, such property owner must first present evidence at a public hearing conducted by the

---


220 See supra Part II.

221 N.Y. EM. DOM. LAW § 204 (Consol. 2011).

222 Id. § 201.

223 Id. § 203.

224 Id. § 207 (emphasis added).


226 *Brody v. Vill. of Port Chester*, 434 F.3d 121, 133 n.10 (2d Cir. 2005).
condemning authority.\textsuperscript{227} The record that would normally be established in a trial court before a judge\textsuperscript{228} is thus developed under the guidance of the agency seeking to exercise eminent domain. When the property owner challenges the agency’s decision to condemn, the appellate court hears evidence developed at such hearing.\textsuperscript{229} This system creates procedural impediments to challenge the stated public purpose for a taking and is thus prone to abuse by condemning authorities.\textsuperscript{230}

The procedures outlined in the EDPL have been upheld as constitutional in New York and thus do not, on their own, violate a challenger’s due process rights under the Fifth Amendment.\textsuperscript{231} The Fifth Amendment mandates that the government shall not deprive citizens of property without due process,\textsuperscript{232} the adequacy of which is determined by considering the private interest at risk of deprivation by the procedure, the risk of erroneous deprivation of that interest, and the government’s interest in implementing the procedure.\textsuperscript{233} The Second Circuit ruled in \textit{Brody v. Village of Port Chester} that the challenger of a taking “has no constitutional right to participate in the [agency’s] initial decision to exercise its power of eminent domain, and the post determination review procedure set forth in EDPL § 207 is sufficient” to provide challengers adequate process.\textsuperscript{234} Because challengers can raise their claims at the public hearing required under the EDPL, they have an opportunity to make a record by presenting their views and submitting evidence.\textsuperscript{235} As such, the procedure here described is not itself a violation of the due process right “to be heard ‘at a meaningful time and in a meaningful manner.’”\textsuperscript{236}

However, Justice Kennedy’s hypothetical in \textit{Kelo} does not require a violation of due process to trigger heightened scrutiny when there are allegations that the stated public purpose for a

\begin{footnotes}
\item[227] N.Y. E.M. DOM. LAW § 204.
\item[228] See supra note 219.
\item[230] Id.
\item[232] U.S. CONST. amend. V.
\item[234] Brody v. Vill. of Port Chester, 434 F.3d 121, 133 (2d Cir. 2005); see also Vill. Auto Body Works, Inc., 454 N.Y.S.2d at 743 (holding that EDPL § 207 “does not violate either the procedural or substantive due process rights of the property owner” (citation omitted)).
\item[236] Mathews, 424 U.S. at 333 (quoting Armstrong v. Manzo, 380 U.S. 545, 552 (1965)).
\end{footnotes}
taking is mere pretext to bestow a private benefit. Rather, his standards for denying a presumption of validity require only that “the procedures employed [are] prone to abuse . . . .” Though a public hearing allows challengers to establish a record that can later be heard by an appellate court, there are “practical impediments” inherent in this non-adversarial forum to demonstrating impermissible favoritism. For example, in Kaur, ESDC closed the record despite the fact that the challenging landowners were engaged in FOIL litigation to retrieve documents from ESDC in support of their public use challenge. Had the record been established under the supervision of a neutral arbiter rather than by the condemning authority itself, the plaintiffs in Kaur may have been able to use these documents to state a more compelling case of impermissible favoritism. Although the EDPL does not facially violate the Due Process Clause, it sufficiently burdens challengers to trigger heightened judicial review under Kennedy’s analysis for mere pretext. Therefore, the New York Court of Appeals should have applied a heightened standard of review in Kaur.

C. Possible Legislative Solutions to Eminent Domain in New York

If the New York courts continue to allow takings for unconstitutional private purposes, then the New York State Legislature must take steps to prevent the abuses that occurred in Kaur from recurring in the future. Most obviously, New York should join the other forty-nine states in requiring trial level review of eminent domain challenges under the Public Use Clause to ensure procedures that are not “prone to abuse.” Also, if the courts insist on maintaining different standards of review, the application of which depend on whether a taking is intended to eliminate blight or solely to promote economic development, then the statutory scheme must redefine blight so as to avoid the
manipulations devised by ESDC and Columbia to elude heightened judicial scrutiny.\textsuperscript{246} Thus, New York should amend the UDC Act to construe “blighted area” narrowly. Specifically, the legislature should adopt Vermont’s statutory limitation that “[n]o area shall be determined to be a blighted area solely or primarily because its condition and value for tax purposes are less than the condition and value projected as the result of the implementation of any . . . private redevelopment plan.”\textsuperscript{247} Finally, New York should reduce the hardships imposed on persons displaced and disinherited by takings that seek solely to promote economic development by providing increased compensation for the property taken.\textsuperscript{248} The legislature should adopt Kansas’s approach, which requires the condemnor in an economic development taking to pay the landowner 150 percent of the subject property’s fair market value.\textsuperscript{249} By adopting these new laws, the New York State Legislature can reduce the instances and inequalities of eminent domain abuse despite the courts’ unwillingness to oppose private development interests and their political enablers.

CONCLUSION

The Supreme Court should have reversed the New York Court of Appeals in \textit{Kaur} and remanded the case for review using the heightened scrutiny required by \textit{Kelo}. Such a decision would have both defended the vulnerable populations harmed by the taking itself\textsuperscript{250} and settled the jurisdictional split regarding public use challenges to takings that are purportedly intended to remediate blight.\textsuperscript{251} Instead, the Second Circuit and New York courts will continue to defer to the judgments of condemning authorities that seek to abuse their power of eminent domain and who, in doing so, contribute to the widening gap between rich and poor throughout the state.\textsuperscript{252}

\textsuperscript{246} See supra Part IV.
\textsuperscript{248} See supra Part I.
\textsuperscript{249} KAN. STAT. ANN. § 26-501b (2009); see also R.I. GEN. LAWS § 42-64.12-8 (2011) (“Owners of property taken for economic development purposes shall be compensated for . . . [a] minimum of one hundred fifty percent of the fair market value of the real property.”).
\textsuperscript{250} See supra Part I.
\textsuperscript{251} See supra Part III.
The Supreme Court may have denied certiorari in *Kaur* because the New York Court of Appeals held that, besides eliminating blight, the expansion of Columbia University's campus in Manhattanville would serve a public purpose in and of itself as a "civic project." Under the UDC Act, ESDC is empowered to take land, not only to remediate blight, but also to "undertake the acquisition . . . of a [civic] project [when] . . . there exists . . . a need for the educational . . . facility . . . [and that] the project shall consist of . . . facilities which are suitable for educational . . . purposes." Thus, the Supreme Court may have declined to discuss the suspicious facts surrounding ESDC's finding of blight in *Kaur* due to an unwillingness to prevent the use of eminent domain in support of an important and beneficial educational institution. If so, the Court should take the next available opportunity to address the improper judicial deference expressed in *Goldstein* and displayed in *Kaur* in order to prevent the New York courts from further abdicating their duties as defenders of individual rights against the excesses of governmental power. However, if the New York courts continue to sanction eminent domain abuse despite the dictates of *Kelo*, the legislature must take action to protect vulnerable populations from takings that only benefit private parties.

*Justin B. Kamen†*

---

† B.A., Columbia University, 2008; J.D. Candidate, Brooklyn Law School, 2012. I wish to thank Norman Siegel, my friend and mentor whose tireless crusade to protect civil rights inspired this article. I also wish to thank my fellow research assistants, Whitney Philips and Ellie Merle, for their help on the state surveys. And of course, thank you to my friends, family, and especially to my parents, Roy and Marina Kamen, without whom I would be nothing.
Smoking Out Racism in the FDNY

THE DWINDLING USE OF RACE-CONSCIOUS HIRING REMEDIES

INTRODUCTION

The New York City Fire Department (the Fire Department or FDNY), despite its proud history, remains an organization unwelcoming to minorities. In 2007, the United States—later joined by the Vulcan Society and individual plaintiffs—brought suit against the Fire Department alleging that it discriminated against blacks and Hispanics in violation of Title VII of the Civil Rights Act of 1964.\(^1\) The U.S. District Court for the Eastern District of New York found that the Fire Department’s entry-level hiring examinations had both a disparate impact and was the product of systematic disparate treatment.\(^2\) However, the initial remedy—an interim race-conscious hiring plan—did not require the hiring of a sufficient number of black and Hispanic applicants to address the severe underrepresentation of minority firefighters.\(^3\) Nevertheless, the City of New York (the City) refused to implement any of these plans.\(^4\) As a result, the court imposed an extensive oversight plan that prematurely restrained the City from developing its own solution.\(^5\)

Before resorting to extensive oversight, the court should have adopted a race-conscious hiring plan that requires the City to hire a sufficient number of black and Hispanic applicants.

\(^1\) See Complaint ¶ 1, United States v. City of New York, No. 07-CV-2067 (E.D.N.Y. May 21, 2007), ECF No. 1. The Vulcan Society also brought a similar state claim under New York Human Rights laws. Intervenors’ Complaint ¶ 1, City of New York, No. 07-CV-2067 (E.D.N.Y. July 17, 2007), 2007 WL 3117053, ECF No. 48.


\(^3\) See infra Part V.A; City of New York, No. 07-CV-2067, 2010 WL 3709350, at *1 (E.D.N.Y. Sept. 13, 2010).


applicants. If such a plan were enforced, the Fire Department
would have had incentive to reform its hiring and recruiting
methods without having to continuously seek court approval.
The court likely avoided this option because it was sensitive to
an increasing sentiment that affirmative action is no longer an
appropriate method of relief. Unless and until this social and
political movement results in an amendment to Title VII,
however, future courts should embrace the benefits of race-
conscious hiring remedies.

Part I of this note provides a background of Title VII of
the Civil Rights Act of 1964, discussing the theories of
disparate treatment and disparate impact, as well as the broad
range of remedies available. Part II provides a summary of the
current affirmative-action debate, and concludes that court-
ordered, race-conscious relief is still a legitimate and important
remedy in Title VII claims. Part III discusses the history of
racial discrimination in the FDNY. Part IV discusses the
procedural history and current posture of United States v. City
of New York. Part V critiques the court in United States v. City
of New York for its extensive oversight plan, and suggests that
if it enforced an adequate ratio of race-conscious hiring at the
outset, it would have provided sufficient incentive for the Fire
Department to more permanently remedy the discrimination
on its own terms. Finally, this note concludes that future courts
should continue to embrace race-conscious injunctive relief,
especially where deep-seated racial discrimination is found.

I. TYPES OF CLAIMS AND RELIEF UNDER TITLE VII OF THE
CIVIL RIGHTS ACT OF 1964

Under Title VII of the Civil Rights Act of 1964, the court
has broad authority to grant relief. The court in United States v.
City of New York should have used this broad authority to
impose race-conscious interim hiring. Title VII makes it illegal
“for an employer . . . to fail or refuse to hire or to discharge any
individual, or otherwise to discriminate against any
individual . . . because of such individual’s race, color, religion,
sex, or national origin.” Title VII is “a law triggered by a
Nation’s concern over centuries of racial injustice and intended
to improve the lot of those who had been excluded from the

See infra Part II.

American dream for so long.” There are essentially two theories under which a plaintiff can prove an unlawful employment practice: disparate treatment and disparate impact.

A. Disparate Treatment and Disparate Impact Claims

In race-based disparate treatment claims, the principal issue is whether the employer “treats some people less favorably than others because of their race, [or] color.” These cases typically involve discrete instances of intentional discriminatory conduct, rather than general practices. Liability is found where the employer’s decisions are “actually motivated” by the employee’s protected trait. In other words, the protected trait must have “had a determinative influence” on the employer’s decision.

Disparate treatment claims are further subdivided into single-motive, mixed-motive, and systematic disparate treatment (or “pattern or practice”) theories. Single-motive and mixed-motive claims are the more traditional and more easily supported claims because they tend to focus on discrete events. Systematic disparate treatment, on the other hand, is in many ways more difficult to prove. Nevertheless, the court found systematic disparate treatment in United States v. City of New York.

A systematic disparate treatment (“pattern or practice”) case must establish that the employer has put in place an overall system that naturally (and purposely) leads to adverse employment actions that are based on employees’ protected class. Statistical evidence alone may establish such a case. For example, in a test-taking situation, if the ratio of minority applicants who pass the exam is two or three standard

---

8 United Steelworkers of Am. v. Weber, 443 U.S. 193, 204 (1979) (citation omitted) (internal quotation marks omitted).
12 Id.
13 See McDonnell Douglas, 411 U.S. at 801.
18 See id. at 339-40 & n.20.
deviations below the ratio of minority applicants who took the exam, that may suffice.\(^{19}\) Typically, however, courts look for anecdotal evidence of discrimination to buttress the statistical showing.\(^{20}\) The employer could attack the statistical showing by questioning the accuracy of the data collected, or by arguing the particular labor pool that the plaintiff used was not appropriate.\(^{21}\) The employer could also assert a nondiscriminatory reason for the disparity—for example, that the general population within the protected class simply does not prefer the occupation—but this defense is difficult to make credibly since the employer must support it with extensive evidence and cannot base its reasoning on stereotypical inferences.\(^{22}\) Once a pattern or practice is found, each plaintiff has the opportunity to present evidence of individual damages, for which the employer can offer a legitimate, nondiscriminatory reason.\(^{23}\)

The court in *United States v. City of New York* also found that the City discriminated against black and Hispanic applicants under the theory of disparate impact.\(^{24}\) Disparate impact cases, unlike disparate treatment cases, do not require discriminatory motive.\(^{25}\) Instead, the plaintiff must show that a particular policy, while not discriminatory on its face, disproportionally affects a protected class.\(^{26}\) This theory was developed to combat employment procedures that are not predictive of future job performance and have “built-in headwinds” that work against minority groups.\(^{27}\) However, applicants claiming injury under the disparate impact theory must still have been qualified for the job; the theory does not

\(^{19}\) See Anderson v. Douglas & Lomason Co., Inc., 26 F.3d 1277, 1291 n.26 (5th Cir. 1994); cf. Kadas v. MCI Systemhouse Corp., 255 F.3d 359, 362 (7th Cir. 2001) (noting that not all cases find two to three standard deviations to be sufficient and that there is no bright-line rule).

\(^{20}\) See, e.g., *Int'l Bhd. of Teamsters*, 431 U.S. at 339 (noting that plaintiffs did not rely on “statistics alone,” but instead “brought the cold numbers convincingly to life”).


\(^{22}\) See infra note 249 and accompanying text.

\(^{23}\) See *Int'l Bhd. of Teamsters*, 431 U.S. at 361-62.


\(^{26}\) See Griggs, 401 U.S. at 431.

\(^{27}\) See id. at 432 (internal quotation marks omitted).
simply allow “the less qualified [to] be preferred over the better qualified simply because of minority origins.”

B. Forms of Relief and the Value of Race-Conscious Measures

The wide-ranging forms of relief available under Title VII demonstrate that the court in United States v. City of New York both failed to utilize and subsequently abused this broad authority. The goal of relief in an employment discrimination action is always a combination of deterrence and compensation. Forms of relief include preliminary and permanent injunctions, back pay, and front pay. Disparate treatment claims, unlike disparate impact claims, also allow for compensatory damages and “any other equitable relief.”

1. The Development of Race-Conscious Injunctive Relief

Injunctive relief has been interpreted broadly—a district court “has not merely the power but the duty” to “bar like discrimination in the future.” Therefore, courts will often issue remedies that allow the court to monitor—but not dictate—the methods of—compliance. For example, it is regularly required that any future examination or “selection device” be reviewed by the court and approved before its use. Other court-ordered procedures typically require applicants to be

---

28 Id. at 436.
31 See, e.g., 42 U.S.C. §§ 2000e-5(g), 2000e-16(d), 12117(a); 29 U.S.C. § 216(b), (c).
34 Id. § 2000e-5.
hired at a certain ratio, usually reflecting “the applicant pool or the relevant work force,” which ensures that no disparate impact could later be found.” Courts understand that combating future discrimination will almost invariably require a race-conscious effort in order to be effective.38

The scope of injunctive relief in Title VII cases has expanded greatly over the last thirty-plus years. At first, while employers were not forced to engage in affirmative action measures, they were given permission to do so by courts without the threat of future litigation based on “reverse discrimination.” In 1979, the Court in *United Steel Workers of America v. Weber* explicitly allowed the employers to engage in “private, voluntary, race-conscious affirmative action plans.”39 To deny this form of relief, the Court opined, “would ‘bring about an end completely at variance with the purpose of the statute’ and must be rejected.”40 Since the original purpose of Title VII was to aid in the “the plight of the Negro in our economy,”41 “[i]t would be ironic indeed” to use it against this purpose.42

Specifically, the Court in *Weber* allowed the employer to set racial-equality goals by reserving half of the positions in its craft-worker training programs for blacks.43 The Court found that the plan did not “unnecessarily trammel the interests of the white employees” since it did not go so far as to require white employees to be discharged.44 The Court also found it important to emphasize that the plan was temporary; the plan ended the moment the percentage of blacks in the labor force was properly represented in the plant.45

In 1987, *United States v. Paradise* laid the foundation for injunctive relief in the form of court-ordered affirmative-action plans.46 The fundamental shortcoming of the court in *United States v. City of New York* was to discount the significance of *Paradise* in some instances and overextend its

---

37 Id. at *7 (quoting Guardians, 630 F.2d at 109).
38 See infra notes 51-54 and accompanying text.
40 Id. at 202 (quoting United States v. Pub. Utils. Comm’n, 345 U.S. 295, 315 (1953)).
41 Id. (quoting 110 CONG. REC. 6548 (1964) (remarks of Sen. Humphrey)).
42 Id. at 204. But see Ricci v. DeStefano, 129 S. Ct. 2658 (2009) (finding that employers engaged in “reverse discrimination” when the employer voluntarily invalidated an entrance exam under the belief that the exam had a disparate impact on minority applicants).
43 Weber, 443 U.S. at 198.
44 Id. at 208.
45 Id. at 208-09.
Paradise affirms that “courts have the authority and the duty not only to order an end to discriminatory practices, but also to correct and eliminate the present effects of past discrimination.”47 Almost twelve years after the district court ordered its initial decision, the court forced the employer “to take affirmative and substantial steps to open the upper ranks to black troopers.”48 The court’s order required 50 percent or more of corporal promotions be given to qualified black troopers until the employer developed its own nondiscriminatory plan.49 When granting relief, the corrective plan “must unavoidably consider race.”50

Paradise shows that broad, court-ordered injunctive relief is necessary because “the effects of past discrimination . . . ‘will not wither away of their own accord.’”51 For instance, allowing an employer to come up with its own integration plan may do little to prevent “the continuing effects” of discrimination if the courts have no mechanism to enforce the plan in the future.52 The injunctive approach is particularly essential when an employer has failed to correct the problem for decades; the FDNY has failed to do so arguably since its inception and at least since the district court recognized the Fire Department’s discriminatory practices in 1973.53 The relief must be race-conscious because the alternatives, such as “an outright ban on hiring or promotions, or continued use of a discriminatory selection procedure,” are less appealing.54

Of course, courts do not have unlimited leeway in determining what injunctive relief is appropriate, and at times the court in United States v. City of New York failed to pay heed to these limitations. Once the court determines that it should take some form of action, it must take care to ensure that its ordered plan is sufficiently narrow.55 The Supreme Court has articulated what factors are relevant in determining whether the plan is sufficiently narrow:

(i) the efficacy of alternative remedies; (ii) the planned duration of the remedy; (iii) the relationship between the percentage of minority

47 Id. at 154 (quoting NAACP v. Allen, 340 F. Supp. 703, 705 (M.D. Ala. 1972)) (internal quotation marks omitted).
48 Id. at 163 (quoting Paradise v. Prescott, 585 F. Supp. 72, 74 (M.D. Ala. 1983)).
49 Id.
50 Id. at 194 (Stevens, J., concurring).
51 Id. at 163 (majority opinion) (quoting Paradise, 585 F. Supp. at 75-76).
52 Id. (quoting Paradise, 585 F. Supp. at 75-76).
53 See id. at 166-67; supra Part III.
55 See Paradise, 480 U.S. at 187 (Powell, J., concurring).
workers to be employed and the percentage of minority group members in the relevant population or work force; (iv) the availability of waiver provisions if the hiring plan could not be met; and (v) the effect of the remedy upon innocent third parties.\footnote{56 Id. (citations omitted).}

Under this analysis, the Court in \textit{Paradise} found the lower court’s order to be sufficiently narrow. The Court emphasized that the length of the plan was appropriate because it was “contingent upon the Department’s own conduct.”\footnote{57 Id. at 178 (majority opinion) (“The requirement endures only until the Department comes up with a procedure that does not have a discriminatory impact on blacks—something the Department was enjoined to do in 1972 and expressly promised to do by 1980.”).} The fifty percent hiring requirement acted as “an end date, which regulated the speed of progress toward fulfillment of the hiring goal.”\footnote{58 Id. at 180 (citing \textit{Sheet Metal Workers}, 478 U.S. at 487-88 (Powell, J., concurring)).} In addition, the order did not place an absolute bar on white employees because it did not require that any of them be discharged,\footnote{59 Cf. \textit{Ricci v. DeStefano}, 129 S. Ct. 2658 (2009) (finding disparate treatment where, by throwing out an entry-level exam after the applicants’ scores were already submitted, the white applicants were in a sense discharged).} and it did not reduce the quality of workers promoted because they must all still be qualified.\footnote{60 \textit{Paradise}, 480 U.S. at 183.} Ultimately, the court-imposed hiring scheme in \textit{Paradise} was successful. After the order was implemented, the defendant had sufficient incentive to promptly submit a nondiscriminatory promotion method, and as a result, the race-conscious court order was lifted shortly thereafter.\footnote{61 \textit{Id.} at 179.} Thus, the relief granted in \textit{Paradise} established a successful template that the court in \textit{United States v. City of New York} failed to adequately consider.

Limits on injunctive relief are particularly salient when dealing with government entities. \textit{Missouri v. Jenkins} set the outer limit for court control over local government through injunctive relief.\footnote{62 \textit{Id.} at 180 (citing \textit{Sheet Metal Workers}, 478 U.S. at 487-88 (Powell, J., concurring)).} There, the Supreme Court considered whether the district court had the power to direct an increase in local government taxes to ensure adequate funding for a desegregation plan.\footnote{63 \textit{Id.} at 39.} The Court noted that a “proper respect for the integrity and function of local government institutions” was a substantially important consideration.\footnote{64 \textit{Id.} at 51.} The Court found that
the lower court, by imposing a tax increase in local
government, completely circumvented local authority. The
Court noted that the lower court had an equally viable and
significantly less invasive alternative; it could have simply
required local government “to levy property taxes at a rate
adequate to fund the desegregation remedy.” The Court found
that the “difference between these two approaches is far more
than a matter of form.” Allowing local government to create its
own remedy “protects the function of those institutions” and
“places the responsibility for solutions . . . upon those who have
themselves created the problems.” Jenkins demonstrates that
the court in United States v. City of New York should have
more seriously contemplated its alternatives before resorting to
extensive oversight of the FDNY.

There is leeway for a district court to influence the
actions of local government, however, particularly when other
methods proved unsuccessful. In United States v. Yonkers Board
of Education, the Court of Appeals for the Second Circuit
addressed whether the district court abused its discretion by
refusing to adopt the City’s alternate proposal for remedying
unconstitutional housing segregation, and by appointing a
Housing Special Master. In rejecting the City’s contention, the
Second Circuit thought it particularly significant that it had
been eight years since the first order regarding remedial
measures was entered, and there was little progress to show for
it. The court also found that the City must come up with an
alternative that the district court will find to
“realistically . . . work now.” Otherwise, the district court
would not be obligated to consider those alternatives. With
regards to a Special Hiring Master, the court found that
although the City did not have final authority when
disagreements arose, it did have the ability to appeal those
decisions. Thus, the court found that the plan did not
unnecessarily restrict the local government. By contrast, in

---

65 Id.
66 Id.
67 Id.
68 Id.
69 Id.
70 29 F.3d 40 (2d Cir. 1994).
71 Id. at 44.
72 Id. at 43 (quoting Green v. Cnty. Sch. Bd. of New Kent Cnty., Va., 391 U.S. 430, 439 (1968)).
73 Id. at 44.
74 Id. Still, a court cannot simply refuse to entertain alternatives proposed by
a local government. See Schwartz v. Dolan, 86 F.3d 315 (2d Cir. 1996) (finding that the
United States v. City of New York, there was little demonstration that other alternatives did or would have failed, and it was apparent that the court heavily restricted the City’s affairs.74

2. Ricci v. DeStefano and the Ebbing Support for Race-Conscious Injunctions

The court in United States v. City of New York, by choosing not to enforce race-conscious hiring relief, was likely influenced by Ricci v. DeStefano.75 Ricci, while not explicitly contradicting or limiting race-conscious remedies—conveyed an attitude that Paradise-type relief may be falling out of favor. In Ricci, the Fire Department discarded a written exam after it had already been administered and the results submitted.76 Consequently, white firefighters from the City of New Haven, Connecticut, filed suit for disparate treatment discrimination.77 The Fire Department’s main defense was that it has been caught in a catch-22: in an effort to avoid discriminating against minorities it was forced to reject qualifying white applications.78

In a five-to-four decision, Justice Kennedy, writing for the Court, held in favor of the disparate treatment claim, stating the decision to throw out the exam was made “because of race.”79 The Court held that an employer must have a “strong basis in evidence” that use of the test will lead to disparate-impact liability, and the employer in Ricci did not meet that standard.80 Justice Scalia, concurring, made a prediction that disparate-impact claims may soon be extinct altogether.81 He argued that the majority’s decision only “postpones the evil day on which the Court will have to confront the question: Whether, or to what extent, are the disparate-impact provisions of Title VII of the Civil Rights Act of 1964 consistent with the Constitution’s guarantee of equal protection?”82 In furtherance of his argument, Scalia noted that the Equal Protection Clause of the Fourteenth Amendment prevents the

---

74 See infra Part V.
75 129 S. Ct. 2658 (2009).
76 Id. at 2664.
77 Id.
78 Id.
79 Id. at 2676 (citation omitted) (internal quotation marks omitted).
80 Id.
81 Id. at 2682-83 (Scalia, J., concurring).
82 Id. at 2682.
government from discriminating on the basis on race. In
addition, Scalia expressed his own distaste for disparate-
impact claims which, in his view, “place a racial thumb on the
scales, often requiring employers to evaluate the racial
outcomes of their policies, and to make decisions based on
(because of) those racial outcomes.”

Lower courts sympathetic to Scalia’s views might be
hesitant to implement any race-conscious injunctive relief for
fear of being viewed as placing a “racial thumb” on the scales.
The court in United States v. City of New York actually
attempted to separate itself from Ricci. There, the court had a
narrow interpretation of Ricci: whether an employer could
engage in disparate treatment, yet still avoid liability by
arguing it did so in order to avoid disparate impact liability.”
As a result, the court found that Ricci has no application—
including in United States v. City of New York—where the
question is whether the employment practice “actually had a
disparate impact.” While the court attempted to limit Ricci, its
aversion to race-conscious hiring measures strongly suggests
that the court sided with Scalia’s anti-affirmative action
sentiments.” As the next section demonstrates, however, the
court was at best premature to adopt such an outlook.

II. AFFIRMATIVE ACTION AND ITS CRITICS

Affirmative action has a complex history in U.S.
jurisprudence. Ever “since its inception in 1961, [it] has been
under siege.” It has been attacked on several fronts: Presidential
Executive Orders, theories of “color-blind[ness]” and “reverse
discrimination,” hostility exemplified in academic articles and
studies and state constitutional amendments banning
affirmative action." Barack Obama’s presidential election in 2008 made many believe that “the United States ha[d] officially entered a postracial era.”° Until this political movement directly and explicitly changes the broad range of Title VII remedies available, however, courts should not incorporate this sentiment into Title VII doctrine. In any event, the movement is ultimately flawed and should not be followed on its merits.

While the movement does appear to have had some recent success in implementing these types of statutory changes on the state level, there has not been significant change on the federal level. Ward Connerly, a black Republican political activist who “spearheaded” the proposals for state constitutional amendments, believes that “[w]ithout any doubt, we have to understand that race preferences are on the way out.” Connerly argues that “it does not bode well for a civilized society that professes to believe in equality to countenance treating its citizens differently based on traits over which they have no control as a result of acquiring them by reason of birth.”° Connerly also feels there are negative societal consequences in a society where “people can’t talk honestly about issues without somebody screaming about it. . . . [and] those who are the beneficiaries [of affirmative action] never want to let go. . . . [I]t becomes a crutch.”° On November 2, 2010, Arizona became the fifth state in the country to pass an anti-affirmative action amendment to its state constitution.°

Proponents of a postracial society, or “[c]olorblindness,” hope to place race in a “vacuum,” removing it from societal consciousness.°° They argue that “racism is irrational” since it is “unconnected from social reality” and is nothing more than mere “physical presence.”°°° Adopting this philosophy would

°° See, e.g., id. at 1277 & n.6.
°° Id. at 1277 & nn.7-8.
°° Id.
°°°°° Id. (quoting Neil Gotanda, A Critique of the Constitution Is Colorblind, 44 STANFORD L. REV. 1, 48 (1991)).
require courts to provide “a consistent and categorical application of the law” across racial lines.\textsuperscript{102} One goal of postracialists is to achieve a “pure meritocracy,” where one’s social standing is truly a result of inherent skill and perseverance, irrespective of race.\textsuperscript{103} Affirmative action arguably presents an “egregious affront” to this structure;\textsuperscript{104} by definition, affirmative action reduces the need for minorities to achieve, based on the traditional standards of merit.\textsuperscript{105} Because of this supposed imbalance, many proponents argue that affirmative action leads to unqualified candidates and leaves the “successful” minority candidates nevertheless stigmatized by their peers.\textsuperscript{106} Postracialists also reject race-based remedies because they believe that class-based injuries are often at the core of what we traditionally view as based on race.\textsuperscript{107} They also postulate that any advantages given to one race necessarily harms all others, resulting in a “zero-sum game.”\textsuperscript{108} Finally, they suggest that an imbalance is created when whites, who fear they will otherwise be confronted with “false accusations of racism,” decline to pursue reverse discrimination claims.\textsuperscript{109}

Although the movement has had some material gains in some states, their philosophy is ultimately flawed. At best, Connerly and his followers ignore how the implementation of a so-called postracial society will impact a nation that has focused on race throughout its history.\textsuperscript{110} At worst, postracialists actually help reinstate “a formal regime of White privilege.”\textsuperscript{111}

\textsuperscript{102} Id.
\textsuperscript{103} Id. at 487 (internal quotation marks omitted).
\textsuperscript{104} Id. at 488.
\textsuperscript{105} Especially when compared to their majority counterparts.
\textsuperscript{106} Beydoun, supra note 100, at 488; see also Tanya Kateri Hernandez, “Multiracial” Discourse: Racial Classifications in an Era of Color-Blind Jurisprudence, 57 Md. L. Rev. 97, 139-40 (1998).
\textsuperscript{108} Id.
\textsuperscript{109} Id. (citing Richard Thompson Ford, The Race Card: How Bluffing About Bias Makes Race Relations Worse 339 (2008)).
\textsuperscript{110} See generally Mario L. Barnes, Erwin Chemerinsky, & Trina Jones, A Post-Race Equal Protection?, 98 Geo. L.J. 967 (2010).
\textsuperscript{111} Beydoun, supra note 100, at 486. Critics have questioned the legitimacy and noted the hypocritical nature of the movement’s tactics. For example, Ward Connerly was likely chosen as a spokesperson against affirmative action “because of his identity as a Black man.” Id. at 489. In addition, critics note that Connerly receives a yearly salary of $1 million to perform these duties, suggesting that his motives may not necessarily be sincere. Id. Critics also note that the supporters of this movement largely stem from “a small group of wealthy and powerful rights [sic] wing corporate tycoons,” which suggests their motives may be in line with “trying to turn back the clock on civil rights . . . .” Id. at 490-91 (citing Lee Cokorinos, The Big Money Behind
First, a postracial society is simply too utopian an ideal to be successful in our current society. Second, even if a truly postracial society is possible, deep inequalities that are based on race still exist today.\textsuperscript{112} While postracialists point towards symbolically significant but nonetheless anecdotal evidence—for example, President Obama’s presidency—as proof of a postracial society,\textsuperscript{113} significant statistical disparities still exist across the nation that are attributable to race.\textsuperscript{114} In addition, the forms of discrimination have become “more subtle and covert,” not lending themselves to straightforward detection by the public or statistical analysis.\textsuperscript{115} Unconscious discrimination permeates our daily lives, especially during “the course of reasoned evaluations.”\textsuperscript{116} Additionally, colorblindness, or “race-neutral universalism,” is rarely applied universally and tends to benefit nonminorities disproportionately.\textsuperscript{117} Finally, the idea that we have entered a postracial era has been raised many times before—practically every time minorities have made significant strides—and those proclamations have almost universally been discredited.\textsuperscript{118} There is nothing to suggest that we have now developed the awareness and foresight necessary to determine whether race has become a nonissue.

In sum, the politics of today need not erode the purposes and goals of remedies developed under federal and state law. While the tides of public opinion may have swayed as of late,

\footnotesize
\begin{itemize}
  \item \textsuperscript{113} See Barnes et al., supra note 110, at 981-82.
  \item \textsuperscript{114} See id. at 982-91 (noting that significant statistical disparities still exist in terms of poverty, income and wealth, homeownership, employment, education, and criminal justice statistics); see also Pager, supra note 112, at 107 (noting that “blacks, and young black men in particular, have become increasingly likely to drop out of the labor market altogether when faced with the prospect of long-term unemployment or marginal employment opportunities”).
  \item \textsuperscript{115} Pager, supra note 112, at 105 (“It could be the case . . . that discrimination remains fairly routine in certain contexts, despite infrequent public exposure.”).
  \item \textsuperscript{116} Id. at 108 (citations omitted).
  \item \textsuperscript{117} Cho, supra note 107, at 1602 (citing john a. powell, Post-Racialism or Targeted Universalism?, 86 DENV. U. L. REV. 785, 789-98 (2009)). For example, model recipients for the “G.I. Bill, welfare, and social security” are “white, able-bodied, and male.” Id. (citing powell, supra, at 794-98).
  \item \textsuperscript{118} See Barnes et al., supra note 110, at 972 (“Almost from the moment the Civil War ended and the Thirteenth Amendment abolished slavery, there were declarations that the United States had moved beyond race.”).
\end{itemize}
there is nothing to suggest that courts—particularly federal courts and the great majority of state courts where affirmative action is still permitted—should lessen their resolve to provide race-conscious remedies. These remedies help achieve the very thing postracialists believe we already have—a society in which one’s race does not hamper one’s equal opportunity to succeed.

III. HISTORY OF RACE DISCRIMINATION IN THE FDNY

The long history of racial discrimination in the FDNY only emphasizes the need to implement impactful race-conscious hiring in United States v. City of New York. In the 1920s, Black firefighters needed strong resolve to last in the Fire Department. They lived—and literally slept—with reminders that they were not welcomed. For instance, while on overnight duty they were assigned to a “black bed” at the firehouse, “which whites refused to use even when no black firefighters were on duty.” Being a New York City Firefighter “tends to be a family business.” As a result, black and Hispanic citizens do not have the same access or encouragement to become part of the Fire Department.

Black and Hispanic firefighters, however, were not without their champions. In 1940, Wesley Williams, the third black firefighter in the Fire Department, helped establish the Vulcan Society in an effort to draw the black-firefighter community together. By 1944, the Vulcan Society had successfully lobbied for regulations that banned racial discrimination—at least officially—within the Fire Department.

By the 1970s, overt racial discrimination had been supplanted by more subtle, yet equally virulent, racial discrimination. A typical experience for a black firefighter at this time was akin to that of “being in Archie Bunker’s living room.” They often found “racial epithets written on the

---

122 History of the Vulcan Society, supra note 121; see also GOLWAY, supra note 119, at 204.
123 History of the Vulcan Society, supra note 121; see also GOLWAY, supra note 119, at 204. Although, “the black bed . . . persisted into the immediate postwar years . . . .” Id. at 205.
124 GOLWAY, supra note 119, at 277 (internal quotation marks omitted).
firehouse blackboard,” and it was apparent that their “white colleagues were not particularly fond” of the “black neighborhoods.” In fact, in 1973 the United States District Court for the Southern District of New York found that the Fire Department discriminated against minorities, holding that the City’s entry-level exams had a disparate impact on minority firefighters.

Minority representation in the FDNY has not appreciably improved since 1973. In that year, only 5 percent of blacks and Hispanics made up the Fire Department while making up 32 percent of the general population. Between 1991 and 2007, the percentage of black firefighters never exceeded 3.9 percent, and by 2007, had dropped to 3.4 percent. In a 1999 census of minority firefighter representation in large American cities, New York ranked at the bottom of the list; New York minority firefighters were only one-tenth as represented in the FDNY relative to the city’s general population. Not surprisingly, a discrimination suit was again brought in 2007.

IV. BACKGROUND AND ORDERS ISSUED THUS FAR IN UNITED STATES V. CITY OF NEW YORK

A. Facts and Prior History

In 2007, the United States, followed by the Vulcan Society of the New York Fire Department and individual plaintiffs, filed suit in District Court within the Eastern District of New York alleging that the Fire Department’s hiring practices violated Title VII of the Civil Rights Act of 1964. The focus of the complaint concerned two written exams that ranked prospective entry-level firefighters and significantly impacted hiring. Plaintiffs argued that the tests had a disparate impact on black

125 Id.
126 Id.
127 For a general discussion on the theory of disparate impact, see supra Part I.A.
130 Id.
131 Id. at 242 (citations omitted).
132 Complaint ¶ 1, City of New York, No. 07-CV-2067 (E.D.N.Y. May 21, 2007), ECF No. 1; Intervenors’ Complaint ¶ 1, City of New York, No. 07-CV-2067 (E.D.N.Y. July 17, 2007), 2007 WL 3117053, ECF No. 48.
133 Complaint ¶ 1, City of New York, No. 07-CV-2067 (E.D.N.Y. May 21, 2007), ECF No. 1.
and Hispanic applicants. The Vulcan Society also argued that the continued racial disparities in the Fire Department amounted to disparate treatment and that defendants “have been long-aware of the discriminatory impact” their examination process has had on blacks. The Vulcan Society alleged that defendants “continued relying on and perpetuating . . . these racially discriminatory hiring processes.”

The entry-level exams for New York City firefighters have been under scrutiny for some time. In 1973, the city’s entry-level exams were similarly found to have had a disparate impact on minority firefighters. In fact, the court in United States v. City of New York found the testing procedures used during the 1973 decision to be “strikingly similar to the testing procedures in this case.” In addition, the defendants in the 1973 decision “failed to . . . demonstrate[ ] that the examination was job-related,” which had “the quality of déjà vu” when compared to the current action. The 1973 court ordered the City to hire one-third minority applicants until they were able to develop a new, nondiscriminatory test. The City also hired a private consulting firm to help develop the tests.

Only three years later, however, the City abandoned their relationship with the private consulting firm purportedly because of a “fiscal crisis.” The City also failed to follow the minority hiring requirement and instead “instituted a hiring procedure that required . . . minimum appointment requirements such as college credits, a driver’s license, and certified first responder with defibrillation training.” The 2012 court found that any effect the 1973 decision had on minority hiring “constituted little more than a brief departure from an otherwise relentless pattern” of discrimination.

---

134 Id.
137 See City of New York, 683 F. Supp. 2d 225, 238 (E.D.N.Y. 2010) (citing Vulcan Soc’y of N.Y.C. Fire Dep’t, Inc., 360 F. Supp. 1265, 1269 (S.D.N.Y. 1973)). The court notes this case “furnishes proof of an old adage: the more things change, the more they remain the same.” Id.
138 Id. at 239.
139 Id. at 240.
140 Id. (citing Berkman v. City of New York, 536 F. Supp. 177, 184 n.1 (E.D.N.Y. 1982)).
141 Id. (citing Berkman, 536 F. Supp. at 184).
142 Id. (citation omitted).
143 Id.
144 Id.
B. Disparate Impact Decision

On July 22, 2009, the court in United States v. City of New York found that the entry-level exams—Written Examinations 7029 and 2043—had a disparate impact on black and Hispanic applicants in violation of Title VII of the Civil Rights Act of 1964. 145

Both exams were “an 85-question, paper-and-pencil multiple choice test” administered from 1999 through 2002 and from 2002 through 2007, respectively. 146 Passing either exam was a prerequisite for taking the physical performance test (PPT). 147 The applicants were ranked in order by combining the results of both the written and physical tests. 148 About 1750 black applicants and 2125 Hispanic applicants took exam 7029. 149 Only 104 (3.2 percent) blacks were hired from those who took the exam. 150 Also, black applicants passed the exam only 67 percent as often as white applicants. 151 This “disparity [amounted] to 33.9 units of standard deviation,” 152 which means the probability the disparity “occurred by chance [was] less than 1 in 4.5 million-billion.” 153 In addition, only 274 (8.5 percent) Hispanics who took the exam were hired. 154 Hispanic candidates passed the exam 85.3 percent as often as white candidates. 155 “[T]his disparity is equivalent to 17.4 units of standard deviation, meaning that the likelihood it occurred by chance is less than 1 in 4.5 million-billion.” 156 A similar set of statistics was present for Exam 2043. 157

The court found that there were no “material factual disputes sufficient to preclude summary judgment on job-relatedness.” 158 The court stated that “the City tested for tasks and abilities that could be learned on the job,” 159 rather than

145 City of New York, 637 F. Supp. 2d 77, 82-83 (E.D.N.Y. 2009).
146 Id. at 84-85.
147 Id. at 85.
148 Id.
149 Id. at 86.
150 Id.
151 Id. at 88.
152 Id.
153 Id. (citations omitted).
154 Id. at 86.
155 Id. at 88-89.
156 Id. (citations omitted).
157 See id. at 89-90.
158 Id. at 110.
159 Id. at 113 (citations omitted).
measuring tasks and abilities that “are needed at entry.” In addition, the court found that “the City [did] not offer[ ] any evidence of a competent test construction process . . . . Instead, . . . . the City appear[ed] to be relying on the same practices for which it was criticized by the Second Circuit thirty years ago.” The court also found the City “fail[ed] to test various cognitive and non-cognitive abilities . . . [and] fail[ed] to show that the examinations had an appropriate reading level.” Finally, the court found that a reasonable fact-finder could not conclude that exams had the “reliability or validity” necessary to refute the statistical claims. In sum, the court found that the exams—in which a disproportionate number of black and Hispanics had failed—did not adequately predict future performance. Therefore, the court found that the exams had a disparate impact on black and Hispanic applicants.

C. Disparate Treatment Decision

On January 13, 2010, the court also found that the City engaged in systematic disparate treatment. The court stated the exams were “part of a pattern, practice, and policy of intentional discrimination against black applicants that has deep historical antecedents and uniquely disabling effects.” The plaintiffs were able to establish their prima facie case primarily through “undisputed statistical and anecdotal evidence.” The large standard deviation in racial disparity—from 10.5 to 33.9 units—made the statistical evidence legally significant, and the statistical evidence was “supplemented . . . [by] showing with extensive historical, anecdotal, and testimonial evidence that intentional discrimination was the City’s ‘standard operating procedure.’”

The City made no attempt to show that the plaintiffs’ proof was either “inaccurate or insignificant” by questioning its

---

160 Id. (citations omitted).
161 Id. at 116.
162 Id. at 123.
163 Id. at 131.
164 Id.
166 Id. at 273.
167 Id. at 251.
168 Id. at 236.
169 Id. at 250 (quoting Int’l Bhd. of Teamsters v. United States, 431 U.S. 324, 336 (1977)).
“source, accuracy, or probative force.” Instead, the City attempted (and failed) to show that it had no “subjective intent to discriminate.” Therefore, the court found that the defendants engaged in disparate treatment based on a pattern and practice of discrimination and left the form of relief to the remedial stage.

D. Preliminary Injunction Decision

On June 29, 2010, the City stated that it “expect[ed] to hire approximately 300 new firefighters” from a new test, “Exam 6019.” Plaintiffs were able to demonstrate, however, that this new exam would still produce a disparity between black and white applicants’ pass rates—estimated at 22.70 units of standard deviation. The plaintiffs also demonstrated that Hispanics and white applicants’ pass rates would be separated by 11.35 units of standard deviation. The court was persuaded that this potential disparity was more than sufficient to show that a prima facie case would be met if litigation was brought regarding the new exam.

The court also found that the City would be unable to raise a proper business necessity defense, since it still could not successfully demonstrate that a higher test score, by as much as three points, detected any marked increase in an applicant’s actual skill as an entry-level firefighter. In addition, the City failed to demonstrate that the new exam was “content valid.” As a result, the court enjoined any further hiring by the City under the new exam until October 1, 2010, giving the court time to develop an interim hiring plan that would not violate federal or state discrimination laws.

170 Id. at 253-55 (quoting Robinson v. Metro-North Commuter R.R., 267 F.3d 147, 158 n.4 (2d Cir. 2001)).
171 Id. at 251.
172 Id. at 255.
173 Id. at 273.
174 Letter from James M. Lemonedes, City of New York, No. 07-CV-2067 (E.D.N.Y. June 29, 2010), ECF No. 456.
176 Id.
177 Id. at 301.
178 Id. at 315.
179 That is, the exam failed to test specific knowledge necessary for entry-level firefighters to possess. Id.
180 Id.
E. Interim Hiring Decision

On September 13, 2010, the court offered several hiring alternatives from which the City was to choose, most or all of which contained race-conscious adjustments to the Exam 6019 rankings. The court emphasized it had the power to issue broad relief to combat past discrimination; “the district court ‘has not merely the power but the duty’ to ‘bar like discrimination in the future.’” The proposals included variations of “random selection” procedures, where a specified pool of qualified applicants (avoiding the lowest-scoring candidates) would be hired at random, as well as “applicant flow” procedures.

The “random selection” proposals included race-conscious adjustments. For example, one proposal allowed the white applicants who were ranked at the bottom 2500 candidates to be replaced by minority candidates who were just below the 2500 mark, and this new pool would be used to select candidates at random. While the court admitted that “the rank-adjustment proposal [was] a race-conscious compliance measure,” the court had no doubt that the proposal was lawful. Indeed, “[a]n unbroken string of Supreme Court and Second Circuit case law confirm[ed] that race-conscious remedial compliance measures are permissible under Title VII.”

The “applicant flow” procedures arguably had a more blatant race-conscious component; the City could hire using “any criteria [it] desired” as long as it did so in proportion to the minority representation of the applicant pool. The court recognized that “these proposals [also] strongly resemble[d]
racial hiring quotas,” but believed that precedent allowed this procedure.\footnote{189} The court was partially persuaded by the fact that the City proclaimed its “current fiscal condition may ... be such that it cannot reasonably bear the costs of further firefighter hiring delays.”\footnote{190} A major advantage of the court-proposed interim hiring procedures was that they could be implemented in a matter of weeks rather than months or even years.\footnote{191}

\section*{F. The City’s Rejection of Interim Hiring and the Court’s Response}

On September 17, 2010, the City enigmatically ignored all of its previous financial and safety concerns. In a letter to the presiding judge, the City stated that “[e]very one of the five proposals from which the Court [was] allowing the city to select involve[d] some form of race-based quota.”\footnote{192} The defendants recognized that, as a consequence, the City would not be “permitted to hire any entry level firefighters for the duration of the ‘temporary injunction.'”\footnote{193} The City would simply try to make a valid examination “as expeditiously as possible.”\footnote{194}

The court—with unambiguous distaste for the City’s tactics—“permanently enjoin[ed] the City from hiring firefighters based on the results of Exam 6019.”\footnote{195} The court found that the City’s refusal to follow the court’s order was “compelling evidence that enjoining the City from hiring off that test, except according to one of the Hiring Options, would not unduly burden the City ... .”\footnote{196}

Thus, the court did not require the City to implement any of the offered interim hiring procedures.\footnote{197} Instead, the

\footnotesize
\begin{itemize}
\item \footnote{189} Id. at *14 (citing \textit{Local 28 of Sheet Metal Workers’ Int’l Ass’n}, 478 U.S. at 450-51, 464; \textit{Paradise}, 480 U.S. at 166; \textit{Guardians Ass’n of the N.Y.C. Police Dep’t, Inc. v. Civil Serv. Comm’n}, 630 F.2d 79, 109 (1980)).
\item \footnote{190} Id. at *15.
\item \footnote{191} Id.
\item \footnote{192} Letter from Michael A. Cardozo, \textit{City of New York}, No. 07-CV-2067 (E.D.N.Y. Sept. 17, 2010), ECF No. 532.
\item \footnote{193} Id.
\item \footnote{194} Id.
\item \footnote{195} Id.
\item \footnote{196} Id. at *2.
\item \footnote{197} Id. at *8.
\item \footnote{198} Id. at *2.
court enjoined any immediate hiring and gave the City the option to change its mind at any time, leaving the plaintiffs, and everyone else who spent a significant amount of time preparing and taking entry-level exams 7029 and 2043, without any legitimate hiring opportunities in the interim.\textsuperscript{199} Although the court was aware that the City’s “shifting and contradictory positions”\textsuperscript{200} were “simply the latest episode in the City’s long campaign to avoid responsibility for discrimination in its Fire Department, whatever the cost,”\textsuperscript{201} it left the City to come up with its own nondiscriminatory procedure. The court chose not to enforce any of its temporary hiring procedures despite the fact that even the City “now implicitly accept[ed] that the court has the authority to order quotas in the appropriate circumstances.”\textsuperscript{202}

G. Subsequent Motion by the Plaintiffs for Compensatory and Injunctive Relief

On December 19, 2010, the plaintiffs filed a motion for a proposed order for injunctive and monetary relief.\textsuperscript{203} The plaintiffs requested that the court appoint a “Special Monitor to oversee compliance with the Court’s Order.”\textsuperscript{204} The plaintiffs also requested specific injunctive relief, including a requirement that the Fire Department administer an entry-level exam every two years rather than every four.\textsuperscript{205} This frequency would alleviate the fact that blacks and Hispanics are less likely to be aware of the exam or exam date.\textsuperscript{206} In addition, the plaintiffs moved for the court to order the Fire Department to enhance their minority recruitment and publicity programs.\textsuperscript{207} In doing so, the Fire Department should “ensure that blacks, Hispanics, and whites are equally informed about employment opportunities . . . .”\textsuperscript{208} The goal

\begin{itemize}
\item \textsuperscript{199} See id.
\item \textsuperscript{200} Id. at *1.
\item \textsuperscript{201} Id.
\item \textsuperscript{202} Id. at *7.
\item \textsuperscript{203} Memorandum of Law in Support of Plaintiffs-Intervenors’ Proposed Order for Injunctive Relief, City of New York, No. 07-CV-2067 (E.D.N.Y. Dec. 9, 2010), 2010 WL 6917514, ECF No. 596.
\item \textsuperscript{204} Id. at 2, 6 (arguing that “District Courts have the authority to appoint a Special Master or Monitor when broad injunctive relief is ordered” (citing United States v. Yonkers Bd. of Educ., 29 F.3d 40, 44 (2d Cir. 1994))).
\item \textsuperscript{205} Id. at 7-8.
\item \textsuperscript{206} Id. at 8.
\item \textsuperscript{207} Id. at 8-9.
\item \textsuperscript{208} Id. at 9.
\end{itemize}
would be to have an applicant pool that is “representative of the racial and ethnic makeup of the age-eligible residents of New York City.” Such a plan included retaining an expert recruitment consultant, employing thirty full-time recruiters, and providing for an adequate budget. The plan would also include quality test-preparation courses that are “[f]ree and [a]ccessible” to minority applicants.

The plaintiffs also sought to implement changes to the postexam candidate screening process. The plaintiffs noted that minority candidates who were eligible for consideration were screened out “at much higher rates than white candidates.” This disparity occurred because the Fire Department often failed to successfully notify minority candidates about their “initial candidate screening intake interview,” and others were disproportionately scrutinized for “arrests that did not result in convictions.” In addition, the plaintiffs requested that the Fire Department should have a system in place for monitoring and preventing retaliation and workplace discrimination against minority firefighters.

H. The Court Finds Continued Supervision Necessary to Eliminate Recruiting and Hiring Racial Discrimination in the Fire Department

On September 30, 2011, the court issued a Memorandum and Findings of Fact highlighting many of the issues that remain with regards to the Fire Department’s hiring and recruiting processes, creating a strong impression that the entry-level exams only reach the surface of the Fire

209 Id.
210 Id. at 12-14.
211 Id. at 15.
212 Id. at 16-17.
213 Id. at 17.
214 Id. at 17-18 (noting that minorities were generally harder to reach because their addresses changed more often, but little was done to try to ensure notice when it was apparent the addresses were incorrect).
215 Id. at 17; see also Dave Saltonstall, Bravest’s Hiring Under Fire Minorities Seek Larger Numbers on Force That’s 93% White, DAILY NEWS (N.Y.C.), May 2, 1999 (noting that family members of firefighters receive help with “training and surviving the required background checks that minority firefighters say are often used to keep them out of the department. Such checks . . . have snared black candidates for jumping subway turnstiles as kids, while whites with more troubling records are allowed to pass.”).
Department’s discriminatory practices.  

The court found the need for an attrition-mitigation plan for its entry-level hiring process. The elongated hiring process made black candidates 40 percent more likely than white candidates to eventually be disqualified. The court noted several other areas of concern that needed to be addressed, including setting measurable goals and ensuring continued support.

In essence, the court found the City unwilling to address the issue and thus felt the need to impose a structure for remedial measures. Indeed, the court felt that if the City had “shown the least bit of concern . . . this would be a much different order.” The court based this conclusion in part on the amount of time that had passed in this litigation—four years—rather than the amount of time that an actual plan has been implemented, which had not yet occurred. Thus, although the remedial phase of the litigation was in its opening stages, the City was not given an opportunity to implement its own plan.

The Order itself called for extensive oversight by multiple independent parties, and required the City to explain itself when it disagreed with the independent parties. For example, an independent Court Monitor was appointed, and the City could not take “any step in any process for the selection of entry-level firefighters . . . without first obtaining the approval of the Court Monitor.” In addition, the City must “retain an independent recruitment consultant . . . subject to the approval of the Court Monitor,” who, among other things, “identifies best practices for the recruitment of black and Hispanic” applicants. The City must also come up with a plan to prevent “voluntary” attrition.

The Order also imposes extensive record-keeping and document retention measures and gives the Court Monitor broad authority to request on “short notice” any document or investigation of any individual it deems relevant. Any employee, including unpaid interns, who are “formally or
informally” involved in the hiring process of an entry-level firefighter must “immediately create a written record of all oral communications in which they are involved” relating to such conversations.\footnote{Id. at *8.} In addition, the City must appoint an independent EEO consultant, who will produce a report that, among other things, “identifies all tasks the EEO Office should be performing to ensure the FDNY’s compliance with applicable equal employment opportunity laws.”\footnote{Id. at *11.} Most notably, the Order is enforceable for up to ten years, unless upon the second civil service hiring list, the City is essentially able to show that there currently is no disparate impact or treatment, and there is no reason to believe that such discrimination would present itself in the near future.\footnote{Id. at *17-18.}

The court explicitly refused to entertain alternatives to this extensive oversight. Indeed, despite the court’s previous attempt to implement race-conscious hiring measures, it emphasized that the Order “does not impose hiring quotas in any shape or form,” and that “this court has never endorsed hiring quotas . . . it does not believe quotas would be an effective remedy to the City’s discrimination.”\footnote{Memorandum [and] Findings of Fact at 20, City of New York, 07-CV-2067 (E.D.N.Y. Sept. 30, 2011), ECF No. 741.} From this context, it is apparent that the decision to impose extensive oversight so early in the remedial phase was due to the court’s unjustified aversion to race-conscious hiring.

V. \textit{UNITED STATES V. CITY OF NEW YORK} PROPOSED AN INSUFFICIENT REMEDIAL PLAN AND PREMATURELY IMPOSED RESTRAINTS ON LOCAL GOVERNMENT

In \textit{United States v. City of New York}, the court initially failed to fully exercise its duty to remedy long-institutionalized racial disparity. Subsequently, the court overstepped its bounds by not respecting local government independence over federal courts. When “proposing” its interim plan for relief, the court should have created a plan that required a greater ratio of race-conscious hiring and should have imposed it despite the City’s objections. Instead, after the City rejected the interim plan, the court severely limited the City’s control over its own hiring practices by imposing extensive oversight. While there is
potential for the current plan to address many of the deep-seated racial issues in the FDNY, it unnecessarily does so at the expense of local government's control over its own affairs.

A. The Interim Relief of Race-Conscious Hiring Practices

When the court offered several interim hiring plans from which the City could choose (but ultimately rejected), the plans should have been more expansive in order for them to effectively eradicate racial discrimination. In addition, it was within the court's power to issue a preliminary injunction ordering, rather than asking, that the City follow these proposed hiring procedures.\(^{231}\)

1. The Proposed Hiring Procedures Should Have Required a Greater Ratio of Minority Firefighters

Under each proposal, the number of black and Hispanic candidates passing the test must have been a “Representative Pool”\(^ {232}\) of all the applicants. In other words, the percentage of black and Hispanics who pass the entry-level exam should be equal to the percentage of black and Hispanic applicants who take the exam. On its face, this proposal would provide a measure of remedy for the particular applicants who had brought a claim. The proposal fails, however, to adequately address the extensive discriminatory effects of the Fire Department’s actions in two respects.

First, requiring that the Fire Department hire a “Representative Pool” of the applicants fails to provide relief to the greater minority population because the number of minority applicants was insufficient in itself. Minorities have understandably been discouraged from applying to the Fire Department after witnessing decades of minority applicants who fell short of success.\(^ {233}\) In order to address the long history of discrimination, a greater minority hiring ratio—that of the age-eligible local population—is required.\(^ {234}\) This ratio would

---

\(^{231}\) See supra notes 46-60 and accompanying text; United States v. Paradise, 480 U.S. 149, 180 (1987).

\(^{232}\) City of New York, No. 07-CV-2067, 2010 WL 3709350, at *4 (E.D.N.Y. Sept. 13, 2010) (defining “Representative Pool” to mean “that the subgroup’s racial demographics reflect the racial demographics of the entire applicant pool”).

\(^{233}\) See supra Part III.

\(^{234}\) See Memorandum of Law in Support of Plaintiffs-Intervenors’ Proposed Order for Injunctive Relief at 7-12, City of New York, No. 07-CV-2067, 2010 WL 6917514 (E.D.N.Y. Dec. 9, 2010).
better represent the potential applicants that would have applied had the City not engaged in disparate treatment, which in turn would help alleviate the long-term effects of discrimination.

Second, even considering the court’s modest goal of reaching the ratio of minority applicants, it will take a significant (if not an indefinite) amount of time for the entire Fire Department—rather than just the newest, entry-level firefighters—to reach the desired minority ratio.\textsuperscript{235} In essence, most nonminority firefighters hired before this plan would have to retire or resign in order for the appropriate ratio to be reached. In addition, considering that this measure, if implemented, would have been only temporary, minority representation would not have significantly increased during that limited time. As a result, the ultimate goal of the plan would not have been reached since the City would have had little incentive to promptly come up with an examination procedure that lacked disparate impact.

Indeed, the “relevant population or work force,” rather than the relevant applicant pool, is traditionally used as the standard ratio for race-conscious hiring.\textsuperscript{236} It is within the court’s scope to use the relevant population or work force as a point of comparison, since “eradicat[ing] race discrimination may include reasonable race-conscious relief that benefits individuals who were not actual victims of discrimination.”\textsuperscript{237} This type of ratio would do much to address the long history of discrimination in the Fire Department. It would ensure that

\textsuperscript{235} See \textit{Paradise}, 480 U.S. at 180 (holding that a “one-for-one” promotion requirement would “have determined how quickly the Department progressed toward [the] ultimate goal”). There, the Court found that “[s]ome promptness in the administration of relief was plainly justified” where the previous “use of deadlines or end dates had proved ineffective.” \textit{Id.} In that case, the implementation of a “one-for-one” ratio, where it was determined that the labor pool was twenty-five percent black, “was crafted and applied flexibly, [and] was constitutionally permissible.” \textit{Id.}

\textsuperscript{236} \textit{Id.} at 187. \textit{See, e.g.}, \textit{Firefighters Local Union No. 1784 v. Stotts, 467 U.S. 561, 565 (1984)} (where the employer agreed to “adopt[] the long-term goal of increasing the proportion of minority representation in each job classification in the Fire Department to approximate[] the proportion of blacks in the labor force” (emphasis added)); \textit{United Steelworkers v. Weber}, 443 U.S. 193, 199 (1979) (upholding an agreement that “50% of the new trainees were to be black until the percentage of black skilled craftworkers in the Gramercy plant approximated the percentage of blacks in the local labor force” (emphasis added)).

\textsuperscript{237} \textit{Local Number 93, Int’l Ass’n of Firefighters v. City of Cleveland, 478 U.S. 501, 516 (1986)}; \textit{see also Paradise}, 480 U.S. at 184 (recognizing that “the choice of remedies to redress racial discrimination is ‘a balancing process left, within appropriate constitutional or statutory limits, to the sound discretion of the trial court’” (citing \textit{Fullilove v. Klutznick}, 448 U.S. 448, 508 (1980) (Powell, J., concurring))).
integration of minority firefighters would occur at a sufficient pace. In addition, the rate of integration would remain consistent even if minority applicants continued to be discouraged from becoming firefighters.\footnote{238}

The greater ratio also helps prevent future disparate impact discrimination, which was one of the court’s main goals.\footnote{239} If the minority applicant ratio were applied, the Fire Department would be incentivized to dissuade minorities from applying in the first place since it would reduce the impact of the interim hiring plan. If the Fire Department engaged in this activity, however, it would create an acute possibility of future disparate impact and treatment liability.\footnote{240} On the other hand, a hiring ratio based on the general population would eliminate that incentive. Indeed, the general population ratio would actually dissuade the City from such a practice because it would only serve to reduce the pool of minority firefighters from which they could choose—thus lowering the chances that the new hires would be the highest qualified.

This higher ratio would not “trammel the rights” of white employees,\footnote{241} nor would it put the qualifications of the accepted applicants into question. The court’s reasoning for interim relief still holds in this respect.\footnote{242} There, the court felt the hiring procedures were not “likely to upset the expectations of the Qualified Candidates, white and minority alike” because “a high percentage of the [so-called] Qualified Candidates will be appointed to one of the next two firefighter classes.”\footnote{243} By ensuring that all the candidates are “qualified”\footnote{244} and that

\footnotesize\hspace{1\textwidth}

\footnoterule

\footnotesize\textsuperscript{238} A hiring ratio based on the number of applicants might incentivize the Fire Department to discourage minorities from applying; reducing the pool of minority applicants would in turn reduce the number of minority entry-level firefighters they would have to hire. This tactic would have no impact, however, on a hiring ratio based on minority representation in the local population.

\footnotesize\textsuperscript{239} See supra Part I.B; Albemarle Paper Co. v. Moody, 422 U.S. 405, 418 (1975).

\footnotesize\textsuperscript{240} Plaintiffs could bring a disparate treatment or disparate impact claim against the Fire Department regarding its recruitment of minority applicants. Even if the Fire Department ultimately hired a sufficient ratio of minorities, this event would not preclude a disparate impact claim because the “bottom line defense” is not a legitimate defense against disparate impact claims. See Connecticut v. Teal, 457 U.S. 440, 442 (1982).

\footnotesize\textsuperscript{241} See Paradise, 480 U.S. at 183 (noting that any promotion requirement would give an advantage to black applicants, the “situation is only temporary, and is subject to amelioration by the action of the Department itself”).


\footnotesize\textsuperscript{243} Id.

\footnotesize\textsuperscript{244} Paradise, 480 U.S. at 183 (“[T]he basic limitation, that black troopers promoted must be qualified, remains. Qualified white candidates simply have to compete with qualified black candidates.”).
white employees do not face an “absolute bar” from employment, the court would be able to institute temporary relief that allows for a greater ratio of black and Hispanic workers to be hired.

In addition, such a hiring procedure would actually help keep race-conscious relief a temporary measure. Because the hiring would create a substantial and predictable inflow of minority firefighters, the procedure would have a substantial effect—enough to encourage the City to develop a nondiscriminatory test that the court could accept. The incentives to develop an acceptable test would persist even if the City’s intentions for doing so were dubious. For example, because a nondiscriminating test does not necessarily mean that the ratio of the eligible population will be hired, the City—in an effort to reduce minority integration—may calculate that developing their own court-approved exam will actually slow the integration of minority firefighters.

The City opposed hiring blacks at a ratio matching the percentage of blacks that are age-eligible within the City. The City first argued that some of these members will inevitably “self select out because they know they do not meet some of the unassailably objective criteria for being a firefighter.” This argument is at best incomplete, because it does not provide any legitimate reason why potential black applicants would “select out” at a higher rate than white applicants. The City also suggests that “[n]o matter how effective the FDNY’s recruitment efforts are, and no matter how positively an FDNY career is viewed in diverse communities, at the end of the day firefighting will not be everyone’s career choice.” This statement implies that blacks are simply not as interested in becoming firefighters. However, such an argument must be

---

245 See id. at 182 (stating that an “absolute bar” to white advancement” does not exist where “50% of those elevated were white” and the procedure “does not require the layoff and discharge of white employees” (citing Wygant v. Jackson Bd. of Educ., 476 U.S. 267, 283 (1986))). The Court in Wygant notes that “layoffs impose the entire burden of achieving racial equality on particular individuals, often resulting in serious disruption of their lives.” 476 U.S. at 283.

246 The disparity will generally need to be between two to three standard deviations away from the eligible population for a disparate impact claim to be successful, which leaves room for the “neutral” test to have results that are somewhat below the ratio of minorities in the relevant population. See supra note 19 and accompanying text.

247 Defendants’ Memorandum of Law in Opposition to Plaintiffs-Intervenors’ Motion for a Proposed Order for Injunctive Relief at 18, City of New York, No. 07-CV-2067 (E.D.N.Y. Dec. 9, 2010), 2010 WL 6917516, ECF No. 599.

248 Id.
supported by strong evidence that is not based on stereotypical belief.\textsuperscript{249} Here, not only does the City fail to provide such evidence, but it also fails to rebut the presumption that many minorities have been discouraged specifically because of the intentional discrimination\textsuperscript{250} that has existed in the Fire Department for decades.\textsuperscript{251}

2. The Court Should Have Required Implementation of an Interim Hiring Procedure Rather Than Issue a Permanent Injunction and a Subsequent Intensive Oversight Program

A race-conscious hiring procedure (using either ratio) would have been preferable over the court’s permanent injunction because of the injunction’s effect upon third-party bystanders. The injunction forced all the applicants who took entry-level exams 7029 and 2043 to bear the costs of the finding of discrimination. None of the applicants who took these exams will be able to begin employment for an indefinite period of time. Even if a hiring procedure came close to “trammell[ing] the rights” of white employees\textsuperscript{252}—for example, by hiring little or no white firefighters for a specified time period—there would still be a substantial benefit in allowing the current plaintiffs some form of immediate relief. The permanent injunction, on the other hand, leaves all the applicants with no prospects of employment, as now they all must wait for the City to develop an adequate exam.\textsuperscript{253} If the court intended to use this delay to allow the City to work on a more effective plan, it could have instead simply implemented a temporary race-conscious hiring procedure and let the City use the subsequent time to come up with a more permanent solution.

In addition, the injunction did not create sufficient incentive for the City to develop a nondiscriminatory exam. If the court does not accept the City’s new proposals, the court

\textsuperscript{249} EEOC v. O & G Spring & Wire Forms Specialty Co., 38 F.3d 872, 876-77 (7th Cir. 1994) (an employer cannot simply assert that the racial disparity is based on lack of interest without providing closely scrutinized evidence); cf. EEOC v. Sears, Roebuck & Co., 839 F.2d 302, 313-14 (1988) (employer successfully argued with extensive evidence that women were not interested in becoming salespersons).

\textsuperscript{250} The court acknowledges that such discrimination exists. See City of New York, 683 F. Supp. 2d 225, 241 (E.D.N.Y. 2010).

\textsuperscript{251} See supra Part III.

\textsuperscript{252} See supra note 241 and accompanying text.

would be again forced to offer interim race-conscious hiring procedures before the City could commence hiring. Given the fact that the court’s proposals as constituted would not have a significant impact over time, the consequences for the City are small if it does not produce an adequate exam.

B. The Court Overzealously and Prematurely Imposed Restraints on Local Government

Despite the apparent difficulty the court faced in getting the City to cooperate with its interim hiring plan, the court was overzealous and premature in developing an extensive oversight system. While the plan allows the City to have its proposals heard before the court and provides for the ability to take appeal—and thus does not on its face violate the limits on local government outlined in Missouri v. Jenkins and Schwartz v. Dolan—the plan was implemented too soon and puts into question the City’s practical ability to effect such proposals.

The limits on injunctive relief set forth in Paradise shed light on this plan. A significant factor in determining whether injunctive relief is too broad is the planned duration of the remedy. Here, the plan could last as long as ten years. Admittedly, the length of the plan is somewhat “contingent upon the [City’s] own conduct,” as it was in Paradise; the court could relinquish its control if the City were able to meet certain goals. The goals the City must reach, however, are unrealistic. For example, the City must prove, by a preponderance of the evidence, that it does not and will not use “any examination that in any way results in a disparate impact” and must be job-related or required by business necessity. Even with earnest intentions and generous resources, no amount of evidence can definitively pronounce that future examinations will not have a disparate impact. A proposed change in any step of the examination procedure, up to and including the examination questions, can have an unexpectedly profound effect on the results.

---

254 See infra Part V.A.1.
256 86 F.3d 315 (2d Cir. 1996); see supra notes 62-73.
258 Id. at 187.
259 Id. at 178.
261 Id. at *17.
In addition, even if the Fire Department were able to meet these goals, the court would enforce the plan at least while the next two entry-level exams were administered. Currently, the Fire Department administers the exam every four years, but plaintiffs are hoping to condense it to every two years. Therefore, the court will likely enforce the plan for at least four years—if not eight—regardless of the Fire Department’s progress.

More importantly, the court overstepped its authority by failing to properly address the “efficacy of alternative remedies.” Specifically, the court refused to address the efficacy of more direct race-conscious hiring practices, and in fact outright rejected this possibility. A potent race-conscious interim hiring measure has the potential to be just as successful in attacking racial discrimination without encroaching on local government. In the event that these measures were unsuccessful, only then could the court find that more intensive oversight is necessary.

C. Regardless of Who Is in Control, Minority Focused Recruitment Measures Should Be Implemented

Providing for recruitment policies will lead to many of the advantages of race-conscious hiring without directly imposing a race-conscious remedy. Therefore, whether the court or the Fire Department is in control of developing the hiring program, it should be apparent to both parties that enhanced recruiting measures would be both beneficial and not politically costly. Thus, even if the court did not impose strict oversight, it is highly likely that the Fire Department would have implemented its own recruiting procedures.

---

262 See supra note 218.
263 Paradise, 480 U.S. at 187.
264 See supra Part V.A.
265 While minority recruiting efforts are “race-conscious” in the general sense of the word, such efforts are not as controversial as affirmative action techniques that account for race during or after the application process itself. Minority recruiting, for example, does not raise any possible doubt as to the qualifications of the accepted applicants, because they objectively scored well on the test without adjusting for race. Similarly, it poses little threat to white applicants who fear that minority applicants who did less well on the exam will nevertheless be hired in place of them.
Although minority recruitment efforts in the Fire Department began as early as 1991, the sincerity of the effort and its actual successes are open to question. In 1991, the City’s first black mayor and first Hispanic fire commissioner stated they would increase firefighter diversity.\footnote{William Murphy & Joseph W. Queen, Recruiting Force FDNY Minority Drive Described as Sadly Lacking, NEWSDAY (N.Y.C.), June 25, 1996, at A04.} Fire Department officials, however, would rarely if ever attend meetings when minority recruitment was the topic of discussion and future meetings were eventually cancelled.\footnote{Id. (noting that “[r]ecruiters could not attend some strategy meetings because their supervisors gave them ‘priority’ assignments that took them elsewhere”). The authors derived their information from the “city’s Equal Employment Practices Commission, a mayoral agency, in its 1994 annual report.” Id.} The Fire Commissioner at the time, Carlos Rivera, complained that City Hall “was not really committed to minority hiring and . . . never gave him authorization to spend more money for recruiting.”\footnote{Id.} The recruiting drive was not successful by any calculation; not only were minority applicants who passed the exam still underrepresented as compared with the total number of minority applicants, but the total number of minority applicants continued to be underrepresented when compared to the local labor pool.\footnote{See Saltonstall, supra note 215 (noting that in February 1999, “only 10% of those who took the February exam were black and only 12% Hispanic, an increase of only a few percentage points from 1992, when the last test was given”). The lack of minority turnout resulted “despite a much-ballyhooed minority recruiting drive the department undertook in 1994.” Id.}

Although unsuccessful in the past, if the Fire Department were required to implement a race-conscious hiring procedure until it could provide a reasonable alternative, it would have incentive to conduct a recruitment campaign properly. Under those circumstances, a successful recruitment campaign would be politically desirable because it would allow the Fire Department to avoid the public backlash from using quotas or race-conscious hiring; if they successfully recruit and train a sufficient number of qualified black and Hispanic applicants to meet the imposed hiring ratios, they would not need to engage in any race-conscious hiring after the scores were submitted. Thus, the desire to avoid race-conscious relief should make developing a recruitment policy an especially attractive option to both the court and the Fire Department.\footnote{The court in United States v. City of New York has shown a propensity to limit race-conscious relief when possible. See supra notes 85-88 and accompanying text.}

There are many other advantages to developing a successful recruitment program to the benefit of all parties.
would naturally increase the number of minorities who choose to join the applicant pool, which no amount of race-conscious hiring could ever achieve directly. It would also lessen the negative impact of using the pool of applicants as the “Representative Ratio,” since the ratio of minority applicants could equal and possibly even eclipse the ratio of minorities in the local population after successful recruitment.

A recruitment program would also help prevent the type of disparate impact liability that would stem from an underrepresentation of minority applicants. Lieutenant Rod Lewis believes “the recruiting problem is, in part, a matter of perception in the city’s minority neighborhoods . . . [y]ou see the Fire Department . . . [b]ut you don’t see black firefighters, so many kids don’t think it’s an option.” By reaching out to minorities, and sending the message that they have legitimate opportunities to become firefighters, the Fire Department will be able to break the “Irish and Italian firehouse culture in which fathers routinely pass on their knowledge and contacts to sons eager for work.” As a result, the City will be better protected from racial discrimination claims regarding its recruiting programs.

The methods of recruitment suggested in plaintiffs’ memorandum should largely be implemented. It is especially important to develop outreach programs that help train minority candidates for the entry-level exam, since current firefighters’ families have been enjoying access to similar preparatory programs for some time. Such a program would be directly related to the subject of the litigation because it would decrease the disparate impact of any future testing. It may also be a more cost-effective and reliable way to reduce disparate impact than relying on consultants to create a test that is reliable, predictive, and consistent with business necessity. Standardized tests, such

272 Saltonstall, supra note 215.
273 See id.
274 See generally Memorandum of Law in Support of Plaintiffs-Intervenors’ Proposed Order for Injunctive Relief, City of New York, No. 07-CV-2067 (E.D.N.Y. Dec. 9, 2010), 2010 WL 6917514, ECF No. 596; see also supra Part IV.G.
275 See Memorandum of Law in Support of Plaintiffs-Intervenors’ Proposed Order for Injunctive Relief at 15, City of New York, No. 07-CV-2067 (E.D.N.Y. Dec. 9, 2010), supra note 274.
276 See Saltonstall, supra note 215.
277 The City argued that their recruitment process was not the subject of the litigation, despite acknowledging that the court has great leeway in developing a remedy. See Defendants’ Memorandum of Law in Opposition to Plaintiffs-Intervenors’ Motion for a Proposed Order for Injunctive Relief at 2-7, City of New York, No. 07-CV-2067 (E.D.N.Y. Dec. 9, 2010), 2010 WL 6917516, ECF No. 599.
as the SATs and LSATs, have proven that money and intense research can only go so far in creating a successful, nondiscriminatory exam. After more than thirty years of waiting for the City to develop a test that accomplishes this feat, the court should recognize that such an effort may not ever be wholly successful, even if the City acts in good faith. Preparing minorities for the exam, on the other hand, attacks the racial disparity at its source; it educates minorities in order for them to have the tools already available to nonminorities.

CONCLUSION

Although recent political movements have attempted to rid the nation of affirmative action, courts should not acquiesce to them because doing so undermines well-established doctrines in Title VII. The court in United States v. City of New York, however, has given full consideration to the movement in its decision. In an effort to avoid race-conscious hiring measures, the court instead instituted an extensive oversight program. While this program gives the plaintiffs, past victims of discrimination, and future minority applicants hope of recourse for the systematic disparate treatment that plagued the Fire Department for decades, this achievement could have been reached without imposing heavy oversight on local government by a federal court.

Future courts, when dealing with deep-seated discrimination, should instead require defendants to engage in interim hiring at a ratio equal to that of the local minority population. If such a plan were implemented here, the City would have had sufficient incentive to develop an unbiased exam and genuine recruitment efforts within a short time.

Hal Budnick†

† J.D. Candidate, Brooklyn Law School, 2012; B.S., Cornell University, 2008. I would like to thank the editors and staff of the Brooklyn Law Review for all their hard work. And a special thanks to my mom and dad, who would always break my writer’s block when it came to vocabulary homework.
Cleaning Up the Muck
A TAKINGS ANALYSIS OF THE MORATORIUM ON DEEPWATER DRILLING FOLLOWING THE BP OIL SPILL

INTRODUCTION

On April 20, 2010, the Deepwater Horizon drilling platform exploded in the Gulf of Mexico. The blast killed eleven workers and triggered the worst oil spill in America’s history.¹ The platform was owned by Transocean Services, Ltd., and was under lease to British Petroleum, PLC (BP), for the purpose of drilling an exploratory well five thousand feet below the ocean’s surface off the coast of Louisiana.² Aside from the unfortunate deaths of the workers, another tragedy unfolded as it became clear that the spill could not be stopped for weeks or even months. Around-the-clock video feeds of the oil spewing from the leak showed viewers the enormity of the disaster, capturing the hearts and interest of the nation.³ It was estimated that over two hundred million gallons escaped from the leaking well before it was finally capped in mid-July.⁴ The environmental effects on the region’s wildlife—and on the people who depended on that wildlife for their subsistence—were staggering.⁵

³ In fact, the oil spill was the most searched-for term in Yahoo’s search engine in 2010, the first time since 2005 that a celebrity had not topped the list. Michael Liedtke, BP Oil Spill Swamps Yahoo Search Engine in 2010, YAHOO!FINANCE, (Dec. 1, 2010, 7:54 AM), http://finance.yahoo.com/news/BP-oil-spill-swamps-Yahoo-apf-1847949440.html?x=0&.v=6.
⁴ Editorial, Gulf Leak Is Over, Impacts Still Uncertain, DAY (New London, Conn.), Sept. 22, 2010. To put this astronomical number in perspective, the infamous Exxon Valdez disaster that occurred off the coast of Alaska in 1989 resulted in a spill of about eleven million gallons. David Dipino, Researcher Warns that Current Could Still Bring Oil to the Area, SUN SENTINEL (Fort Lauderdale, Fla.), Sept. 8, 2010, at 1.
⁵ Hornbeck, 696 F. Supp. 2d at 630 n.2 (“As a result [of the spill], nearly one-third of the Gulf of Mexico has been closed to commercial and recreational fishing.”).
Following the Deepwater Horizon disaster, it became immediately apparent that a prolonged oil leak posed a significant threat to wildlife living in and near the Gulf of Mexico. The region “is home to more than 400 marine and coastal fish and wildlife species,” including five endangered species of sea turtle, a variety of birds such as brown pelicans and terns, and several marine mammals including sperm whales and bottlenose dolphins. Additionally, it is the largest spawning ground in the world for blue fin tuna. About a month after the spill began, more than seven hundred dead animals had already been collected from the Plaquemines, Jefferson, and Lafourche parishes of Louisiana alone. Unlike beach goers and bathers in the area, many animals were ill-equipped or unable to avoid exposure to the spill. One expert expressed concern over the potential threat to the reproductive capabilities of fish in the area, noting, “Fish can swim away from the oil spill . . . , but eggs and larvae cannot.” Even worse, sea turtles in the area were observed attempting to feed on the tar balls that were prevalent in the Gulf.

Aside from those animals affected by direct exposure to the oil, the damage the spill caused to the marine and coastal habitat also posed more long-term, indirect threats. The salt marshes and mangrove coastlines that make up the wetlands surrounding the Gulf were described as practically impossible to clean without doing additional damage. Coast Guard Admiral

---

5 Dipino, supra note 4.
7 See Gulf Coast Oil Spill—pt. 1: Hearing Before the H. Comm. on Transp. and Infrastructure, 111th Cong. (2010) (statement of Larry Schweiger, President/CEO, Nat’l Wildlife Fed’n). For example, young sea turtles were described as “prone to being poisoned or coated by the sticky oil.” Id. Their adult counterparts fared no better as they “show[ed] no natural avoidance behaviors when confronted with an oil slick” and those that remained in the affected area often suffered from malnutrition. Id. Likewise, birds that tried to cope with their exposure to the spill by grooming the oil out of their feathers often exacerbated their problems by inadvertently consuming more oil in the process. Id.
8 Dipino, supra note 4.
Thad Allen, in charge of the federal effort to clean up the spill, described oil in the wetland marshes as a “worst-case scenario.” These wetlands are essential to the water quality of estuaries, with 98 percent of the fish and shellfish of the Gulf relying on them for food, shelter, and breeding. The potential danger was even more ominous considering that “[m]ore than 20 years after the Exxon Valdez spill, oil can still be found on Alaska’s beaches, and many species have not completely recovered.”

In addition to the perils facing the wildlife and the environment, the spill was particularly alarming for those people involved in Louisiana’s seafood, tourism, and recreation industries, which bring in almost $4 billion each year. Even with the spill barely a month old, and before pictures of the coastal impact inundated the American household, USA Today found that 13 percent of people polled would not eat any seafood that originated in the Gulf. Adding to these fears was a concern that the aggressive cleanup efforts might actually increase the damage to seafood and the environment in general. The chemical dispersants being used were of particular concern. The dispersants used by BP to break up and dissolve the oil were assumed to be safe, but, alarmingly, the long-term health and environmental effects are still not well known. Additionally, while the use of dispersants may be an effective means for hastening the degradation of oil, it also “increases the...
risk that aquatic life in the water and on the sea floor will be exposed to oil . . . and add yet more inherently toxic chemicals to the already toxic oil.\textsuperscript{21} The federal government acknowledged this potential danger, at one point ordering BP to “identify and use a less toxic and more effective dispersant.”\textsuperscript{22} Even without proof of actual contamination or danger, the enormous amount of media coverage discussing possible contamination resulted in a general consumer aversion to seafood from the Gulf that, in turn, destroyed the economic prospects of many of those who relied on the seafood industry in the region.\textsuperscript{23}

In response to the spill, the White House issued a moratorium on all oil drilling in the Gulf of Mexico and the Pacific Ocean.\textsuperscript{24} Critics argued that this action did nothing more than exacerbate the impact on the region by creating another class of employees to stand in the unemployment line, namely those who rely on the energy industry.\textsuperscript{25} As a result—much like the blobs of oil that arrived on beaches hundreds of miles from their source—the legal consequences quickly spread throughout the country, touching the lives of thousands of people.\textsuperscript{26} In addition to the seemingly endless procession of environmental liability lawsuits facing BP, companies and individuals in the industry—who were innocent in causing the spill but who felt they were being punished or forced to suffer nonetheless—brought litigation against the federal government.\textsuperscript{27}

One of these cases was \textit{Hornbeck Offshore Services L.L.C. v. Salazar}, in which a group of offshore oil and gas drilling support


\textsuperscript{26} A former chairman of the Alaska Oil Spill Commission’s legal task force after the Exxon Valdez disaster estimated that BP’s clean up and legal liability costs could reach as high as $90 billion. Stempel, supra note 15. The quantity of lawsuits involved in the BP litigation is expected to become so large that it has even been compared to asbestos and tobacco litigation. Id.

\textsuperscript{27} \textit{See} First Supplemental and Amended Complaint for Declaratory and Injunctive Relief at ¶ 89(b), \textit{Hornbeck}, 696 F. Supp. 2d 627 (No. 10-1663).
and service providers located in the Gulf sued the government, claiming that the overly broad moratorium unjustly deprived them of their service contracts with oil exploration companies unassociated with the spill. Further, they claimed that the damage was potentially irreparable if those exploration companies permanently left the Gulf for other waters as a result of the moratorium.

Certainly, quick government action was required to stop and then clean up the leaking oil and to punish those who caused the spill. But the broad and dramatic oil-drilling moratorium, and its subsequent detrimental impact on the oil industry in the Gulf, illustrates the dangerous potential of reactive government regulation that forces innocent parties to bear a burden more rightly placed on others. Protection against this threat can be found in the takings clause of the United States Constitution. Accordingly, this note will focus on a takings claim that was briefly mentioned but never fully argued or ruled on in *Hornbeck* to show that such a claim should provide a valid recourse for future oil industry plaintiffs affected by federal regulations in response to oil spills. Moratoria impose uniquely detrimental burdens on service industry entities that rely on property interests with a definite life span. Incorporating these burdens into a traditional takings analysis will deter the federal government from passing moratoria that are too rash or broad. At a minimum, such an application would provide a compensation mechanism to those who suffer as a result of moratoria that are necessary but nonetheless detrimental. Accordingly, a court applying a takings analysis to a factual situation similar to the events that unfolded in *Hornbeck* should find that a taking occurred and that compensation must be paid—despite an unwillingness to come to such a holding in the past.

Part I of this note discusses the government’s response to the Deepwater Horizon disaster. Part II analyzes the legal

---

29 *Id.* at 638.
30 See Plaintiffs’ Original Complaint and Application for Temporary Restraining Order and Injunctive Relief at ¶ 33, *Hornbeck*, 696 F. Supp. 2d 627 (No. 10-1663(F)(2)); First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, *supra* note 27, ¶ 91; Diamond Offshore’s Brief in Support of Emergency Motion to Intervene at ¶ 11, *Hornbeck*, 696 F. Supp. 2d 627 (No. 10-1663(F)(2)). Instead of focusing on the takings claim, the bulk of the plaintiffs’ argument and the focus of the court rested on grounds that the moratorium was invalid under the Administrative Procedure Act (APA) because it was arbitrary and capricious. *See Hornbeck*, 696 F. Supp. 2d at 636-38.
proceedings stemming from the *Hornbeck* plaintiffs’ lawsuit against the federal government. Part III describes the history of the takings claim in order to frame the proper regulatory takings analysis for the oil-drilling moratorium. Part IV presents previous examples of federal government action that affected the property rights of legitimate leaseholders to explore and drill for oil on their property, and describes the litigation that arose from it. Part V applies current takings jurisprudence to determine whether the moratorium in *Hornbeck* amounted to a Fifth Amendment taking of property that required just compensation. Finally, this note concludes by arguing that *Hornbeck* further demonstrates that, while takings jurisprudence remains seriously muddled, case law on the subject provides at least theoretical latitude for plaintiffs to bring takings challenges against federal moratoria. Further, the conclusion asserts that the courts have been too conservative in their analysis of temporary takings and that the oil-drilling moratorium presents a factual scenario where fairness and justice require a more liberal and inclusive takings analysis.

I. THE GOVERNMENT’S RESPONSE

The government took decisive action in response to the unprecedented environmental disaster caused by the Deepwater Horizon explosion. President Obama formed a bipartisan commission dubbed the National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (Commission).\(^3\) The Commission consisted of a seven-member team led by former Florida governor and former U.S. senator Bob Graham and former Environmental Protection Agency administrator William Reilly.\(^4\) It was tasked with “investigating the facts and circumstances concerning the cause of the blowout.”\(^5\) Investigations are still ongoing at the time of this writing, but it has been estimated that BP’s civil and criminal liabilities for long-term restoration of the Gulf will likely exceed $15 billion\(^6\) and could balloon as high as $90 billion.\(^7\)


\(^5\) *Hornbeck*, 696 F. Supp. 2d at 630.

\(^6\) John M. Broder, *Panel Wants BP Fines to Pay for Gulf Restoration*, N.Y. TIMES, Sept. 28, 2010, at A17. In addition to the investigation by the Commission, several other investigations commenced. Eilperin & Lebling, *supra* note 31. These included investigations by the Marine Board of Investigation to identify the factors leading to the explosion, the House Energy and Commerce Committee to determine the
In addition to the various investigations into the causes of and liabilities for the spill, President Obama ordered the Secretary of the Interior, Kenneth Salazar, to report and recommend any additional precautions and regulations that should be required to improve the safety of oil exploration and production on the outer continental shelf. The examination was conducted by the Department of the Interior in conjunction with a panel of experts from various levels of “state and federal governments, academic institutions, and industry and advocacy organizations.” After a thirty-day examination, the White House issued a report purportedly based on the panel’s findings and ordered a six-month moratorium halting all offshore exploratory drilling in depths of more than five hundred feet of water.

The moratorium met with sharply divided reviews from politicians and the media alike. Supporters argued the moratorium was necessary because the government simply could not risk the possibility of another spill. Among these supporters was Representative Edward Markey of Massachusetts, who noted, “The only thing worse than one oil

---

35 Stempel, *supra* note 15. For its own part, BP has estimated that $40 billion should cover its liabilities for the spill. Tom Bergen, *Special Report: How BP’s Oil Spill Costs Could Double*, Reuters (Dec. 1, 2010), http://www.reuters.com/article/2010/12/01/us-special-report-how-bps-oil-spill-cost-idUSTRE6B02PA20101201. However, this number has been disputed as overly optimistic and as an attempt to underestimate the costs. *Id.* Instead, an analysis by Reuters stated that the “fines, damages, costs related directly to the leak, compensation and the damage to BP’s business suggests the final spill bill could, over the long term, end up [being] twice as much.” *Id.*


37 *Hornbeck*, 696 F. Supp. 2d at 630.

38 *Id.* at 630-31.


rig in the bottom of the Gulf of Mexico would be two oil rigs in the bottom of the Gulf of Mexico.” Likewise, Senator Bill Nelson praised the moratorium stating, “Until we know what happened with the Deepwater Horizon, and we'll know very soon, it makes sense not to put Gulf Coast residents and the economies there at further risk.” The economic consequences of the moratorium on the Gulf Coast region, particularly on the oil industry, were acknowledged by supporters but seen as a necessary evil to prevent additional future harm.

Critics of the moratorium saw the situation dramatically differently, finding the undeniably immense economic consequences impossible to ignore in a region still recovering from Hurricane Katrina. One estimate predicted the moratorium would cause a nationwide loss of over twelve thousand jobs, $2.8 billion in economic activity, and $219 million in tax revenue. Further, critics worried that drilling companies currently located in the Gulf, or in the process of applying for a future lease to drill, would instead leave for foreign countries rather than wait for the moratorium to end. Once gone, the concern was that drilling companies would “not return for several years, if ever.” The long-term effects on the national economy and scientific progress were lamented as being equally dire.

---

42 Ann Woolner, Editorial, Overreaching Times Two: A Judge Goes Too Far to Overturn a Deepwater Drilling Moratorium that Went Too Far, PITTSBURGH POST-GAZETTE, June 25, 2010 (internal quotation marks omitted).
43 Hughes & Power, supra note 25 (internal quotation marks omitted).
44 Id.
47 Hughes & Power, supra note 25. The CEO of Diamond Offshore, one of the plaintiffs in Hornbeck, told a presidential commission that his company had already sent two deepwater rigs to foreign waters as a result of the moratorium, and warned that “there won’t be much of a U.S. industry left” if the moratorium remains in place. Id.
49 Rebecca Terrell, Oil Leak Outrage, 26 NEW AMERICAN, July 19, 2010, available at 2010 WLNR 15967962. One senator even compared the moratorium to “the aftermath of the Three Mile Island nuclear power plant disaster . . . that brought all
II. LEGAL Fallout

The moratorium also sparked significant legal debate. On the extreme end of the spectrum, the moratorium was condemned as everything from a blatant executive overreach lacking reason and spurred by fear, to an unconstitutional regulation of commerce by the executive branch in violation of the separation of powers. Armchair debaters aside, there are real and tangible legal issues stemming from the actions of the federal government in the wake of the Deepwater Horizon accident. On June 7, 2010, Hornbeck Offshore Services, L.L.C. (Hornbeck) filed suit in the United States District Court for the Eastern District of Louisiana against the Secretary, the Department of the Interior, the Minerals Management Service (MMS), and the Director of the MMS seeking declaratory and injunctive relief to end the moratorium. Subsequently, additional plaintiffs joined the litigation. Judge Martin Feldman, presiding over the case, issued a preliminary injunction against the enforcement of the moratorium. The court held that, based on the administrative record, the blanket moratorium on all drilling wells of more than five hundred feet of water was likely to be found “arbitrary and capricious” and was thus invalid under the Administrative Procedure Act (APA) and the Outer Continental Shelf Land Act (OCSLA) and its implementing regulations. The following section will further explore that decision.

A. The Hornbeck Decision

Judge Feldman framed the issue to be decided in Hornbeck as “whether the federal government’s imposition of a general moratorium on deepwater drilling for oil in the Gulf of Mexico was imposed contrary to law.” The statutes governing the nuclear power plant applications to a screeching halt.” Id. The senator added, “In hindsight that was not the right decision. Today, we are 30 years behind the French in nuclear technology.” Id.

Woolner, supra note 42.


Hornbeck, 696 F. Supp. 2d at 632.

Id.

Id. at 630.

Id. at 639.

Id. at 630.
outcome were OCSLA, which provides authority to the Secretary to suspend leases in the Gulf under certain circumstances,\(^\text{57}\) and the APA, which authorizes the federal courts to review final agency action.\(^\text{58}\) The plaintiffs generally alleged that the moratorium by the Secretary as well as the Notice to Lessees (NTL) implementing the moratorium were “arbitrary, capricious, an abuse of discretion and otherwise not in accordance with the APA, OCSLA and its implementing regulations.”\(^\text{59}\) Additionally, they made a brief allegation that the moratorium was an impermissible “taking” of . . . property rights in violation of the 5th Amendment to the United States Constitution.\(^\text{60}\)

The focus of the plaintiffs’ complaint was directed at the arbitrary and capricious claim.\(^\text{61}\) They claimed that the moratorium was unwarranted based on the Report provided to the Secretary by a panel of experts.\(^\text{62}\) They alleged that the Secretary had exaggerated or entirely invented the experts’ support and recommendations for the moratorium.\(^\text{63}\) Further, they alleged the Secretary had failed to adequately explain the reasons behind the suspension of operations\(^\text{64}\) or why he had chosen a general depth limit of five hundred feet for the drilling ban.\(^\text{65}\) The plaintiffs pointed to a lack of individualized justification for the moratorium, stating,

The Report itself does not contain any facts, data, analysis or risk assessment concerning why the Secretary imposed a Moratorium on further drilling by the "33 [existing] wells." Twenty-nine of these

\(^{57}\) Id. at 632-33 (citing 43 U.S.C. § 1334(a)(1)).

\(^{58}\) Id. at 634 (citing 5 U.S.C. §§ 702, 704).

\(^{59}\) First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 22. One of the expert panelists stated that if anybody had made the suggestion of a moratorium on existing drilling, “we’d have said that’s craziness.” Crude Politics, WALL ST. J., June 17, 2010 (internal quotation marks omitted).

\(^{60}\) Diamond Offshore’s Brief in Support of Emergency Motion to Intervene at 4, Hornbeck, 696 F. Supp. 2d 627 (No. 10-1663(F)(2)).

\(^{61}\) See generally First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27.

\(^{62}\) Id. ¶ 52.

\(^{63}\) Id. ¶ 83. At least some of this skepticism was confirmed when it “was exposed that an important White House official had changed the Safety Report before its public release, which created the misleading appearance of scientific peer review.” Hornbeck Offshore Servs., L.L.C. v. Salazar, No. 10-1663(F)(2), 2011 WL 454802, at *2 (E.D. La. Feb. 2, 2011).

\(^{64}\) Hornbeck, 696 F. Supp. 2d at 631.

\(^{65}\) The Deepwater Horizon operation was conducted at a depth of over five thousand feet. Hornbeck’s Memorandum of Law in Support of Its Motion for Preliminary Injunction, supra note 48, at 10. The Report issued to the Secretary noted that, compared to drilling in shallow water, risks were greater after one thousand feet. First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 89(d).
wells had been subjected to additional inspections following the Incident. According to the “MMS Deepwater Drilling Rig Inspection Report” . . . , issued on May 11, 2010, MMS found no violations of governing regulations or existing permit terms on 27 of the 29 drilling rigs inspected and only minor violations on the two others. Further, each of the 33 rigs had previously satisfied the rigor of the MMS permitting process."

Additionally, they expressed concern about the injurious economic effects of the moratorium, exclaiming that “lost wages for direct and indirect jobs lost could be over $165 million to $330 million per month for every month the 33 platforms are idle.” The long-term effects were viewed as similarly alarming. The Report stated that the offshore operations provide employment for approximately 150,000 people. The moratorium put many of these jobs at risk. Further, without robust and continuous drilling activities, this labor force would lack incentive to remain in the region, thus reducing the ability of companies like the Hornbeck plaintiffs to find workers. Finally, the plaintiffs also pointed to the possibility that the moratorium might actually last substantially longer than six months, an unacceptable possibility for an industry that relied on contracts and equipment with a limited useful life.

The government countered by citing the relevant portions of OCSLA that specifically authorize the Secretary to direct a suspension of drilling whenever it determines that “activities pose a threat of serious, irreparable, or immediate harm or damage” to human or animal ‘life, property, [ ] mineral deposit, or the marine, coastal, or human environment.” The Secretary highlighted that the moratorium was needed to “address critical

66 First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 46.
67 Hornbeck’s Memorandum of Law in Support of Its Motion for Preliminary Injunction, supra note 48, at 23. But see Woolner, supra note 42 (noting that “five days before filing suit, Hornbeck Offshore Services Inc. assured the Securities and Exchange Commission and the investing public that the ban would have little effect on it”).
68 First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 89a.
69 Id.
70 Id. ¶ 54.
71 Id. ¶¶ 93-98. Plaintiffs noted that they could lose as much as $500,000 per day for each rig while the moratorium remained in place. See Plaintiffs’ Original Complaint and Application for Temporary Restraining Order and Injunctive Relief at ¶ 10, supra note 30. Additionally, some contractual parties were cancelling their contracts with plaintiffs altogether. First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 67.
72 Defendants’ Opposition to Plaintiff’s Motion for Preliminary Injunction at 6, Hornbeck Offshore Servs., L.L.C. v. Salazar, 696 F. Supp. 2d 627 (E.D. La. 2010) (No. 10-1663(F)(2)) (quoting 30 C.F.R. § 250.172(b)).
spill containment and response deficiencies” and warned that there were “insufficient available response resources should another deepwater spill occur while the containment and clean up efforts [were ongoing] . . . .” The government pointed out that courts must defer to agency decisions that are supported by a thorough administrative record, and in this case, “the interim safety measures in the Safety Report and the corresponding suspension of deepwater drilling [were] appropriately supported by the Administrative Record.” The defendants spent little time addressing the Fifth Amendment takings claim, asserting only that it was “both wholly without merit and outside of the jurisdiction of this Court to adjudicate.”

Judge Feldman issued his decision on June 22, 2010, holding that the moratorium was contrary to law and that he was “unable to divine or fathom a relationship between the findings and the immense scope of the moratorium.” He noted that the Report—supposedly the supporting basis for the moratorium—focused narrowly on the Deepwater Horizon incident alone. In contrast, the resulting moratorium was exceedingly broad, applying to rigs that had exemplary safety records and that drilled in significantly shallower water than the Deepwater Horizon. Judge Feldman found it hard to believe that such a suspension would be deemed appropriate in other contexts, asking, “If some drilling equipment parts are flawed, is it rational to say all are? Are all airplanes a danger because one was? All oil tankers like Exxon Valdez? All trains? All mines? That sort of thinking seems heavy-handed, and

---

73 Memorandum in Support of Defendants’ Motion to Dismiss Complaint at 5, Hornbeck, 696 F. Supp. 2d 627 (No. 10-1663(F)(2)).
74 Defendants’ Opposition to Plaintiff’s Motion for Preliminary Injunction, supra note 72, at 16. The government argued that an agency’s investigation and choice of methodology are entitled to particularly broad deference when the agency is responding to an emergency. See id.
75 Defendants’ Response to Diamond Offshore’s Motion to Intervene at 3 n.4, Hornbeck, 696 F. Supp. 2d 627 (No. 10-1663(F)(2)).
76 Hornbeck, 696 F. Supp. 2d at 637.
77 See id. “[The Report] is incident-specific and driven: Deepwater Horizon and BP only. None others.” Id.
78 See id. at 637-38 & n.11; see also Hornbeck’s Memorandum of Law in Support of Motion for Preliminary Injunction, supra note 48, at 10. The Deepwater Horizon well was drilled in nearly five thousand feet of water, and the Report addressed wells in depths greater than one thousand feet, yet the NTL set the moratorium at the significantly shallower depth of five hundred feet. Hornbeck’s Memorandum of Law in Support of Motion for Preliminary Injunction, supra note 48, at 10.
rather overbearing.”

Accordingly, the court held that the government’s actions in implementing the moratorium had been “arbitrary and capricious” and were thus contrary to the requirements of the APA and OSCLA. Therefore, the court granted the plaintiffs’ motion for a preliminary injunction preventing the moratorium from being enforced. Because the parties had failed to fully argue it, and perhaps to avoid entering the difficult and muddled jurisprudence of takings analysis, the court did not analyze or even mention the merits of the plaintiffs’ takings claim.

B. The Government’s Response to the Injunction

The decision by Judge Feldman led to additional controversy surrounding the moratorium. Only days after the ruling, the Secretary publicly announced that the government was working on passing a second moratorium. The government reiterated this intention when—just hours before the district court’s decision was appealed before the United States Court of Appeals for the Fifth Circuit—a senior administration official announced that the government “would immediately issue a new moratorium” regardless of the outcome of the appeal. The maneuver sparked outrage from critics who claimed that the statements were made in a brazen attempt to intimidate the court. Nevertheless, on July 12, 2010, the Secretary issued a

---

79 Hornbeck, 696 F. Supp. 2d at 637. Also of import was the fact that since 1969 there had only been three deepwater blowouts before the Deepwater Horizon, none of which were in the Gulf of Mexico. See id. at 638 n.11.
80 Id. at 639. The court was careful to note that “a suspension of activities directed after a rational interpretation of the evidence could outweigh the impact on the plaintiffs and the public” but that here the facts of the case could not support such a determination. Id.
81 See id.
82 See generally Hornbeck, 696 F. Supp. 2d 627.
83 Plaintiffs’ Response in Opposition to Defendants’ Motion to Dismiss Complaint at 2, Hornbeck, 696 F. Supp. 2d 627 (No. 10-1663(F)(2)). After the first injunction was ordered, the Secretary stated, “The decision to impose a moratorium on deepwater drilling was and is the right decision” and that ‘I will issue a new order in the coming days that eliminates any doubt that a moratorium is needed, appropriate, and within our authorities.’” Id. at 5.
84 Id. at 6.
85 See Kingsley Guy, Op-Ed., Obama’s Over-Reach: President Remaking Courts, SUN SENTINEL (Pt. Lauderdale, Fla.), July 25, 2010, at F5 (comparing the Obama administration’s response to Judge Feldman’s injunction to the court packing of FDR and expressing that “regardless of their views on offshore drilling, Americans should be concerned about the heavy-handed action by the Obama administration of reinstating a moratorium. It demonstrates contempt for the judicial system and the attitude of, ‘I’m the president and I can do anything I want.’”).
memorandum rescinding the first moratorium but ordering a new—yet similar—blanket suspension on offshore oil drilling. Additionally, the government moved to dismiss the original suit on the grounds of mootness since the original moratorium was no longer in effect. Counsel for the plaintiffs, incensed by the government’s actions, invoked *Marbury v. Madison* and exclaimed that the decision to pass a new moratorium with the same practical effects as the now enjoined original one constituted executive interference with the judicial branch and the judicial review process.

The motion for dismissal was addressed on September 1, 2010, when Judge Feldman again ruled against the government, holding that mootness did not apply and stating that the second moratorium was essentially the same as the first one. In addressing the issue of whether the Secretary had the authority to rescind the first moratorium, he noted that the proper procedure for an agency seeking to reconsider a decision that is under judicial review is for the agency to move the court to remand. The court voiced its concern that “if agencies are not required to move to remand, they may use rescission and reissuance of their decisions as a way to manipulate the federal jurisdiction of U.S. courts.” Ultimately, Judge Feldman concluded the rescission did have “some administrative force,” but this was not enough to save the defendants’ motion to dismiss. The court criticized their maneuvering, expressing that, “In reality, the new moratorium covers precisely the same rigs and precisely the same deepwater drilling in the Gulf of Mexico as did the first moratorium.” The court did not specifically decide whether the second moratorium was again

---

87 See id.
88 See Plaintiffs’ Response in Opposition to Defendants’ Motion to Dismiss Complaint, *supra* note 83, at 16. “Simply put, the law does not allow for the manipulation of the ‘orderly operation of the federal judicial system,’ . . . .” Id. (quoting U.S. Bancorp Mortg. Co. v. Bonner Mall P’ship, 513 U.S. 18, 26 (1994)). According to the plaintiffs, it did not matter how necessary or right the defendants believed their actions to be; instead, “It is emphatically the province and duty of the judicial department to say what the law is. Those who apply the rule to particular cases, must of necessity expound and interpret that rule.” Plaintiffs’ Response in Opposition to Defendants’ Motion to Dismiss Complaint, *supra* note 83, at 16 (quoting *Marbury v. Madison*, 5 U.S. 137, 177 (1803)).
90 See id. at *4-5.
91 Id. at *4.
92 Id.
93 Id. at *1.
arbitrary and capricious (the sole issue before the court was whether the case surrounding the first moratorium was now moot), but instead focused on whether the harm imposed by the first moratorium would also be imposed by the second. Under the voluntary cessation exception to mootness claims, a federal court will only find a case to be moot if the subsequent government action makes it clear that the initial harm could not reasonably be expected to recur. Judge Feldman noted that the government’s public announcements immediately following his initial ruling sharply undermined their argument that the second moratorium was based on a significantly supplemented administrative record. More importantly, these public announcements and posturing indicated that there was a reasonable expectation the harm to the plaintiffs could recur and thus the government’s repeal of the first moratorium did not render the action moot. Accordingly, Judge Feldman denied the defendants’ motion to dismiss.

For some time, while the Hornbeck suit was underway, new litigation continued to emerge as a result of the moratorium. Additional plaintiffs brought claims that the moratorium had effectively ended drilling in shallow water located in entirely different parts of the country. But it now

---

94 See id. at *7. Defendants argued that the new moratorium was based on new information and a new administrative record and that by lifting the old moratorium all plaintiffs’ claims had become moot. See id. at *1.
95 See id. at *5. Plaintiffs argued that their claims were not moot because the new moratorium applied to the same rigs, in the same area, for the same amount of time, and thus the new moratorium would cause them the same harm as the old one. See id. at *2.
96 Id. at *7. Commenting on the Secretary’s announcement promising a second moratorium just moments after the first injunction was entered, the court stated, “It is difficult to square such public expressions of resoluteness, with the government’s assertion that its rescission of the first moratorium and its issuance of a new moratorium is entitled to solicitude and should not be considered litigation posturing.” Id. at *8.
97 Id. at *7.
98 Id. at *8.
99 Margaret Fisk, Alaska Claims in Suit U.S. Government Improperly Banned Off-Coast Drilling, BLOOMBERG (Sept. 10, 2010), http://www.bloomberg.com/news/print/2010-09-09/u-e-improperly-banned-drilling-off-alaska-coast-state-alleges-in-lawsuit.html. For instance, the state of Alaska brought suit against the Secretary, Alaska v. Salazar, 3:10-cv-00205 (D. Alaska 2010), in early September 2010 claiming that the Secretary improperly banned drilling off the coast of Alaska following the Deepwater Horizon oil spill. No drilling permits have been issued in the Arctic since the incident, even though an Interior Department spokeswoman acknowledged that “[t]he moratorium is on deep-water drilling and there is no deep-water drilling in Alaska.” Id. Additionally, suits were also brought closer to the site of the spill. For example, Exxon Mobile Corporation sued the federal government in August 2011 claiming that it was being deprived of its right to drill in the Julia field in the Gulf of Mexico, an area estimated to contain billions of barrels of oil. Jonathan Stempel, Exxon
appears that any formal need for the courts to enjoin the moratorium has largely passed; the moratorium was lifted on October 12, 2010, several weeks before it was scheduled to terminate. Following the lifting of the moratorium, the *Hornbeck* plaintiffs continued to evaluate their legal options, but it was generally believed that “this [was] a dispute that [had] run its course.” There was lingering concern, however, that a de facto moratorium remained in place.

Todd Hornbeck (CEO of Hornbeck) stated,

> [The industry hasn’t seen the final requirements for what we would have to do to be able to actually get a permit issued. . . . Until that is done, lifting the moratorium may be just a moot or perfunctory act. . . . I’m skeptical that it will be anytime soon that permits will be issued . . . .]

Critical politicians also exuded skepticism as to the practical effects of lifting the moratorium. These concerns proved to be legitimate. In a later decision on February 2, 2011, stemming from the *Hornbeck* litigation, Judge Feldman stated, “Still . . . no drilling permits have been issued for activities barred by [the moratorium] as of this date.” Indeed, more than a year after the spill, the offshore oil exploration and

---


102 Id., supra note 100.

103 Id. Likewise, the executive director of the Shallow Water Energy Security Coalition warned his colleagues about the practical effect of the ending of the moratorium, telling them that “as soon as they try to pop the champagne bureaucrats will be there to stick the cork back in the bottle.” Gerard Shields, *Deep-Water Drilling Ban Lifted*, BATON ROUGE ADVOC., Oct. 13, 2010, available at 2010 WLNR 204881285.

104 Daly, supra note 100. For example, Louisiana Senator Mary Landrieu placed a hold on a Senate vote to confirm President Obama’s nomination of Jacob Lew for head of the Office of Management and Budget until drilling activity actually resumed. Id.

production industry was only “slowly opening up once more” in the Gulf of Mexico.\textsuperscript{106}

The new regulations instituted “after the spill strengthened safety measures and reduced the risk of another catastrophic blowout.”\textsuperscript{107} But as the Secretary stated, “there will always be risks involved with deep water drilling.”\textsuperscript{108} Because future oil spills remain a likely possibility, it is necessary to clearly define the rights and responsibilities of the government in responding to these spills with blanket, albeit temporary, moratoria or similar regulations. Although the arbitrary and capricious arguments presented in Hornbeck proved to be an effective protection against an improper restriction of property rights, such claims provide better protection against a flawed decision-making process than they do against an unjust decision or result. Future problems may instead arise in circumstances where the government’s decision to implement a moratorium is supported by an adequate Administrative Record, limiting the protection provided by the APA. These situations pose a threat to innocent parties whose property rights are unfairly burdened by that moratorium. Alternatively, there may be situations where the circumstances require a proper moratorium but where notions of justice and fairness nonetheless require some form of compensation to those negatively affected. Accordingly, takings claims should serve to fill this gap in protection, even though historically they have met with little success.

III. THE HORNECK TAKINGS CLAIM IN THE CONTEXT OF TAKINGS CLAUSE JURISPRUDENCE

Although the focus in Hornbeck began with a claim that the moratorium was arbitrary and capricious,\textsuperscript{109} the plaintiffs


\textsuperscript{107} Daly, supra note 100.

\textsuperscript{108} Id. In addition to the Secretary’s acknowledgement of the inherent risk associated with deepwater drilling, a presidential panel reported to President Obama that absent significant reform in both the industry and the government’s policies, a similar oil spill “might well recur.” Wendy Koch, Panel Warns Gulf Oil Spill Could Happen Again, USA TODAY (Jan. 6, 2011), http://content.usatoday.com/communities/greenhouse/post/2011/01/panel-gulf-oil-spill-happen-again/1.

\textsuperscript{109} As previously discussed, see supra Part II.A, the court focused on the arbitrary and capricious allegations despite a takings claim briefly made by the plaintiffs. The court, in fact, did not mention any takings claim. The reason for this omission is unclear.
also briefly asserted that the moratorium constituted an unconstitutional taking of private property. Additionally, after joining the case in July 2010, Diamond Offshore alleged the following:

By virtue of their actions, Defendants have violated Plaintiffs' rights under the Fifth Amendment to the U.S. Constitution. That amendment provides that no person shall suffer a “taking” of private property without due process of law or just compensation. As set forth above, the actions of Defendants herein constitute a taking of Plaintiffs' contract rights without due process of the law for which Plaintiffs seek non-monetary relief.

The relevant provision of OSCLA requires the Secretary to manage the offshore leasing program, stating, “Leasing activities shall be conducted to assure receipt of fair market value for the lands leased and the rights conveyed by the Federal Government.” Certainly, the moratorium prevented leaseholders in the Gulf, even those operating safe rigs, from enjoying the fair market value of their property while the moratorium was in place. Undoubtedly, it can be said that much of this value in terms of access to the oil and gas was restored as soon as the moratorium was lifted, but this fails to account for the fact that entities in the oil industry rely on contracts and equipment that often have a limited lifespan. The industry as a whole was likewise threatened if oil rig operators and their crews chose to take their business to other, more business-friendly waters.

The defendants' only response to the takings allegation was relegated to a footnote claiming that it should be “dismissed... because the second claim for relief, which purports to assert a Fifth Amendment Takings claim, is both

---

110 The Hornbeck takings claim may present procedural questions of ripeness and proper jurisdiction. These issues are outside the scope of this note, which will assume that future plaintiffs would properly address the concerns associated with them. See generally Robert Meltz, Inverse Condemnation and Related Government Liability, SC 43 ALI-ABA 57 (1998) (discussing common problems and obstacles arising in takings claims against the federal government).

111 Plaintiffs' Original Complaint and Application for Temporary Restraining Order and Injunctive Relief, supra note 30, ¶ 33. Before Diamond Offshore joined Hornbeck, Hornbeck more generally alleged that the moratorium infringed on its property rights as protected by the Fifth Amendment but did not specify a takings claim. First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 91.


113 Hornbeck's Memorandum of Law in Support of Its Motion for Preliminary Injunction, supra note 48, at 6; Plaintiffs' Original Complaint and Application for Temporary Restraining Order and Injunctive Relief, supra note 30, ¶ 34.
wholly without merit and outside the jurisdiction of this Court to adjudicate.”114 Yet, it appears that the issue may not be as cut and dry as the defendants asserted. Takings law is extremely unsettled and has been described as “both lacking in theory and unpredictable in application.”115 The United States Supreme Court has acknowledged as much, with Justice John Paul Stevens commenting that “[e]ven the wisest lawyers would have to acknowledge great uncertainty about the scope of this Court’s takings jurisprudence.”116 This means that future court decisions are required to settle the area of takings law and that future plaintiffs have some latitude to persuade these courts to expand the protections afforded by the takings clause. Because future cases may not involve factual circumstances amenable to alternative legal remedies such as the arbitrary and capricious claims presented in Hornbeck, the takings clause could serve as an alternative means for protecting against overly broad and unfairly burdensome regulations.

The Supreme Court has established two main categories of unconstitutional takings: per se takings and regulatory takings.117 The two types require very different analytical approaches to determine whether a taking has occurred.118 According to Tahoe-Sierra Preservation Council, Inc. v. Tahoe Regional Planning Agency,119 Hornbeck must be analyzed under the ad hoc balancing test established for regulatory takings in Penn Central Transportation Co. v. City of New York.120 Traditionally, courts have been hesitant to find a taking under this framework in circumstances similar to Hornbeck.121 But, the ad hoc test courts have been applying is not ad hoc enough. These courts have failed to differentiate between the value taken in the case of land development moratoria (that

---

114 Defendants’ Response to Diamond Offshore’s Motion to Intervene at 3 n.4, Hornbeck Offshore Servs., L.L.C. v. Salazar, 696 F. Supp. 2d 627 (E.D. La. 2010) (No. 10-1663(F)(2)).
116 Id. at 1007 (quoting Nollan v. Cal. Coastal Comm’n, 483 U.S. 825, 866 (1987) (Stevens, J., dissenting)).
118 Id.
temporarily affect the value of real property or a potential home, for example) compared to the value taken by moratoria on a service industry that relies on property rights with limited lifespans. In the former case, property values may continue to rise while the moratorium is in place, or, even if that proves not to be the case, full value should be restored upon rescission of the moratorium. In the latter circumstance, however, moratoria do not simply result in a temporary diminution in the resale value of the property. Instead, any rebound in value after rescission fails to mitigate the significant lost time and investment, potentially resulting in irreparable harm including the destruction of the industry entirely.

Despite the unsettled nature of takings jurisprudence and courts’ reluctance to engage in takings analysis, the more recent additions to the Supreme Court’s takings framework provide the opportunity for enforceable takings claims in these Hornbeck-like circumstances. In the interest of “fairness and justice,” the long-term threat posed by a moratorium and the nature of the property right at issue should be taken into greater consideration by courts as they apply the Penn Central balancing test. Thus, the circumstances that led to Hornbeck provide a tangible example of a claim that—based on the theoretical justifications for temporary takings law presented in Tahoe-Sierra—could be held a taking.

A. Regulatory Takings and the Penn Central Balancing Test

The text of the Takings Clause of the United States Constitution reads as follows: “nor shall private property be taken for public use, without just compensation.” This seemingly straightforward text, however, has proven exceedingly difficult in practice and application. Part of the problem stems from a lack of evidence indicating the Framers’ meaning behind the takings clause or the reasons for including it. Thus, courts have needed to flesh out the meaning and
One of the first cases to distinguish between physical and regulatory appropriation of property was *Mugler v. Kansas*,[129] but it was the rise of the modern government and the case of *Pennsylvania Coal Co. v. Mahon*[130] that really began to define the scope of the clause.[131] *Pennsylvania Coal* involved a mining company that had sold the surface rights of a plot of land but had expressly and contractually reserved the rights to remove any coal found under it.[132] Subsequent to the sale, Pennsylvania passed a statute forbidding the mining of coal that would result in the subsidence of any structure used for human habitation, essentially voiding the contractual reservation.[133] The Supreme Court determined that the statute made it commercially impracticable to mine the coal and had “very nearly the same effect for constitutional purposes as appropriating or destroying it.”[134] Accordingly, the Court held that the statute was invalid because it amounted to a taking without just compensation.[135] In a later case, the Supreme Court would describe *Pennsylvania Coal* as being the “leading case for the proposition that a state statute that substantially furthers important public policies may so frustrate distinct investment-backed expectations as to amount to a taking.”[136] Thus, after *Pennsylvania Coal* it was clear that the Court would deem some legislative acts as going “too far” and rising to the level of a taking of private property.[137] The obvious question that remained before the Court, and a question that still remains unsettled today, is how far is too far?[138] More than fifty years later, in *Penn Central*, the Supreme Court would finally attempt to set forth a framework for answering that question.[139]

---

[128] Id. at 136. Because of the lack of precedent in the area, most of the early Supreme Court decisions accord with early state decisions. Id.


[131] Fee, supra note 115, at 1009.


[135] Id. at 414-15.

[136] *Penn Central*, 438 U.S. at 127 (internal quotation marks omitted).


[138] See Fee, supra note 115, at 1010.

The question presented in Penn Central was whether a city could, as part of a program to preserve historic landmarks, place restrictions on the development of these landmarks without the restriction amounting to a constitutional taking requiring payment of just compensation. The case again addressed the issue of regulatory takings, and the Court implicitly rejected the “proposition that a ‘taking’ can never occur unless government has transferred physical control over a portion of a parcel.” In a landmark decision, the Court established what would become known as the Penn Central balancing test, later described as a “multi-factor test for determining whether a regulation restricting the use of property effects a taking.” The test consisted of a case-by-case analysis of three factors: (1) the economic impact of the regulation on the plaintiff; (2) the extent to which the regulation has interfered with distinct investment-backed expectations; and (3) the character of the governmental action. The Court may have laid out the framework for regulatory takings, but the application of that framework to this day remains disjointed. In fact, critics have condemned it as being both “convoluted and seemingly arbitrary.”

B. Lucas and Its Progeny: Per Se Takings

Although the evolution of takings jurisprudence now makes per se takings inapplicable to the case at hand, it is useful to briefly discuss this second category because previous cases have used such analysis under similar factual

---

140 Id. at 107.
141 Id. at 122 n.25.
142 Stop the Beach Renourishment, Inc. v. Fla. Dep't of Envtl. Prot., 130 S. Ct. 2592, 2603 n.6 (2010).
143 The Court never actually announced a specific set of factors to be used in the balancing test, but these three were “especially prominent” and were again cited by the Court a year later, solidifying their importance in regulatory takings analysis. Joshua P. Borden, Derailing Penn Central: A Post-Lingle, Cost-Basis Approach to Regulatory Takings, 78 GEO. WASH. L. REV. 870, 875-76 (2010).
144 Later cases have changed the word “distinct” to “reasonable.” Christopher Serkin, Existing Uses and the Limits of Land Use Regulations, 84 N.Y.U. L. REV. 1222, 1251 (2009). The Court made this change without acknowledging any distinction between the terms, but the change has had tangible effects on subsequent courts’ analysis of takings claims. See id. “[T]his factor is now principally used to distinguish a property owner’s reasonable expectations from pie-in-the-sky development dreams.” Id.
145 See Penn Central, 438 U.S. at 124.
146 Borden, supra note 143, at 870-71 (exclaiming that “one cannot help but believe that a better, sounder, approach must exist”).
147 See, e.g., Bass Enters. Prod. Co. v. United States, 45 Fed. Cl. 120, 123 (1999) (finding that a per se taking had occurred where government delay in deciding
circumstances. Shortly after instituting the *Penn Central* balancing test, the Supreme Court began to carve out exceptions that required a different analytical framework. These exceptions reflected circumstances where the claimant was automatically entitled to just compensation, without an “inquiry into the public interest advanced in support of the restraint.” The per se takings exception to the *Penn Central* balancing test applies to the following: (1) takings that amount to a physical occupation of the property by the government and (2) regulations that result in the total loss of value of the property. The case that defined the first exception was *Loretto v. Teleprompter Manhattan CATV Corp.*, where the Court held that permanent physical occupation of property—even in the case of two measly four-inch by four-inch metal cable boxes—is always a taking, “without regard to whether the action achieves an important public benefit or has only minimal economic impact on the owner.” The Court justified its holding by noting that a permanent physical occupation destroys many strands from the “bundle of property rights” that have historically been protected by property law, including the right to possess, use, dispose, and exclude. It also noted that a balancing test was largely unnecessary because such cases present “relatively few problems of proof” compared with regulatory takings.

The second category of per se exceptions to the *Penn Central* balancing test—those regulations that deny all economically beneficial or productive use of the land—was defined in *Lucas v. South Carolina Coastal Council*. The Court noted that these types of regulations “carry with them a heightened risk that private property is being pressed into some form of public service under the guise of mitigating serious public harm.” Further, like permanent physical occupations, the Court explained,

---

149 Id.
152 Id. at 437.
153 See generally Lucas, 505 U.S. 1003. The plaintiff in Lucas, an owner of two undeveloped parcels of land, argued that by passing a law barring the erection of any permanent habitable structures on certain beachfront property, the state legislature had taken his property and that he was thus entitled to compensation. Id. at 1009.
154 Id. at 1018.
We believe similar treatment must be accorded confiscatory regulations, i.e., regulations that prohibit all economically beneficial use of land: Any limitation so severe cannot be newly legislated or decreed (without compensation), but must inhere in the title itself, in the restrictions that background principles of the State’s law of property and nuisance already place upon land ownership.\textsuperscript{155}

Thus, “[w]hen . . . a regulation that declares ‘off-limits’ all economically productive or beneficial uses of land goes beyond what the relevant background principles [of property and nuisance law] would dictate, compensation must be paid to sustain it.”\textsuperscript{156}

C. Temporary Takings and Moratoria: The Importance of Tahoe-Sierra

The fact that a moratorium lasts for only a limited period of time\textsuperscript{157} is not fatal to a regulatory takings claim.\textsuperscript{158} Instead, the duration of the restriction is only one factor for courts to consider.\textsuperscript{159} The Supreme Court has held that the “effect of a regulation must be measured against the ‘parcel as a whole,’” however the Court has failed to fully define this term.\textsuperscript{160} Earlier cases usually focused on the amount of physical or spatial portions of a land parcel impaired by a restriction to determine if a taking had occurred.\textsuperscript{161} The Supreme Court in

\textsuperscript{155} Id. at 1029.

\textsuperscript{156} Id. at 1030. The Court attempted to lay out a clear framework for per se takings analysis, stating:

The “total taking” inquiry we require today will ordinarily entail . . . analysis of among other things, the degree of harm to public lands and resources, or adjacent private property, posed by the claimant’s proposed activities, the social value of the claimant’s activities and their suitability to the locality in question, and the relative ease with which the alleged harm can be avoided through measures taken by the claimant and the government (or adjacent private landowners) alike.

\textsuperscript{157} Id. at 1030-31 (citations omitted).

\textsuperscript{158} In Hornbeck, the moratorium lasted less than six months, although at the time the claim was originally filed, there was concern that it might last much longer. First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 54.


\textsuperscript{160} Id. at 342.

\textsuperscript{161} Fee, supra note 115, at 1029-30.

\textsuperscript{153} See Glenn P. Sugameli, Takings Law Symposium: Lucas v. South Carolina Coastal Council: The Categorical and Other “Exceptions” to Liability for Fifth Amendment Takings of Private Property Far Outweigh the “Rule,” 29 ENVTL. L. 939, 948-53 (1999) (discussing how various cases have dealt with the fact that regulations “are three dimensional [and] have depth, width, and length” (quoting First English Evangelical
Tahoe-Sierra, however, focused on a different slice of the property—the temporal dimension affected by the regulation. Specifically, the Court addressed the issue of how to analyze a temporary land development moratorium and how to determine if such a regulation could ever amount to a taking. Unfortunately, because the Court answered the question by responding “neither ‘yes, always’ nor ‘no, never,’” the analysis is not exactly straightforward. The Court’s description of and focus on “fairness and justice” as central tenants of takings analysis, though, should provide at least some future takings victims the opportunity to bring a successful claim.

The Court in Tahoe-Sierra built upon the principle established in First English Evangelical Lutheran Church v. County of Los Angeles, where the Court previously addressed temporary takings in general. The Court in First English held, “[W]here the regulation has already worked a taking of all use of property, no subsequent action by the government can relieve it of the duty to provide compensation for the period during which the taking was effective.” In that case, though, the Court was concerned with determining compensation once a taking had concededly occurred, and thus did not specifically address the threshold question of whether the temporary denial of land use had constituted a taking in the first place. In Tahoe-Sierra, the Court took that first step and established the framework to be used for analyzing temporary moratoria in the takings context. The Court rejected the petitioners’ claim that a temporary taking that resulted in the total deprivation of all economic use while the moratorium was in place was

Lutheran Church v. Cnty. of L.A., 482 U.S. 304 (1987)). At least in terms of physical restrictions, landowners are not permitted to conceptually sever individually affected segments from the property as a whole in order to argue that the use or value of these segments has been totally destroyed by the regulation. Fee, supra note 126, at 1030. Yet, courts have sometimes struggled in determining what the relevant parcel to be analyzed is in order to determine both whether a taking has occurred, and if so, how much compensation is required. See Melz, supra note 110, at 72; Fee, supra note 126, at 1029-32.

Tahoe-Sierra, 535 U.S. at 331-32.

Id.

Id. at 321.


See generally First English, 482 U.S. 304.


Tahoe-Sierra, 535 U.S. at 328.

See id. at 334-38.
subject to the per se takings analysis established in *Lucas*.170 The Court worried that such a categorical rule would open the floodgates to takings litigation and would apply to even “normal delays in obtaining building permits . . . as well as to orders temporarily prohibiting access to crime scenes.”171 Rather, the Court stated that Justice Brennan’s “parcel as a whole” theory of takings analysis established in *Penn Central* must be applied to the temporal dimensions of a property the same way it applies to the physical dimensions.172 Accordingly, the Court held that the “better approach to claims that a regulation has effected a temporary taking requires careful examination and weighing of all the relevant circumstances” by using the *Penn Central* balancing test.173 But, important to a *Hornbeck*-type scenario, the Court made it clear that “[i]n rejecting petitioners’ per se rule, we do not hold that the temporary nature of a land-use restriction precludes finding that it effects a taking.”174 Instead, “the duration of the restriction is one of the important factors that a court must consider in the appraisal of a regulatory takings claim . . . .”175 Further, throughout the opinion, the Court repeatedly referenced notions of fairness and justice as instructive to the holding.176

The question of how long a restriction is too long is one that remains unclear and confusing. For instance, courts have held, “A permanent physical occupation does not necessarily mean a taking unlimited in duration. [Instead, a] ‘permanent’ taking can have a limited term.”177 Interestingly, many temporary regulations held to not constitute takings at all are longer in duration than the “permanent” deprivation that required application of the per se rule in *Lucas*.178 One thing is

---

170 Id. at 334.
171 Id. at 335 (citations omitted).

> “Taking” jurisprudence does not divide a single parcel into discrete segments and attempt to determine whether rights in a particular segment have been entirely abrogated . . . [T]his Court focuses rather both on the character of the action and on the nature and extent of the interference with rights in the parcel as a whole . . . .

*Id.*

173 *Tahoe-Sierra*, 535 U.S. at 335 (internal quotation marks omitted).
174 Id. at 337.
175 Id. at 342.
176 See *id.* at 321, 332-36; see also Eagle, *supra* note 165, at 505.
177 Eagle, *supra* note 165, at 456 (quoting Skip Kirchdorfer, Inc. v. United States, 6 F.3d 1573, 1482 (Fed. Cir. 1993)).
178 See *id.* The delay in development in *Lucas* lasted only two years. *Id.*
clear, however: courts have tended to put great weight on the
durational factor of temporary takings.\footnote{See Bass Enters. Prod. Co. v. United States, 381 F.3d 1360, 1366 (Fed. Cir. 2004) (noting that an extraordinary delay in governmental decision making may constitute a taking, but citing delays of eight years, seven years, and forty-five months that were not held to be takings).}

The impact of *Tahoe-Sierra* on takings jurisprudence
cannot be overemphasized. For example, its effect can clearly
be discerned from the two opposing decisions reached by the
United States Court of Federal Claims in *Bass Enterprises
Production Co. v. United States*.\footnote{Compare id., with Bass Enters. Prod. Co. v. United States, 45 Fed. Cl. 120 (1999).}
The litigation commenced in *Bass Enterprises*
after the Bureau of Land Management denied
the plaintiff lessees’ application to explore and drill for oil and
gas on their leased property for forty-five months while the
Environmental Protection Agency (EPA) determined the
environmental impact that such development posed.\footnote{Bass Enters. Prod. Co., 381 F.3d at 1362-64.}
The Court of Federal Claims initially found a temporary (but per
se) taking, stating, “Plaintiffs have not been permitted to use
their leases for a substantial period of time. Their loss during
that period was absolute.”\footnote{Bass Enters. Prod. Co., 45 Fed. Cl. at 123.}
The court cited *Lucas*, explaining
that the limited duration of the regulation did not bar
constitutional relief, and ordered damages in the amount of the
interest that the plaintiffs would have earned on the oil and
gas profits during the delay period.\footnote{Id. at 123-24.}
Following the Supreme
Court’s decision in *Tahoe-Sierra*, however, the government’s
motion for reconsideration was granted.\footnote{Id. at 1364.}
This time the Court of Federal Claims applied a *Penn Central* balancing test and
found that “the economic impact on Bass was de minimis and
that the Government’s delay was reasonable given the
importance of protecting the public . . . .”\footnote{Id.}
After “[w]eighing the factors and the circumstances surrounding the delay as a
whole,” the court concluded that there had not been a taking.\footnote{Id. at 1365.}

*Bass Enterprises* is an example of the heavy emphasis
courts have placed on the duration of the delay when applying
a temporary takings analysis following *Tahoe-Sierra*.\footnote{See id. at 1366-69.}

Accordingly, under a similar judicial application, Hornbeck,
plaintiffs similarly situated, may have difficulty bringing a successful takings claim despite the significant value that was lost—and which could not be recovered—during and following the moratorium. Still, this result may not be fair or justified based on takings law as it currently stands, and it therefore warrants a closer look. Additionally, while there are many factual similarities between Bass Enterprises and the circumstances leading to Hornbeck, there are enough distinctions to warrant further analysis. To begin this analysis, it is useful to look at how courts have previously treated takings claims in cases following moratoria on oil drilling. This examination will illustrate how the evolution of takings jurisprudence requires a different outcome here than was reached in both Bass Enterprises and these previous moratoria cases.

IV. PREVIOUS CASES INVOLVING MORATORIA BY THE FEDERAL GOVERNMENT FOLLOWING OIL SPILLS

Moratoria have been used by administrative agencies as a common means to “preserve the status quo while formulating a more permanent . . . strategy.” Moratoria have been employed in a wide variety of contexts, from use of the death penalty, to the prohibition of killing marine mammals, to the suspension of mining of valuable fossil fuels. Likewise, moratoria and regulations suspending operations involving gas, oil, and mineral rights have been the subjects of takings analysis in the past. In fact, the Deepwater Horizon incident is not the first time that the Department of the Interior has

---

188 See infra Part V.B.
192 See Sam Kalen, The Devolution of NEPA: How the APA Transformed the Nation's Environmental Policy, 33 WM. & MARY ENVTL. L. & POL'Y REV. 483, 524 (2009) (discussing the Department of the Interior's informal moratorium on new coal mining leases and permits until “the Department could develop a coherent approach to the leasing and development of the nation's coal resources”).
broadly suspended drilling rights via moratorium following an oil blowout.²⁹⁴ Nor is it the first time that—in response—a takings claim was brought by an aggrieved plaintiff.²⁹⁵

The massive oil spill in Santa Barbara, California, in January 1969 spurred several cases that explored the property rights of leaseholders in the Santa Barbara Channel.²⁹⁶ The spill “caused severe property and environmental damage” and prompted the Secretary of the Interior to order all companies in the Santa Barbara Channel to cease all drilling and production regardless of their involvement in the spill.²⁹⁷ This line of cases focused on the authority of the Secretary to pass regulations that amounted to a taking and determined that the duration of the regulation was a requisite factor in determining that authority.²⁹⁸

In *Gulf Oil Corp. v. Morton*, the Secretary of the Interior followed up the initial moratorium with a second one in 1971 that was to last until 1973.²⁹⁹ He justified the second suspension by stating that it was necessary to give Congress time to consider whether it wished to pass legislation to terminate the leases in the interest of conservation.³⁰⁰ The plaintiffs in the case—who had paid $153 million for the leases—filed suit, seeking, among other things, declaratory judgment that the suspension was outside the scope of the Secretary’s authority and must be revoked.³⁰¹ In addressing the first suspension the court declared that “[a]fter the leases in question were made, events occurred in the Santa Barbara Channel that were both unexpected and very dangerous to the environment . . . [causing] the Secretary to reconsider the dangers to the natural resources of the area if drilling were to proceed under the leases.”³⁰² As was the case in *Hornbeck*, the case focused on whether the suspension was arbitrary and capricious and contrary to the

---

²⁹⁴ See, e.g., *Pauley Petroleum*, 591 F.2d at 1312.
²⁹⁵ See, e.g., *Union Oil Co. of Cal. v. Morton*, 512 F.2d 743, 750 (9th Cir. 1975).
²⁹⁶ See generally *Gulf Oil Corp. v. Morton*, 493 F.2d 141 (9th Cir. 1973); *Union Oil*, 512 F.2d at 743; *Pauley Petroleum*, 591 F.2d 1308.
²⁹⁷ *Pauley Petroleum*, 591 F.2d at 1312.
²⁹⁸ See generally *Gulf Oil*, 493 F.2d 141; *Union Oil*, 512 F.2d 743; *Pauley Petroleum*, 591 F.2d 1308.
²⁹⁹ *Gulf Oil*, 493 F.2d at 143.
³⁰⁰ *Id.* The decision was supported by the director of the United States Geological Survey, who noted that continued operation under the leases posed the following risks: “[T]he possibility of another blowout; the possibility that wells would be improperly plugged . . . should [they later be abandoned]; [and] the possibility that geologic structures such as the one which contributed to the 1969 spill would be encountered and fractured, thus causing large quantities of oil and gas to escape.” *Id.*
³⁰¹ *Id.*
³⁰² *Id.* at 146.
APA; the court, however, also touched on several aspects associated with a takings analysis, including just compensation.\textsuperscript{203} In addressing a possible takings claim, the court noted that in a letter accompanying the proposed bill to terminate the leases, the Secretary wrote that the legislation would “offer[] a mechanism for determining and paying just compensation to the lessees . . . .”\textsuperscript{204} The court held that “Congress authorized the Secretary to suspend operations under existing leases whenever he determines that the risk to the marine environment outweighs the immediate national interest in exploring and drilling for oil and gas.”\textsuperscript{205} The court also held that, since the lessees had not yet begun drilling, the circumstances at hand permitted a suspension while Congress weighed the merits of the proposed bill.\textsuperscript{206} The court cautioned, however, that the Secretary could not “continue to issue comparable orders one after another and justify them by repeatedly having his proposed legislation introduced in the Congress . . . at some point, if Congress does not act, there must be an end to the matter.”\textsuperscript{207}

The same facts that led to \textit{Gulf Oil} spurred litigation in \textit{Union Oil Co. of California v. Morton.}\textsuperscript{208} The Secretary of the Interior had initially granted plaintiffs the right to build a new floating drilling platform.\textsuperscript{209} But, following the spill, the Secretary announced that Union Oil would not be permitted to build the additional platform.\textsuperscript{210} Unlike \textit{Gulf Oil}, which discussed a temporary suspension, the issue in this case was whether the Secretary had the power to permanently suspend the plaintiffs’ lease.\textsuperscript{211} The plaintiffs claimed that the Secretary had denied them the full exercise of their rights under their lease with the federal government\textsuperscript{212} and that the suspension amounted to a permanent taking without compensation.\textsuperscript{213} The

\textsuperscript{203} See id. at 146-47.
\textsuperscript{204} Id. at 147.
\textsuperscript{205} Id. at 144.
\textsuperscript{206} Id. at 146.
\textsuperscript{207} Id. at 148. Indeed, on petition for rehearing the court held that the bill had been before Congress for four sessions without any substantial action to push the bill forward towards law and that therefore the Secretary’s power to suspend the leases vanished on October 18, 1972. Id. at 149.
\textsuperscript{208} 512 F.2d 743 (9th Cir. 1975).
\textsuperscript{209} A provision in plaintiff’s lease specifically allowed for the erection of floating drilling platforms. Id. at 746.
\textsuperscript{210} Id.
\textsuperscript{211} See id. at 751.
\textsuperscript{212} Id. at 746.
\textsuperscript{213} Id. at 750.
court acknowledged that, while the Secretary had the authority to suspend the lease under certain circumstances, the executive branch had no intrinsic power of condemnation and thus could not suspend the lease indefinitely. Further, a suspension that “deprived Union of all benefit from the lease in that particular area” would be a permissible taking if enacted by Congress (as long as just compensation was provided), but was outside the scope of the Secretary’s power. Accordingly, the court analyzed whether the Secretary had impermissibly taken Union’s property by determining whether the suspension was temporary or was instead an indefinite suspension amounting to a “pro tanto cancellation of [the] lease.” A suspension that was limited in time by the “occurrence of new events or the discovery of new knowledge which can be anticipated within a reasonable period of time” would not constitute a taking, according to the court. Ultimately, the court determined that the facts were insufficient on the record to determine the answer to that question and remanded the case to the district court.

The Santa Barbara oil spill also led to Pauley Petroleum Inc. v. United States. In that case, the plaintiffs were a consortium of oil corporations who acquired leases from the federal government to explore and drill off the coast of Santa Barbara. In conjunction with the general moratorium, the Secretary promulgated a regulation relating to the level of liability for general lessees involved in oil spills. The Secretary also required all drilling companies in the area to submit “all geological, geophysical and structural information,” explaining that after this information was studied on a lease-by-lease basis the companies would be permitted to resume drilling. The plaintiffs brought suit claiming, among other things, that the absolute liability requirement and clearance program amounted to a regulatory taking because it rendered their leases “economically worthless and exposed them to
unmeasurable risks." They argued that the regulation applied retroactively and imposed absolute liability for all cleanup costs as well as for any damage to third-party property caused by a spill. The court held that the power to suspend leases could not rise to such a level that it resulted in a total suspension of rights because the Secretary does not have the authority to take property. As was the case in Union Oil, the court determined that the plaintiffs had "fail[ed] to meet at least one of the prerequisites for a constitutional taking—the requirement that the taking be authorized by Congress." Thus, the plaintiffs could not bring a takings claim because the power to take property was solely vested in Congress, and did not extend to actions by the Secretary. According to the court, "Congress clearly did not intend to grant leases so tenuous in nature that the Secretary could terminate them, in whole or in part at will." Further, the court noted that the "short, temporary suspension was plainly not so severe a property deprivation as to constitute a fifth amendment taking." Accordingly, the court held that a takings claim was not the proper means of recourse for the plaintiffs. Again the court required a regulation to be permanent and authorized by Congress for it to afford a valid takings claim.

Many of these cases share similar factual circumstances to the Hornbeck scenario. Despite this precedent, however, the outcome of a takings claim in Hornbeck, or future claims arising under similar circumstances, remains unclear. The courts in the Santa Barbara spill cases held that takings claims were precluded because the Secretary lacked authority to restrict property rights in this manner, often hinging that authority on the duration or permanence of the suspension. This position is outdated, however, in light of subsequent precedent. The Hornbeck takings claim need not hinge on the

---

224 Id. at 1314.
225 Id. at 1313.
226 Id. at 1326.
227 Id.
228 Id. (citing Union Oil Co. of Cal. v. Morton, 512 F.2d 743, 751 (9th Cir. 1975)).
229 Id. (citing Union Oil, 512 F.2d at 751).
230 Id. at 1327.
231 The court also dismissed the plaintiffs' other causes of action, which included breach of contract and mutual mistake claims. Id. at 1326 (citing Union Oil, 512 F.2d at 751).
232 Id.
233 See, e.g., Union Oil, 512 F.2d at 751.
234 See David W. Spohr, "What Shall We Do with the Drunken Sailor?: The Intersection of the Takings Clause and the Character, Merit, or Impropriety of
permanent versus temporary distinction because the duration of the regulation is now only a part of the more fully established regulatory takings framework. And any notion that only actions by Congress, and not those by a member of the executive branch, can amount to a taking is similarly misplaced in today's analysis. Instead, “the Takings Clause bars the State from taking private property without paying for it, no matter which branch is the instrument of the taking.” Accordingly, it is necessary to analyze the Hornbeck claim under the more modern approach to temporary regulatory takings. Although the Supreme Court has attempted to clarify and establish a working framework for this area, Hornbeck illustrates how the unsettled state of takings jurisprudence has made some plaintiffs unwilling to bring these kinds of claims and courts hesitant to venture into such an analysis if provided an alternative. It is precisely because this area is still malleable, though, that plaintiffs and courts should look to the takings clause to protect worthy victims from regulations that go too far.

V. AN AD- HOC ANALYSIS OF THE TAKINGS CLAIM IN HORNBECK

A court may be hesitant to find a takings claim in Hornbeck, despite the expansion of takings jurisprudence since the Santa Barbara oil spills and facts that distinguish Hornbeck from Bass Enterprises. Such a result, however, is both unfortunate and undesirable, and should be modified in light of the concepts of fairness and justice discussed in Tahoe-Sierra. Admittedly, there is still confusion and discord in takings law, but a proper reading of Tahoe-Sierra illustrates

Regulatory Action, 17 S.E. ENVTL. L.J. 1, 21 (2008) (noting that “[a] legislative grant of authority to an agency to exercise discretion over various affairs could lead to acts by [executive] officials that were illegal but still within the [congressionally defined] scope of agency authority” (citing Del-Rio Drilling Programs, Inc. v. United States, 146 F.3d 1358 (Fed. Cir. 1998)); see also Tahoe-Sierra Pres. Council, Inc. v. Tahoe Reg’l Planning Agency, 535 U.S. 302, 342 (2002) (holding that the duration of a regulation is not dispositive regarding the takings question).

235 Tahoe-Sierra, 535 U.S. at 335.
236 The Takings Clause . . . is not addressed to the action of a specific branch or branches. It is concerned simply with the act, and not with the government actor . . . .” Stop the Beach Renourishment, Inc. v. Fla. Dep’t of Envtl. Prot., 130 S. Ct. 2592, 2601 (2010) (discussing a judicial takings claim).
237 Id. at 2602.
238 See Meltz, supra note 110, at 73 (describing confusion as to when cases should be analyzed as a breach of contract instead of a taking and noting the general preference of the federal courts to take the breach of contract route if possible).
239 535 U.S. at 333-34.
the “considerable latitude” available to property owners to assert claims based on these concepts.\footnote{Eagle, supra note 165, at 505.} The facts in Hornbeck provide an example of the danger for plaintiffs in a service industry who may be injured uniquely and significantly because of the permanent effects of a temporary regulation. Accordingly, these types of claims are worthy of relief under a more contextually inclusive three-part \textit{Penn Central} analysis that looks at (1) the economic impact of the regulation; (2) the distinct investment-backed expectations of the plaintiff; and (3) the character of the government regulation.\footnote{See \textit{Penn Cent. Transp. Co. v. City of New York}, 438 U.S. 104, 124 (1978).}

\section{The Economic Impact of the Regulation}

The economic impact is a fact-specific question, and in the case of the Hornbeck plaintiffs it is immense in terms of sheer numbers. The plaintiffs alleged that the moratorium led to the termination of valuable service contracts with the oil and exploration companies in the Gulf of Mexico.\footnote{First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, \textit{supra} note 27, ¶¶ 58-67. The allegations were not mere conjecture; as early as June 2010 valuable contracts were being cancelled or reduced. \textit{See id.}} Plaintiffs noted that lost job wages could be more than $330 million for each month the moratorium remained in place.\footnote{Hornbeck’s Memorandum of Law in Support of Its Motion for Preliminary Injunction, \textit{supra} note 48, at 23.} Likewise, the Bureau of Ocean Energy Management (BOEM) predicted “that the moratorium would lead to the loss of more than 23,000 jobs in the . . . region and that oil and gas industry spending in [the Gulf would] be reduced by more than $10 billion.”\footnote{Plaintiff’s Memorandum in Opposition to Defendants’ Cross-Motion for Partial Summary Judgment at 14 n.13, \textit{Enesco Offshore Co. v. Salazar}, 2010 WL 3973222 (E.D. La. Sept. 9, 2010) (No. 10-CV-01941). It was also reported that the moratorium would result in approximately $500 million in lost wages for workers in the oil industry. Steven Shavell, \textit{Should BP Be Liable for Economic Losses Due to the Moratorium on Oil Drilling Imposed After the Deepwater Horizon Accident?}, 64 \textit{VAND. L. REV.} 1995, 1999 (2011).}

The leases,\footnote{Complaint at ¶ 55(b), \textit{Enseo Offshore Co. v. Salazar}, 2010 WL 2812241 (E.D. La. July 9, 2010) (No. 10-CV-01941).} contracts, and vessels that constitute the property rights in the industry last for only a limited time, and thus it is difficult or impossible to recoup any lost expenses as a result of suspensions of operations.\footnote{Hornbeck’s Memorandum of Law in Support of Motion for Preliminary Injunction, \textit{supra} note 48, at 3.} These effects can be distinguished from those imposed by moratoria in other
contexts, and make the plaintiffs more worthy of relief considering the immense level of investment that they have made into both their own equipment and the industry as a whole in the region. Additionally, unlike in other contexts, the plaintiffs in Hornbeck do not enjoy any reciprocity of advantage that might mitigate the damage suffered when they are subjected to a stop-drilling order. Instead, for plaintiffs, and similarly situated companies, the harm suffered by moratoria is without any potentially positive consequences.

Also, the final economic impact may be even greater than initial estimates suggested. The BOEM estimates were based on a six-month moratorium, but at the time the complaint was filed the moratorium was predicted to last much longer. Additionally, the Department of the Interior estimated that long-term job loss might significantly exceed BOEM’s initial estimates due to the potential relocation of rigs to other regions and the potential demise of some drilling companies. Those fears did not fully materialize (the moratorium did in fact end before six months), yet the lifting of the moratorium did not bring about an immediate end to the negative economic effects it created. Thus, the impact of the

247 See infra notes 271-74 and accompanying text.
248 Hornbeck’s Memorandum of Law in Support of Motion for Preliminary Injunction, supra note 48, at 3-6. Regarding investments put into their own equipment, Diamond Offshore explained, for example, that the cost of one of its floating rigs can exceed $500,000 per day. Plaintiffs’ Original Complaint and Application for Temporary Restraining Order and Injunctive Relief, supra note 30, ¶ 10. Likewise, the value of the deepwater vessels built by plaintiff Bollinger Shipyard Company in the last five years exceeds $200 million. First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 38.
249 See Tahoe-Sierra Pres. Council, Inc. v. Tahoe Reg’l Planning Agency, 535 U.S. 302, 341 (2002). The Supreme Court in Tahoe-Sierra noted that all landowners in the community benefited from the moratorium because the development restriction also prevented their neighbors from engaging in unwanted development. Id. In fact, in that case, the moratorium might protect or even increase the property values in the area. Id.
250 First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, supra note 27, ¶ 54.
251 Plaintiff’s Memorandum in Opposition to Defendants’ Cross-Motion for Partial Summary Judgment at 14 n.13, Enseo Offshore Co. v. Salazar, 2010 WL 3973222 (E.D. La. Sept. 9, 2010) (No. 2:10-CV-01941-MLCF-JCW); Hornbeck’s Memorandum of Law in Support of Motion for Preliminary Injunction, supra note 48, at 6 (explaining that “when a deepwater rig or vessel leaves a drilling region, it does so under a long-term contract and will not return for several years, if ever”). At least two deepwater rigs had left for foreign waters as early as July 2010. Hughes & Power, supra note 25. But see Daly, supra note 100 (noting that loss of jobs in the Gulf region is likely temporary).
252 People in the industry feared the possibility that a de facto moratorium would remain in place, which would extend the suspension of operations and increase the injury. See Matthew Daly, Administration Lifts Freeze on Drilling, Official Says New Rules Improved Safety, Cut Risks of Another Disaster, CHARLESTON DAILY MAIL,
regulation could potentially be felt far into the future. More importantly, future moratoria could last significantly longer, thus resulting in an even greater impact and illustrating the need for takings clause protection.

B. Reasonable Investment-Backed Expectations

The reasonable investment-backed expectations of plaintiffs affected by the moratorium provide several distinguishing characteristics from previous precedent that help support the finding of a taking. Whether a regulation results in a change to the status quo is an important factor for the Supreme Court in a takings analysis. For example, the Penn Central Court “hinted that it may have [ruled differently] if the . . . regulation had prevented [the property] from being used as it always had been.” Thus, it is more reasonable for parties to expect their permits and contracts to continue unhindered once they have obtained all necessary authorizations and drilling has already commenced. The plaintiffs here are not seeking to develop property beyond its current state; rather, they simply seek to continue using their property as they have for many previous years. Similarly, they are not seeking an initial permit to begin utilization of their property by exploring and drilling for oil, as was the case in Bass Enterprises. Accordingly, while the delays in both Bass Enterprises and Hornbeck stemmed largely from the need to protect against a “possible environmental and health hazard,” only the Bass Enterprises delay resulted in an extension—rather than a disruption—of the status quo.

The facts in Bass Enterprises represent a situation where the plaintiffs had more reason to expect delays than do the plaintiffs in Hornbeck. In Bass Enterprises, the lease in question was above a portion of an underground nuclear waste storage site. The delay was imposed while the EPA

Oct. 13, 2010; see also Hornbeck Offshore Servs., L.L.C. v. Salazar, No. 10-1663, 2011 WL 454802, at *3 (E.D. La. Feb. 2, 2011). In fact, several months after the moratorium was lifted, drilling permits were still not being issued. Id. at *2.

ELLICKSON & BEEN, supra note 127, at 197. Further, the Supreme Court noted that the type of moratoria at issue in Tahoe-Sierra was one that is routinely used to preserve, not disrupt, the status quo. Tahoe-Sierra, 535 U.S. at 337.


Id. at 1367. The court in Bass Enterprises noted that such a situation justifies delay because “we do not want to ‘encourage hasty decisionmaking’ by the Government.” Id.

Id. at 1362.
determined if it was necessary to condemn and obtain the leasehold in its entirety to protect the integrity of the site. Thus, it was unknown whether any safe use of the lease could ever be engaged in by the leaseholder. In contrast, the leases and contracts in *Hornbeck* were for rigs that had recently passed inspections and had been deemed to conform to all safety regulations by the MMS. Also, there had been no similar blowouts in the Gulf of Mexico since 1969. Thus, there was little reason to anticipate another spill. Instead, it was more reasonable to expect that as long as the parties continued to comply with all safety regulations they would not suffer any delays.

Likewise, although “delay is inherent in complex regulatory . . . schemes,” such a delay is more reasonably expected when operations have not yet commenced. The decision to allow drilling on a portion of a site designated for nuclear waste storage required careful deliberation to determine the effects on the site as a whole. The *Bass Enterprise* plaintiffs had knowledge that these regulations were in place before they began any drilling activities. Thus, they should have expected that they might not immediately be granted the use of their lease, and such a delay would do nothing to alter the status quo. Like the disposal of nuclear waste, the oil industry is highly regulated; still, the oil rigs in *Hornbeck* had conformed to these regulations and procedural requirements and had previously been operating safely and without issue. There was little reason for the plaintiffs to expect any change to these operations, barring individual safety violations by the rig operators. Thus, by banning the operation of all oil rigs, the broad moratorium seriously disrupted the status quo, unfairly burdened innocent parties,

---

257 Id. at 1363.

258 Additionally, the *Bass Enterprise* plaintiffs had the security of knowing they would be compensated if the government determined drilling to be permanently unsafe because the federal government would then condemn the property and pay just compensation. See id. at 1362-63.

259 First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, *supra* note 27, ¶ 46. Twenty-seven of twenty-nine recently inspected rigs had no violations; the other two had only minor violations. Id.


261 *Bass Enters. Prod. Co.*, 381 F.3d at 1366 (quoting *Wyatt v. United States*, 271 F.3d 1090, 1098 (Fed. Cir. 2001)).

262 Takings are generally less likely to be found in industries that are highly regulated. Meltz, *supra* note 110, at 66. This alone does not bar all takings claims though. *Id.* at 67.

263 First Supplemental and Amended Complaint for Declaratory and Injunctive Relief, *supra* note 27, ¶ 46.
and destroyed the investment-backed expectations of the oil service industry in the Gulf.

Further, in determining the safety of oil exploration and drilling, Congress has provided an expectation that the Department of the Interior will proceed expeditiously. The relatively short time frame of thirty days that Congress has allowed for the agency to approve exploration plans submitted under OSCLA supports a reasonable expectation of freedom from extended delay in the industry. Accordingly, it is reasonable for plaintiffs with property interests in the oil industry in the Gulf of Mexico to expect government-imposed delay not to last much longer than thirty days. The length of the moratorium here far exceeded that expectation.

A similar argument draws support from dicta in Tahoe-Sierra. In acknowledging that considerations of “fairness and justice” were the touchstone of moratoria aimed at curbing abusive land development, Justice Stevens indicated that, under the right circumstances, the Court could “craft a narrower rule that would cover all temporary land-use restrictions except those ‘normal delays’ associated with the industry or application process.” There is little precedent for what constitutes a normal delay as it applies to moratoria on drilling following an oil spill. Oil drilling resumed under heightened standards following the Santa Barbara spill, however, after only about two months. Therefore, it is reasonable for the plaintiffs to expect that their investment in the Gulf would not be disturbed by the government for any substantially longer duration.

In contrast, a takings argument is weakened by the fact that the moratorium officially lasted less than six months, a very short period of time compared with other delays that have

---

264 See In re Core Commc’ns., Inc., 531 F.3d 849, 855 (D.C. Cir. 2008).
267 See Pauley Petroleum Inc. v. United States, 591 F.2d 1308, 1312, 1314 (Ct. Cl. 1979). The Secretary immediately requested that drilling operations cease when the spill occurred on January 28, 1969. Id. at 1312. The Secretary gave approval to recommence operations on April 1, 1969. Id. at 1314.
268 Although the moratorium officially lasted only about five months, the added regulations imposed by the government prevented operations from resuming for longer. Daly, supra note 100; see also supra note 252 and accompanying text.
been held to not be takings.\textsuperscript{269} Although the economic impact was drastic and potentially devastating to companies in the oil drilling and exploration business in the Gulf of Mexico, these companies should be able to regain most of the value of their leases in time, now that the moratorium has been lifted. Yet, for those service companies that rely on other entities that have left for foreign waters, this reassurance is not nearly as comforting.\textsuperscript{270} Moreover, the Tahoe-Sierra Court justified its refusal to find the temporary restriction in that case to be a taking by explaining, “[l]ogically, a fee simple estate cannot be rendered valueless by a temporary prohibition on economic use, because the property will recover value as soon as the prohibition is lifted.”\textsuperscript{271} This assertion may be true in the case of temporary planning/development restrictions on real estate (as was the restriction at issue in Tahoe-Sierra), but it fails to account for the significant loss in value imposed in the present situation. Unlike the Tahoe-Sierra landowner, many companies in the oil industry rely on property interests that amount to less than a fee simple estate. Instead, their property interests are vested in contracts\textsuperscript{272} and support vessels that have a limited useful life.\textsuperscript{273} Accordingly, the higher the percentage of these useful lives consumed by the moratorium, the more likely a court should be to find a taking.\textsuperscript{274}

Finally, the Court in Tahoe-Sierra articulated that the petitioners had failed to offer a persuasive account of why moratoria should be treated differently from ordinary permit delays.\textsuperscript{275} Further, the Court condemned the petitioners’ claim

\textsuperscript{269} See Bass Enters. Prod. Co. v. United States, 381 F.3d 1360, 1366 (Fed. Cir. 2004). The Supreme Court has indicated that while the length of the moratorium is not dispositive, the longer the delay, the more likely a taking has occurred. See Tahoe-Sierra, 535 U.S. at 341 (“It may well be true that any moratorium that lasts for more than one year should be viewed with special skepticism.”).\textsuperscript{271} Only two of the thirty-three deepwater oil rigs left the Gulf for other fields, but this result could have been much worse if the first moratorium had not been enjoined in Hornbeck. See Cynthia A. Drew, The Gulf Deepwater Drilling Moratorium: While Merits Still Pending, Already Significant Practical Effect?, ENVTL. L. REP. (2010), available at LEXIS, 40 ELR 11137. Future plaintiffs who do not have the ability to enjoin regulations under alternative means may not be as lucky. See supra Part IV.\textsuperscript{275} Tahoe-Sierra, 535 U.S. at 332.\textsuperscript{272} Complaint at ¶ 1, Ensco Offshore Co. v. Salazar, 2010 WL 2812241 (E.D. La. July 9, 2010) (No. 10-CV-01941).\textsuperscript{273} Hornbeck’s Memorandum of Law in Support of Motion for Preliminary Injunction, supra note 48, at 2.\textsuperscript{274} See Eagle, supra note 165, at 473.\textsuperscript{275} Tahoe-Sierra, 535 U.S. at 337 n.31.
as being too broad when brought as a facial challenge, leaving open the possibility that petitioners may have succeeded on an as-applied claim. Here, the shortened lifespan of the property interests at issue provides this kind of persuasive reasoning. Additionally, as applied to this case, there is another clear reason why a moratorium is different and indeed more burdensome than a normal permit delay—the long-term potential to lose existing business contracts and clients permanently to other regions of the world. This type of threat is unique to the context of property interests that may move, terminate, or disappear in response to government action, such as the mobile deep-water oil rigs that are the source of business for the Hornbeck plaintiffs.

C. The Character of the Government Regulation

The final Penn Central factor is the “character of the governmental action.” It is not entirely clear what the Court had in mind when it described this factor, although it might have been an “attempt to separate out physical invasions from all other types of regulation or [an] attempt to distinguish one subset of permissible regulations from others.” Some courts in the past have given this factor significant weight and have appeared unwilling to find a taking where the moratorium is clearly a regulatory action that “arises from some public program adjusting the benefits and burdens of economic life to promote the common good.” More recently, however, the Court has been reluctant to place much weight on this third factor and has explicitly rejected any analysis looking into whether the regulation serves a public purpose. Accordingly, the third Penn Central factor may now mean little more than whether the government action is characterized as physical as

276 Id. at 334; see also Eagle, supra note 165, at 464, 470, 501 (noting that the Court's hand in ruling was largely forced by procedural issues of the case such as the way the challenge was phrased and presented).
277 Tahoe-Sierra, 535 U.S. at 334.
279 Eagle, supra note 165, at 449. “The need to more readily characterize the physical invasion as a taking lasted only for four years [until the holding in Loretto . . . .]” Id.
280 Penn Central, 438 U.S. at 124. Some courts have also looked for evidence of bad faith on the part of the government in determining whether a taking has occurred. See Eagle, supra note 165, at 476. Such an inquiry is no longer appropriate, though, under Lingle v. Chevron U.S.A., Inc., 544 U.S. 528, 532 (2005).
281 See Lingle, 544 U.S. at 532 (holding that the “substantially advances” for a legitimate public use test is inappropriate for takings analysis).
opposed to regulatory, and will need to be more clearly defined by future court decisions.

Ultimately, any court deciding the outcome under factual circumstances similar to the Hornbeck case would need to weigh all these circumstances\textsuperscript{282} to determine whether the magnitude of the government’s interference with the plaintiffs’ property rights was so severe that it required compensation. Despite many persuasive arguments and facts supporting the finding of a taking in Hornbeck, it is unclear whether the plaintiffs could have succeeded in their takings claim. Courts have exhibited a tendency to put strong emphasis on the length of temporary moratoria and have displayed a general reluctance to find temporary takings.\textsuperscript{283} Nonetheless, Hornbeck provides an interesting example illustrating the arguments that are available to future plaintiffs subjected to similar regulations, at a time in the future when takings jurisprudence has begun to accept and apply the interests in fairness and justice articulated in Tahoe-Sierra.\textsuperscript{284}

CONCLUSION

The limited focus on a potential takings claim by the parties in Hornbeck illustrates that this area of law is still very much uncertain and in flux. Yet, future plaintiffs need not view this uncertainty as a problem. Instead, it should be seen as providing increased flexibility, allowing individual plaintiffs to shape persuasive arguments based on their specific circumstances. A more factually inclusive and flexible approach to the current takings framework would provide better protection and induce the government to more thoroughly internalize the consequences of regulations that may go too far.

Critics have complained that “[t]he persistence of incoherence, instability and incomplete explanations in this area of the law suggests that the [Supreme] Court itself is dissatisfied with the tests it has developed, yet is unable to produce a more satisfying jurisprudence.”\textsuperscript{285} Accordingly, there have been numerous proposals for both the expansion and
contraction of the scope of protection provided by the takings clause. On one side of the argument is Richard Epstein, who has recommended a dramatic expansion of takings protection. According to Epstein, the “government must compensate for every diminution in value it causes to owners by restricting the use of property beyond inherent common law limitations.” On the “other end of the spectrum, Peter Byrne has argued that . . . [b]eyond physical intrusions, the regulatory takings doctrine should be abolished.” Because Tahoe-Sierra expressly left open the prospect of “as applied” challenges to moratoria, there is room for the case law to move towards either end of the spectrum. A takings analysis based on the specific circumstances of each case, determined through a more contextually inclusive application of Penn Central, would establish an appropriate middle ground. Indeed, such an application seems to embody the language used in Tahoe-Sierra, where the Court explained, “we are persuaded that the better approach to claims that a regulation has effected a temporary taking ‘requires careful examination and weighing of all the relevant circumstances.’”

Yet the courts have been overly reluctant to find takings due to a rigid application of the Penn Central test. There are inherent differences between the temporary regulation discussed in Tahoe-Sierra and regulations like the oil drilling moratorium that warrant an end to this reluctance. Thus, instead of applying a single rule for all moratoria, courts should look at how that moratorium is affecting the property right at issue. Courts must account for the fact that temporary regulations place a greater burden on property interests in the oil service industry—because of their limited duration—as compared to property interests in other settings. This is especially true in a case like Hornbeck where the property interests are movable and may not return if prohibited for too long.

Including these contextual circumstances in a Penn Central analysis would more closely adhere to the theories behind takings protection discussed in Tahoe-Sierra. Such an application would prevent the government from unfairly

---

286 Fee, supra note 115, at 1015.
287 Id.
288 Id. at 1016.
289 See Tahoe-Sierra, 535 U.S. at 334.
290 See id. at 335 (quoting Palazzolo v. Rhode Island, 533 U.S. 606, 636 (2001) (O’Connor, J., concurring)).
burdening plaintiffs in service industries; it would also, at a minimum, result in more thoughtfully crafted regulations by forcing the government to internalize the possibility of compensation. Further, it would not prevent the government from instituting moratoria when absolutely necessary, but would simply require payment to those unfairly or disproportionately burdened by those regulations. Finally, adopting a more context-specific application seems appropriate as a concession to the ill-defined state of modern takings law. Doing so would allow courts to worry less about futile attempts to decipher and rigidly apply this jurisprudence and more about finding a result that is fair and just.

Edward W. Thrasher†

† B.A., Political Science, Colgate University, 2006; J.D. Candidate, Brooklyn Law School, 2012. I wish to thank Professor Christopher Serkin for inspiring my interest in the subject matter of much of this note. Additionally, I would like to thank the Brooklyn Law Review staff for their assistance in the editing process, especially Shawna MacLeod, T. Daris Isbell, Jared Gilman, and Stephanie Gal. Finally, I would like to thank all my friends and family, especially Mom, Dad, and Emma, for their continuing love and support in law school and life.