ENGAGE

The Journal of the Federalist Society Practice Groups

Project of the E.L. Wiegand Practice Groups

Nota Bene

Affirmative Action for Men? Strange Silences and Strange Bedfellows in the Public Debate over Discrimination Against Women in College Admissions
by Gail Heriot & Alison Somin

EPA Regulation of Fuel Economy: Congressional Intent or Climate Coup?
by Marlo Lewis

Hot News: The "Hot-News" Doctrine Is Hot Again! Or Is It?
by David L. Applegate & Ryan Schermerhorn

The States and the NLRB: A Study in Comparative Sovereignty
by Thomas M. Christina

Florida Constitutional Challenge to Obamacare: It All Comes Down to Broccoli
by Karen R. Harned

Four Articles on Proposed Civil Procedure Reform
by Ronald J. Allen, E. Donald Elliott, Richard M. Esenberg & Martin H. Redish

Book Reviews

Rebuilding the Ark: New Perspectives on Endangered Species Act Reform
Jonathan H. Adler, Editor

Volume 12, Issue 3 November 2011

The Federalist Society
For Law and Public Policy Studies

Directors/Officers

Steven G. Calabresi, *Chairman*
Hon. David M. McIntosh, *Vice Chairman*
Gary Lawson, *Secretary*
Brent O. Hatch, *Treasurer*
C. Boyden Gray
Leonard A. Leo, *Executive Vice President*
Edwin Meese, III
Eugene B. Meyer, *President*
Michael B. Mukasey
Lee Liberman Otis, *Senior Vice President*
. . .
T. Kenneth Cribb, *Counselor*

Board of Visitors

Hon. Robert H. Bork, *Co-Chairman*
Hon. Orrin G. Hatch, *Co-Chairman*
Prof. Lillian BeVier
Hon. Elaine L. Chao
Mr. Christopher DeMuth
Hon. Lois Haight Herrington
Hon. Donald Paul Hodel
Hon. Frank Keating, II
Mr. Robert A. Levy
Hon. Gale Norton
Hon. Theodore B. Olson
Mr. Andrew J. Redleaf
Hon. Wm. Bradford Reynolds
Prof. Nicholas Quinn Rosenkranz
Mr. Gerald Walpin

Staff

President
Eugene B. Meyer

Executive Vice President
Leonard A. Leo

Senior Vice President & Faculty Division Director
Lee Liberman Otis

Anthony Deardurff, Deputy Faculty Director
Tyler Lowe, Assistant Faculty Director

Lawyers Division
Dean Reuter, Vice President, Practice Groups Director
Lisa Budzynski Ezell, Vice President, Lawyers Chapters Director
Jonathan Bunch, Vice President, State Courts Project Director
Juli Nix, Deputy Director
Maureen Wagner, Deputy Director
David C.F. Ray, Associate Director
Allison Aldrich, Associate Director
Hannah DeGuzman, Assistant Director
Thomas Kraemer, Assistant Director
Jennifer Derleth, Assistant Director

Student Division
Peter Redpath, Vice President, Director
Brandon Smith, Deputy Director
Kate Beer Alcantara, Associate Director
Alexandra Bruce, Assistant Director

Pro Bono Center Director
Peggy Little

International Law and Sovereignty Project
James P. Kelly, III, Director
Ken Wiltberger, Deputy Director

Vice President & Finance Director
Douglas C. Ubben

Development
Emily Kuebler, Director
Cynthia Searcy, Director of Major Gifts
Sophia Mason, Assistant Director

Vice President & Information Technology Director
C. David Smith

Membership Director
Matthew Daniel

Director of Social Media and Alumni Relations
Justin Shubow

Office Management
Rhonda Moaland, Director
Matt Nix, Assistant Director

Letter from the Editor...

Volume 12, Issue 3

Engage: The Journal of the Federalist Society Practice Groups provides original scholarship on current, important legal and policy issues. It is a collaborative effort, involving the hard work and voluntary dedication of each of the organization's fifteen Practice Groups. Through its publication, these Groups aim to contribute to the marketplace of ideas in a way that is collegial, measured, and insightful—to spark a higher level of debate and discussion than we often see in today's legal community.

Likewise, we hope that members find the work in the pages to be well-crafted and informative. Articles are typically chosen by our Practice Group chairmen, but we strongly encourage members and general readers to send us their commentary and suggestions at info@fed-soc.org.

Volume 12, Issue 3 — Engage — November 2011

Administrative Law & Regulation

Goodbye Tax Exceptionalism
by Kristin E. Hickman 4

Thinking About the "Practically Unthinkable": Energy Infrastructure and the Threat of Low-Probability, High-Impact Events
by Adam J. White 8

Civil Rights

Affirmative Action for Men? Strange Silences and Strange Bedfellows in the Public Debate over Discrimination Against Women in College Admissions
by Gail Heriot & Alison Somin 14

Criminal Law & Procedure

A Comprehensive Strategy Targeting Recidivist Criminals with Continuous Real-Time GPS Monitoring: Is Reverse-Engineering Crime Control Possible?
by Peter M. Thomson 23

Honest Services Fraud After *Skilling v. United States*
by Steven Wisotsky 31

Environmental Law & Property Rights

EPA Regulation of Fuel Economy: Congressional Intent or Climate Coup?
by Marlo Lewis 36

Federalism & Separation of Powers

A Return to "the Heady Days"? The Supreme Court Addresses Whether the *Bivens* Doctrine Should Extend to Employees of Government Contractors in *Minneci v. Pollard*
by Robert T. Numbers, II & Lisa L. Dixon 47

Financial Services & E-Commerce

An Overview and Analysis of the Consumer Financial Protection Bureau
by Sarah Riddell 56

Free Speech & Election Law

Illuminating *Citizens United*: What the Decision Really Did *by William R. Maurer* 62

Intellectual Property

Hot News: The "Hot-News" Doctrine Is Hot Again! Or Is It?
by David L. Applegate & Ryan Schermerhorn 65

International & National Security Law

Multinational Businesses and the Matrix of Human Rights Governance Networks
by James P. Kelly, III 71

Labor & Employment Law

The States and the NLRB: A Study in Comparative Sovereignty
by Thomas M. Christina 78

Two Guiding Trends in Contemporary Labor and Employment Law: Technology and Fairness
by Daniel Morton-Bentley 86

Litigation

Business Cases and the Roberts Supreme Court *by Martin J. Newhouse* 90

Florida Constitutional Challenge to Obamacare: It All Comes Down to Broccoli
by Karen R. Harned 97

How to Think About Errors, Costs, and Their Allocation
by Ronald J. Allen 105

Twombly in Context: Or Why Federal Rule of Civil Procedure 4(b) Is Unconstitutional
by E. Donald Elliott 110

A Modest Proposal for Human Limitations on Cyberdiscovery
by Richard M. Esenberg 138

Pleading, Discovery, and the Federal Rules: Exploring the Foundations of Modern Procedure
by Martin H. Redish 145

Religious Liberties

Perry v. Schwarzenegger: Is Traditional Marriage Unconstitutional?
by George W. Dent. Jr. 161

Perry, Same-Sex Marriage, and Federal Constitutional Guarantees
by Mark Strasser 166

Does Neutrality Equal Secularism? The European Court of Human Rights Decides *Lautsi v. Italy*
by William L. Saunders 170

Book Review

Rebuilding the Ark: New Perspectives on Endangered Species Act Reform, Jonathan H. Adler, Editor
Reviewed by James S. Burling 175

Editor

Paul F. Zimmerman

Administrative Law & Regulation

Goodbye Tax Exceptionalism

*By Kristin E. Hickman**

In the past few decades, the practices and doctrines governing the interpretation and administration of the federal tax code have diverged somewhat from general administrative law doctrines and norms in several ways. No one doubts that the Administrative Procedure Act ("APA") applies to federal tax administration. No one questions that Treasury regulations interpreting the Internal Revenue Code ("IRC") are legally binding on all taxpayers. Nevertheless, while the standard of judicial review for most agency regulations that carry such legal force derives from *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*,[1] at least until very recently, many tax lawyers, the United States Tax Court, and some circuit courts maintained that an arguably less deferential standard articulated prior to *Chevron* in the tax-specific case of *National Muffler Dealers Ass'n, Inc. v. United States* applied to most tax regulations.[2] The APA generally requires that agencies seeking to promulgate legally binding regulations do so by publishing a notice of proposed rulemaking and considering public comments received in response before issuing final regulations.[3] If an agency feels the need to issue such regulations prior to or without pursuing notice and comment, the APA requires the agency to explain why pursuing the default notice-and-comment process is "impracticable, unnecessary, or contrary to the public interest."[4] The Treasury Department ("Treasury"), however, issues a large percentage of tax regulations—more than one third during a recent three-year period—as temporary regulations with only post-promulgation notice and comment and without a contemporaneous finding of good cause for bypassing that process.[5] Finally, since the Supreme Court's decision in *Abbott Laboratories v. Gardner* in 1967, the courts have interpreted the APA as establishing a presumption in favor of pre-enforcement judicial review of agency rulemaking efforts.[6] By contrast, the courts have long interpreted language in the IRC and the Declaratory Judgment Act as precluding judicial review of pre-enforcement challenges against Treasury regulations and IRS rulings.[7]

Deviations from general administrative law norms are not unique to tax. Lawyers in many practice areas tend to over-rely on precedents specific to the agencies with which they frequently interact, leading to deviations from general administrative law principles in other areas of law as well.[8] Congress can and often does adopt specific statutory provisions that alter the requirements of the APA for particular agencies. Nevertheless, tax scholars for years have decried the particularly insular nature of the tax bar and its resulting habit of ignoring potentially relevant nontax legal doctrine.[9] "Tax is different" has been a frequent and often unchallenged meme.

The tide is turning, however. Two recent judicial opinions, one from the Supreme Court and the other from D.C. Circuit sitting en banc, have begun to reverse the trend of treating tax differently from other areas of administrative law.

Mayo Foundation for Medical Education and Research v. United States

The first, and most important, is the Supreme Court's decision this past Term in *Mayo Foundation for Medical Education & Research v. United States*.[10] The case concerned a Treasury Department interpretation of a provision of the IRC exempting students who work for the academic institutions in which they are enrolled from FICA taxes on their wages. Treasury exercised its general authority under IRC § 7805(a) to "prescribe all needful rules and regulations for the enforcement of" the IRC and adopted a regulation declaring that medical residents are not students, reversing a longstanding IRS interpretation to the contrary.[11] Institutions that withheld and paid the taxes unsuccessfully sought refunds and then promptly sued, challenging the validity of the regulation.

For years prior to *Mayo*, the courts and the tax community had debated whether *Chevron* or the tax-specific *National Muffler* provided the appropriate standard of review for evaluating such general authority Treasury regulations, and for that matter whether the two standards were meaningfully different.[12] The Supreme Court's previous discussions of the issue were muddled and contradictory.[13] Finally, the *Mayo* case brought the issue squarely before the Supreme Court. The *National Muffler* standard expressly called for considering an interpretation's consistency and longevity[14]—factors that weighed against Treasury's new regulation—while *Chevron* expressly recognizes the need to allow agencies to change their interpretive positions. Also, unlike in prior tax cases before the Court, briefing in *Mayo* by the parties and by dueling amici clearly raised and thoroughly addressed the question of *Chevron* versus *National Muffler* review.[15]

Upholding the regulation, an undivided Court unequivocally chose *Chevron* and rejected *National Muffler* as the standard of review for general authority Treasury regulations. In reaching that decision, the Court made several observations and conclusions, including that the *Chevron* and *National Muffler* standards "call for different analyses of an ambiguous statute"; that *National Muffler* factors such as an agency's inconsistency or an interpretation's longevity or contemporaneity (or lack thereof) are not reasons for denying *Chevron* deference to a Treasury regulation; and, finally, that "*Chevron* and *Mead*, rather than *National Muffler* . . . , provide the appropriate framework for evaluating" the Treasury regulation at issue.[16]

In the midst of this analysis, the Court also offered a short discussion of the relationship between tax administration and administrative law doctrine with potential implications beyond the standard of review question. First, the Court stated explicitly, "[W]e are not inclined to carve out an approach to administrative review good for tax law only. To the contrary, we have expressly '[r]ecogniz[ed] the importance of maintaining a

* *Kristin E. Hickman is a Professor of Law at the University of Minnesota Law School. Part of this essay draws heavily from analysis posted by Professor Hickman on TaxProf Blog, http://taxprof.typepad.com.*

uniform approach to judicial review of administrative action.'"[17] In making this statement, the Court quoted *Dickinson v. Zurko*, a non-tax (patent) case with an extensive discussion regarding Congress's intent that the APA bring uniformity to the otherwise disparate field of federal administrative action.[18] The Court also cited *Skinner v. Mid-America Pipeline Co.*[19] for "declining to apply 'a different and stricter nondelegation doctrine in cases where Congress delegates discretionary authority to the Executive under its taxing power.'"[20] Other turns of phrase within the *Mayo* Court's analysis reflect a similar orientation toward reconciling the tax and non-tax contexts.

Cohen v. United States

The D.C. Circuit's recent en banc decision in *Cohen v. United States* is less immediately consequential but, consistent with the Supreme Court's policy of administrative law uniformity, represents a further shift in favor of bringing tax administration back in line with administrative law norms.[21] The case grew from several challenges against an old telephone excise tax made defunct by changes in telephone technology and long-distance billing practices. After several circuit courts rejected the IRS's arguments in favor of the continued vitality of the tax,[22] the IRS promulgated special refund procedures for the tax by issuing informal guidance, Notice 2006-50, without notice and comment.[23] Taxpayers who consider the IRS's special refund procedures for the telephone excise tax to be fundamentally flawed challenged Notice 2006-50 on APA procedural grounds, seeking notice and comment as the appropriate forum for requiring the IRS to address the alleged inadequacies.

IRC § 7421(a), also known as the Anti-Injunction Act, generally prohibits any lawsuit "for the purpose of restraining the assessment or collection of any tax" until either the IRS issues a notice of deficiency to a taxpayer or denies a taxpayer-requested refund. Correspondingly, the Declaratory Judgment Act prevents courts from providing declaratory relief for controversies "with respect to Federal taxes."[24] In a series of cases in the 1960s and 1970s, the Supreme Court interpreted these provisions as precluding judicial review of virtually all tax cases except for statutory deficiency or refund actions.[25] Although the Court has never interpreted the Anti-Injunction Act or the Declaratory Judgment Act as precluding pre-enforcement judicial review of APA procedural challenges against Treasury regulations and IRS rulings, a few lower courts interpreted its precedents as requiring that conclusion.[26]

In *Cohen*, after the district court dismissed the case for lack of jurisdiction, a split panel of the D.C. Circuit reversed and remanded the case for consideration of the merits of the taxpayers' APA procedural claim.[27] Several months later, the court granted the government's petition for en banc review and requested briefing on several questions pertinent to interpreting the Anti-Injunction Act and the Declaratory Judgment Act in relation to the APA.[28] This summer, a divided en banc court issued its decision, also in favor of the taxpayers, reaching several conclusions regarding the courts' jurisdiction to consider APA procedural claims in the tax context.[29]

First, the court held that APA § 702 waives sovereign immunity for APA procedural challenges in the tax context, just as it does in other regulatory areas; there is no tax exception from the APA.[30] Picking up the *Mayo* Court's admonition in favor of administrative law uniformity, quoted elsewhere in the majority opinion, the court concluded that "[t]he IRS is not special in this regard; no exception exists shielding it—unlike the rest of the Federal Government—from suit under the APA."[31]

Next, the court held that the Anti-Injunction Act and the Declaratory Judgment Act do not bar judicial review of the taxpayers' APA procedural claim.[32] Citing and quoting extensively from *Hibbs v. Winn*, in which the Supreme Court interpreted a similar provision governing state taxation,[33] the D.C. Circuit adopted a narrow, textualist interpretation of the Anti-Injunction Act's limitation on judicial review. According to the court, the Anti-Injunction Act's prohibition against suits to restrain "the assessment or collection of any tax" does not refer to a "'single mechanism' that ultimately determines the amount of revenue the Treasury retains" and is not "synonymous with the entire plan of taxation."[34] Instead, "assessment" and "collection" are defined terms in the Internal Revenue Code. "Assessment" represents "the trigger for levy and collection efforts," and "collection" is "the actual imposition of tax against a plaintiff."[35] The appellants' APA procedural claim does not concern the assessment or collection of taxes because "[t]he IRS previously assessed and collected the excise tax at issue"; rather, this suit is merely about the procedures under which the IRS will refund taxes that it has already collected.[36] Although the text of the Declaratory Judgment Act is arguably broader in its prohibition of declaratory relief in tax cases, the *Cohen* court held that the Declaratory Judgment Act is to be interpreted coterminously with the Anti-Injunction Act and not as a separate limitation on judicial review.[37]

While the government argued that interpreting the Anti-Injunction Act and the Declaratory Judgment Act in this way would open the floodgates for APA challenges against Treasury and IRS actions, those provisions are not the only potential limitations on judicial review of agency action, whether in the tax context or otherwise. The majority and dissenting opinions considered several. Particularly where (as here) a specific statute provides its own legal mechanisms for seeking judicial review, APA §§ 703 and 704 limit the availability of judicial review under the APA to cases in which the challenging parties otherwise lack an adequate legal remedy.[38] The dissenting judges in *Cohen* contended that statutory refund actions authorized by IRC § 7422 offered the appellants an adequate legal remedy.[39] The majority disagreed on the ground the taxpayers' APA procedural challenge seeks equitable relief rather than a tax refund (even if a refund is their ultimate goal), and IRC § 7422 does not offer that remedy.[40] Both opinions additionally discuss the doctrines of ripeness and exhaustion at some length, while standing and finality limitations make brief appearances as well. In analyzing these different barriers to judicial review, the *Cohen* majority construed its conclusions very narrowly. Indeed, the court labeled the case before it as "*sui generis*" and either assumed or stated outright that judicial review of many if not most APA procedural challenges to Treasury and IRS actions will be limited by one or more of these obstacles.[41] Hence, while the taxpayers' APA claim may not be the only one eligible for judicial review outside of the statutory mechanisms provided by

the IRC, just how many others will be able to run this gauntlet of limitations is unclear. Regardless, the *Cohen* court's insistence upon treating the taxpayer's APA challenge as such, and its interpretation of the Anti-Injunction Act and the Declaratory Judgment Act as interacting with rather than wholly displacing the APA, represent bold statements about tax as part of and not separate from administrative law more generally.

One final point of interest from *Cohen* concerns the court's statement regarding the finality of Notice 2006-50. The initial panel decision in the case determined that Notice 2006-50 represents final agency action because it determines taxpayer rights and obligations and binds the agency.[42] In discussing other issues concerning the justiciability of the taxpayers' APA claim, the en banc court reiterated that conclusion.[43] The IRS does not employ APA notice and comment rulemaking in issuing notices (or other informal guidance documents, like revenue rulings or revenue procedures), taking the position that these pronouncements are exempt from such requirements as either interpretative rules or policy statements. Indeed, the *Cohen* taxpayers' primary claim at this point is that the IRS should have subjected the rules contained in Notice 2006-50 to notice-and-comment rulemaking and failed to do so.

General administrative law doctrine surrounding the interpretative rule and policy statement exemptions from notice and comment procedures is notoriously murky but overlaps substantially with finality doctrine.[44] A conclusion that Notice 2006-50 represents a justiciable final agency action does not automatically compel a decision that the IRS should have used notice and comment in that pronouncement's development, but a contrary holding may be difficult to justify. If, in future proceedings, the district court and the D.C. Circuit ultimately conclude that Notice 2006-50 is procedurally invalid for the lack of notice-and-comment rulemaking, then the same is likely true of other IRS notices, revenue rulings, and revenue procedures, meaning that many such guidance documents may be susceptible to invalidation on APA procedural grounds. Thus, while the D.C. Circuit's recent decision in the *Cohen* case really concerns the timing of and avenues for seeking judicial review in tax cases, rather than its availability under any circumstances, the panel's earlier conclusion that Notice 2006-50 represents final agency action may ultimately be the most significant aspect of this case vis-à-vis future litigation.

Conclusion

Notwithstanding these two very important decisions, so many questions regarding the applicability of administrative law doctrines and norms in the tax context remain unanswered. The *Cohen* taxpayers' claims regarding the substantive and procedural validity of Notice 2006-50, and the implications for other IRS guidance documents, remain unresolved.[45] Additionally, the Supreme Court recently agreed to consider a series of conflicting federal circuit court decisions regarding the validity of yet another Treasury regulation, this one raising issues concerning the applicability of the Court's decision in *National Cable and Telecommunications Ass'n v. Brand X Internet Services*[46] to a Treasury regulation promulgated initially in temporary form with only post-promulgation notice and comment in the course of litigation.[47] The courts may ultimately decide that, in some instances, the IRC authorizes deviations from the APA in the tax context. Nevertheless, *Mayo* and *Cohen* have buried the notion that tax is especially unique among areas of government regulation. The tax community has taken notice, and government officials responsible for administering the tax laws are on notice that they should attend also to administrative law doctrines and norms.

Endnotes

1 467 U.S. 837, 842-43 (1984); *see also* United States v. Mead Corp., 533 U.S. 218, 226-27 (2001) (holding that agency actions carrying "the force of law" are *Chevron*-eligible).

2 440 U.S. 472, 477 (1979).

3 5 U.S.C. § 553(b)-(c).

4 5 U.S.C. § 553(b)(B).

5 *See* Kristin E. Hickman, *Coloring Outside the Lines: Examining Treasury's (Lack of) Compliance with Administrative Procedure Act Rulemaking Requirements*, 82 Notre Dame L. Rev. 1727, 1748-51 (2007) (documenting empirical study results).

6 387 U.S. 136 (1967).

7 *See* Kristin E. Hickman, *A Problem of Remedy: Responding to Treasury's (Lack of) Compliance with Administrative Procedure Act Rulemaking Requirements*, 76 Geo. Wash. L. Rev. 1153, 1164-74 (2008) (summarizing the jurisprudence).

8 *See generally* Richard Levy & Robert Glicksman, *Agency-Specific Precedents*, 89 Texas L. Rev. 499 (2011) (describing several examples).

9 *See, e.g.*, Paul L. Caron, *Tax Myopia, or Mamas Don't Let Your Babies Grow Up to Be Tax Lawyers*, 13 Va. Tax Rev. 517, 518 (1994) ("[T]ax law too often is mistakenly viewed by lawyers, judges, and law professors as a self-contained body of law . . . . [T]his misperception has impaired the development of tax law by shielding it from other areas of law that should inform the tax debate."); Leandra Lederman, *"Civil"izing Tax Procedure: Applying General Federal Learning to Statutory Notices of Deficiency*, 30 U.C. Davis L. Rev. 183, 183 (1996) ("Tax law tends to be uninformed by other areas of law. This insularity has the unfortunate consequence of depriving tax and other fields of cross-fertilization.").

10 131 S. Ct. 704 (2011).

11 *See* Treas. Reg. § 31.3121(b)(10)-2(e); T.D. 9196, 69 Fed. Reg. 76404 (Dec. 21, 2004).

12 *See, e.g.*, Kristin E. Hickman, *The Need for* Mead*: Rejecting Tax Exceptionalism in Judicial Deference*, 90 Minn. L. Rev. 1537, 1556-59 (2006) (summarizing the disagreement among the federal circuit courts and legal scholars).

13 *See* Kristin E. Hickman, *Agency-Specific Precedents: Rational Ignorance or Deliberate Strategy*, 89 Texas L. Rev. See Also 89, 107-08 (2011) (analyzing Supreme Court tax precedents).

14 *See* Nat'l Muffler Dealers Ass'n, Inc. v. United States, 440 U.S. 472, 477 (1979).

15 *See* Brief for Petitioners, Mayo Found. for Med. Educ. & Research v. United States, 131 S. Ct. 704 (2011) (No. 09-837); Brief of Amicus Curiae Professor Kristin E. Hickman in Support of Respondent, Mayo Found. for Med. Educ. & Research v. United States, 131 S. Ct. 704 (2011) (No. 09-837); Brief of Tax Professor Carlton M. Smith as Amicus Curiae in Support of the Petitioners, Mayo Found. for Med. Educ. & Research v. United States, 131 S. Ct. 704 (2011) (No. 09-837).

16 *See Mayo*, 131 S. Ct. at 712-14.

17 *Id.* at 713.

18 527 U.S. 150, 154 (1999).

19 490 U.S. 212, 222-23 (1989).

20 *Mayo*, 131 S. Ct. at 713.

21 2011 WL 2600672 (D.C. Cir. Jul. 1, 2011) (en banc).

22 *See, e.g.*, Am. Bankers Ins. Group v. United States, 408 F.3d 1328 (11th Cir. 2005); Office Max, Inc. v. United States, 428 F.3d 583 (6th Cir. 2005); Nat'l R.R. Passenger Corp. v. United States, 431 F.3d 374 (D.C. Cir. 2005).

23 *See* IRS Notice 2006-50, 2006-1 C.B. 1141.

24 28 U.S.C. § 2201(a).

25 *See, e.g.*, United States v. Am. Friends Serv. Comm., 419 U.S. 7, 8, 12 (1974) (per curiam); Alexander v. "Ams. United" Inc., 416 U.S. 752, 756-57, 763 (1974); Bob Jones Univ. v. Simon, 416 U.S. 725, 727, 735-36 (1974); Enochs v. Williams Packing & Navigation Co., 370 U.S. 1, 5-6 (1962).

26 *See, e.g.*, Stephenson v. Brady, 927 F.2d 596, 1991 WL 22835, at *2, *4-5 (4th Cir. 1991) (per curiam) (unpublished table decision); Reimer v. United States, 919 F.2d 145, 1990 WL 186825, at *1-3 (9th Cir. 1990) (unpublished table decision); *cf.* Inv. Annuity, Inc. v. Blumenthal, 609 F.2d 1, 8-9 (D.C. Cir. 1979) (reading the Anti-Injunction Act and the Declaratory Judgment Act as statutory limitations on pre-enforcement judicial review in the tax context, notwithstanding the APA presumption in favor thereof). *But see* E. Ky. Welfare Rights Org. v. Simon, 506 F.2d 1278, 1285-86 (D.C. Cir. 1974), *vacated on other grounds*, 426 U.S. 26 (1976) (concluding before vacatur that the Anti-Injunction Act, the Declaratory Judgment Act, and the APA do not operate to preclude judicial review of all pre-enforcement tax claims).

27 *See* Cohen v. United States, 578 F.3d 1 (D.C. Cir. 2009), *vacated in part*, 599 F.3d 652 (D.C. Cir. 2010).

28 *See* Cohen v. United States, 599 F.3d 652 (D.C. Cir. 2010).

29 *See* Cohen v. United States, 2011 WL 2600672 (D.C. Cir. Jul. 1, 2011) (en banc).

30 *See id.* at *4.

31 *Id.*

32 *See id.* at *6-7.

33 *See* Hibbs v. Winn, 542 U.S. 88, 99-108 (2004) (interpreting 28 U.S.C. § 1341).

34 *Cohen*, 2011 WL 2600672, at *7.

35 *Id.*

36 *Id.*

37 *See id.* at *8-12.

38 *See* 5 U.S.C. §§ 703 & 704.

39 *See Cohen*, 2011 WL 2600672, at *19 (Kavanaugh, J. dissenting).

40 *See id.* at *12-14.

41 *Id.* at *14.

42 *See Cohen*, 578 F.3d at 8-9.

43 *See Cohen*, 2011 WL 2600672, at * 4.

44 *See, e.g.*, Richard J. Pierce, Jr., *Distinguishing Legislative Rules from Interpretative Rules*, 52 Admin. L. Rev. 547, 547-48 (2000) (citing cases describing the distinction between legislative and interpretative rules as "'fuzzy,' 'tenuous,' 'blurred,' 'baffling,' and 'shrouded in considerable smog'").

45 *See generally* Kristin E. Hickman, *IRB Guidance: The No Man's Land of Tax Code Interpretation*, 2009 Mich. St. L. Rev. 239 (2009) (documenting administrative law questions raised by IRS use of informal guidance documents).

46 545 U.S. 967 (2005).

47 *See* United States v. Home Concrete & Supply, No. 11-139, 2011 WL 3322358 (Sept. 27, 2011) (granting certiorari to evaluate validity of regulation interpreting IRC § 6501(e)); *see also, e.g.*, Grapevine Imports, Ltd. v. United States, 636 F.3d 1368 (Fed. Cir. 2011) (finding IRC § 6501(e) statute ambiguous at *Chevron* step one, deferring to the regulation at *Chevron* step two, and declaring APA procedural challenge against temporary regulation moot because the regulation was finalized using notice and comment); Burks v. United States, 633 F.3d 347 (5th Cir. 2011) (finding IRC § 6501(e) clear in support of the taxpayer at *Chevron* step one but alternatively declaring that the court would have found the regulation unreasonable at *Chevron* step two because the temporary regulation lacked notice and comment); Beard v. Comm'r, 633 F.3d 616 (7th Cir. 2011) (finding IRC § 6501(e) clear in support of the government at *Chevron* step one but declaring that the court would have deferred to the regulation regardless).

Thinking About the "Practically Unthinkable": Energy Infrastructure and the Threat of Low-Probability, High-Impact Events

*By Adam J. White**

The National Environmental Policy Act[1] requires federal agencies to ascertain and evaluate the possible environmental effects of federally regulated energy infrastructure proposals. But this broad statutory requirement leaves great uncertainty as to which hypothetical risks of environmental harm must be evaluated, and which risks may be set aside as too contingent or otherwise improbable to merit review. Recent events—the Japanese tsunami that disrupted a nuclear power plant, for example, or the deepwater Gulf of Mexico oil spill—remind us that seemingly unthinkable disasters can occur, posing a significant threat of harm to the environment. But two recent court of appeals decisions have created a circuit split on the question of precisely how agencies should approach the possibility of low-probability, high-impact events—or, as they have come to be known, "Black Swans."

The Tsunami and the Tepco Nuclear Plant

On March 11, 2011, a 9.0 magnitude earthquake twenty miles below the surface of the Pacific Ocean triggered a tsunami.[2] The thirty-foot waves raced to Japan's shore, unleashing a measure of devastation that stunned the worldwide audiences that watched the aftermath in unforgettable television footage.

Of the earthquake's innumerable casualties, the most significant was Tokyo Electric Power Co.'s ("Tepco's") Fukushima Daiichi nuclear plant. When the tsunami struck, three of Fukushima's six boiling-water nuclear reactors were actively operating, their plutonium rods fueling the reactions that supplied electric power to Tepco's customers. Fukushima was utterly unprepared to handle the events that ensued: the earthquake and tsunami cut off the plant's outside power supply, as well as its backup diesel generators, leaving Fukushima unable to sufficiently cool the reactor cores or spent fuel pools.[3] As the crisis continued, hydrogen buildup led to large explosions at the three reactors, releasing radiation into the atmosphere.[4] Each of the three reactors suffered a meltdown,[5] and the seawater that was allowed to flood the reactors, in a desperate attempt to cool them, spread radiation into the ocean.[6]

Tepco and Japanese regulators had designed Fukushima to handle the tectonic events that regularly strike Japan, but even they did not design a plant capable of handling this earthquake, Japan's largest in *three centuries*.[7] "The disaster plan didn't function . . . It didn't envision something this big."[8] Tepco and the regulators "fail[ed] to envision the kind of worst-case scenario that befell Japan: damage so extensive that the plant couldn't respond on its own or call for help from nearby plants."[9]

Or, as Fukushima's own report on its accident-management protocols acknowledged, "The possibility of a severe accident occurring is so small that from an engineering standpoint, it is *practically unthinkable*."[10]

"Practically Unthinkable"

"Practically unthinkable"—an ironic choice of words. For while virtually every aspect of this tragic chain of events was unprecedented, that particular characterization of an unforeseen disaster is itself *well*-precedented. In just recent memory, the United States has witnessed several major infrastructure catastrophes that were quickly characterized, after the fact, as "unthinkable."

Most recently, of course, was the April 2010 explosion at the Deepwater Horizon oil rig at the Gulf of Mexico's Macondo Prospect. The unexpected failure of a blowout preventer to protect against a methane bubble resulted in an explosion that destroyed the rig and allowed nearly five million barrels of oil to leak into the Gulf.[11] BP urged that the disaster was caused by no single action (or company): "Rather, a complex and interlinked series of mechanical failures, human judgments, engineering design, operational implementation and team interfaces came together to allow the initiation and escalation of the accident."[12] According to BP's CEO, there was "no precedent" for the confluence of events that led to the explosion. Or, as BP's spokesmen stated in Fukushima-esque terms, the blowout preventer's failure "seemed inconceivable."[13]

Now, with the benefit of hindsight, those once-"unthinkable" risks are now very, very thinkable. And so U.S. nuclear facility operators—and nuclear power's critics, and Congress—are working to ensure that U.S. reactors are protected from the particular threats that ultimately doomed Fukushima.[14]

But what about those risks that we cannot see in hindsight—because they have not previously occurred? The purely hypothetical risks for which there is no precedent? How could regulators efficiently and effectively guard against the unprecedented earthquake triggering cascading failures at a coastal nuclear facility, in the years *before* the Fukushima disaster? Or, how could regulators sitting in judgment of a deepwater oil drilling operation meaningfully consider the environmental harm that could ensue if a blowout preventer unprecedentedly failed to protect against a methane bubble, long *before* the Deepwater Horizon incident?

"Black Swans" and the Precautionary Principle

These are no small questions. In fact, the very question of how to anticipate—and guard against—the risk of high-impact, low-probability events has been the subject of a contentious intellectual debate in recent years.

Much of the debate has been sparked by—and captured by the title of—Nassim Taleb's 2007 bestseller, *The Black Swan*.

** Adam J. White (www.adamjwhite.com) is Counsel at Boyden Gray & Associates in Washington, D.C., and a 2011 National Security Fellow for the Foundation for Defense of Democracies. This article represents only Mr. White's views, and is not written on behalf of the firm, its clients, or FDD. The title of this article is adapted from the title of Herman Kahn's classic book on nuclear strategy.*

In that book (which is unrelated to the award-winning 2010 film of the same name), Taleb argues that people inherently underestimate the risk of high-impact, low-probability events in their daily lives. We rely on models, habits, assumptions, and *ex post facto* explanations that rely too heavily on past facts, in a manner that blinds us to events that have not previously occurred, but which could occur—with catastrophic effect.[15] Taleb's choice of title refers to one such example: "[P]eople in the Old World were convinced that *all* swans were white" because only white swans were seen in the known world. But then Australia was discovered, and with it black swans. And for all of the innumerable white swans that had been seen since time immemorial, it took only one black swan to disprove the theory that all swans were white.[16]

Thus, Taleb assigned the name "Black Swan" to events with three attributes: they are (1) outlier events beyond the realm of regular experience, with (2) an extreme impact, but (3) for which human nature leads us to create *ex post facto* explanations that make the event seem explainable and predictable.[17] The concept and title have quickly come to direct the public debate over how best to guard against unforeseen events in financial markets, energy infrastructure, and other critically important aspects of modern life.[18]

Of course, while Taleb is perhaps the most prominent participant in the public debate over risk management, he is not alone in grappling with the question of how to manage the threat of unmanageable risk. Judge Richard Posner, law professor Cass Sunstein, political scientist Francis Fukuyama, and others have weighed in with insightful analyses,[19] but regulators and industry officials seeking practical guidance for the mitigation of unforeseeable or seen-but-highly-unlikely risks will find little help in the academic literature. For example: Judge Posner's variation on his familiar framework of cost-benefit balancing, applied to protect against catastrophic risk, is elegant in theory.[20] But it ultimately offers no real guidance in the absence of reliable evidence of the true risk of harm, the magnitude of harm, or the cost of guarding against the threat.[21]

Given both the lack of reliable guideposts for case-by-case analysis, and the possible risk of harm at stake when considering the threat of low-probability, high-impact events, regulators and advocates often fall back on a heuristic commonly known as the "precautionary principle." As Sunstein summarizes it, the precautionary principle instructs us, "[i]n case of doubt," to "[a]void steps that will create a risk of harm. Until safety is established, be cautious; do not require unambiguous evidence. In a catchphrase: better safe than sorry."[22]

Perhaps the best-publicized—and most controversial—application of the precautionary principle was that offered by Vice President Cheney, with respect to the Bush Administration's strategy in the early days of the Global War on Terror: because an unexpected terrorist attack could have catastrophic consequences, even "a one percent chance" of occurrence would suffice to trigger responding as if the attack were "a certainty."[23]

Or, to take an example from the other side of the political aisle, many critics of climate change and other environmental risks have presented their criticisms in similar terms. Speaking to the question of man-made climate change, *New York Times* columnist Thomas Friedman urged, "When I see a problem that has even a 1 percent probability of occurring and is 'irreversible' and potentially 'catastrophic,' I buy insurance. That is what taking climate change seriously is all about."[24]

But as a heuristic, the precautionary principle suffers no shortage of defects all its own. In its strongest form, the principle is "paralyzing" because it effectively "forbids all courses of action, including inaction."[25] As Sunstein explains, in any case where "there is a possibility of serious harms, and no authoritative scientific evidence suggests that the possibility is close to zero," then if "the burden of proof is on the proponent of the activity or process in question, the precautionary principle would seem to impose a burden of proof that cannot be met."[26] In other words, when a course of action implicates the Precautionary Principle, but so do its alternatives—including inaction—then the Precautionary Principle offers no solution.

From Theory to Practice:
The National Environmental Policy Act

As a law professor, before he joined President Obama's Office of Management and Budget as Administrator of the Office of Information and Regulatory Affairs, Cass Sunstein wrote extensively on this subject, especially with respect to climate change and terrorism.[27] And, in fact, those are two of the areas in which difficult questions of low-probability, high-impact events arise as a practical regulatory matter, as agencies implement the National Environmental Policy Act ("NEPA").[28]

NEPA "is our basic national charter for protection of the environment."[29] But while its prefatory language lays out general substantive environmental goals,[30] NEPA's "mandate . . . is essentially procedural."[31] Unlike other environmental statutes such as the Clean Water Act or Wild and Scenic Rivers Act, or federal infrastructure permitting statutes such as the Natural Gas Act, Federal Power Act, or Atomic Energy Act, it neither imposes new substantive environmental standards nor directs federal or state agencies to impose new substantive environmental standards.[32]

Instead, NEPA works to supplement the agencies' enforcement of those substantive standards in cases of federal action—such as an agency's approval of an application to site a new interstate natural gas pipeline under the Natural Gas Act. Specifically, anytime that an agency proposes to undertake or approve a "major Federal actio[n] significantly affecting the quality of the human environment,"[33] such as the approval of an application to construct a federally-regulated energy infrastructure project under the Natural Gas Act or other statute, the agency must take a "hard look" at the possible environmental effects of the project, and the comparable effects of alternative proposals—including the "no-action" alternative—before approving the project.[34]

By adding that procedural requirement to each agency's administration of its other substantive statutes, "NEPA . . . makes environmental protection a part of the mandate of every federal agency and department."[35] Its procedural requirements ensure that the agency take action "only after responsible decisionmakers had fully adverted to the environmental consequences of the actions, and had decided

that the public benefits flowing from the actions outweighed their environmental costs."[36] Thus, "by focusing the agency's attention on the environmental consequences of a proposed project," including its effect on human populations, before issuing a final approval for the project, "NEPA ensures that important [environmental] effects will not be overlooked or underestimated only to be discovered after resources have been committed or the die otherwise cast."[37]

NEPA accomplishes this primarily by requiring the agency to prepare an "environmental impact statement" ("EIS") for any "major Federal action[] significantly affecting the quality of the human environment."[38] The EIS identifies, *inter alia*, "the environmental impact of the proposed action," "any adverse environmental effects which cannot be avoided should the proposal be implemented," and "alternatives to the proposed action[.]"[39]

Those broad requirements leave no shortage of discretion to the agency; they leave even more discretion to the underlying applicant, with respect to defining the project's purpose and alternatives.[40] But perhaps the most complex and contentious aspect of the EIS is identifying the "adverse environmental effects" that would be caused by the proposed action and thus must be analyzed by the agency—for precisely the reasons discussed above. A proposed energy infrastructure project, for example, has reasonably certain environmental effects, at least with respect to the geographic footprint of the project, but it also can give rise to no shortage of hypothetical environmental impacts, ranging in likelihood from the probable to the nearly impossible. Even assuming that the agency, the project's proponents, and the project's critics could identify all possible contingencies and environmental effects—even the so-called "Black Swans" threats—measuring the actual risk and possible impact of all such threats is effectively impossible.

Under that shadow of uncertainty, NEPA does not impose the Precautionary Principle; it does not require the agencies to consider any and all hypothetical environmental impacts of the project. While "[r]easonable forecasting and speculation" is "implicit in NEPA,"[41] "agencies may not be precluded from proceeding with particular projects merely because the environmental effects of that project remain to some extent speculative."[42] Instead, the agency must strike a pragmatic balance:

> [O]nly those effects that are "likely" (or "foreseeable" or "reasonably foreseeable") need be discussed . . . and, as in other legal contexts, the terms "likely" and "foreseeable," as applied to a type of environmental impact, are properly interpreted as meaning that the impact is sufficiently likely to occur that a person of ordinary prudence would take it into account in reaching a decision.[43]

This resembles a sliding scale of risk and possible harm:

> Danger . . . is not set by a fixed probability of harm, but rather is composed of reciprocal elements of risk and harm, or probability and severity. . . . That is to say, the public health may properly be found endangered both by a lesser risk of a greater harm and by a greater risk of a lesser harm. Danger depends upon the relation between the risk and harm presented by each case, and cannot legitimately be pegged to "probable" harm, regardless of whether that harm be great or small.[44]

Marshaling those considerations, the Supreme Court's ultimate standard is effectively tort law's standard of "proximate causation." In *Metropolitan Edison Co. v. People Against Nuclear Energy*, the leading case on this point, the Court expressly adopted this tort concept as guiding the agencies' inquiry under NEPA. According to the Court, NEPA's concept of environmental impact should "be read to include a requirement of a reasonably close causal relationship between a change in the physical environment and the effect at issue," caused by the federal agency's decision to approve the project at issue.[45] The mere *risk* that the agency-approved project could give rise to environmental impact is not *itself* environmental impact: "[A] *risk* of accident is not an effect . . . . A risk is, by definition, unrealized in the physical world."[46] And so where the agency's action does not directly result in the foreseen environmental effect, but instead is merely the first step in a causal chain with numerous "middle links" preceding the hypothetical outcome, those intervening links may "lengthen[] the causal chain beyond the reach of NEPA."[47]

Applying that standard in *Metropolitan Edison*, the Court held that NEPA did not require the Nuclear Regulatory Commission, in reviewing a nuclear facility license application, to consider purely psychological harm—"anxiety, tension and fear, a sense of helplessness, and accompanying physical disorders"[48]—that could be suffered by persons living near the facility, caused by the increased risk of nuclear accident.[49]

But in rejecting the challenge to NRC's decision in that case, the Court cautioned against over-reading the Court's own decision: "We emphasize that in this case we are considering effects"—*i.e.*, the aforementioned "tension and fear"—"caused by the risk of an accident. The situation where an agency is asked to consider effects that will occur if a risk is realized, for example, if an accident occurs at [the nuclear facility], is an entirely different case."[50]

Two decades later, in cases once again involving the NRC, two federal courts of appeals split on the question of how to apply NEPA under the post-9/11 shadow of nuclear terrorism.

Ninth Circuit: Over-Simplifying the "Chain" Of Events

After the al Qaeda strike of September 11, 2001, the threat of terrorist attack on the nation's nuclear infrastructure was immediately cognizable. Indeed, the "9/11 Commission Report"—or, the Final Report of the National Commission on Terrorist Attacks Upon the United States—reported that both 9/11 mastermind Khalid Sheikh Mohammed and Mohammed Atta, the al Qaeda agent who carried out the attack, had considered targeting U.S. nuclear facilities.[51] And the threat of terrorist attack quickly became a central issue in the NRC's review of applications for nuclear infrastructure.

Just three months after the attacks of 9/11, Pacific Gas & Electric Co. applied to NRC for a license to construct and operate a spent fuel storage installation at the site of its Diablo Canyon nuclear power plant.[52] San Luis Obispo Mothers

for Peace, a non-profit organization opposed to the project, challenged the application and the agency's preliminary approval on the ground that, *inter alia*, they "fail[ed] to address environmental impacts of terrorist or other acts of malice or insanity[.]"[53] But the NRC ultimately approved the project application; it refused to analyze the environment impact of a terrorist attack on the facility, concluding that NEPA review of that issue was unnecessary because, *inter alia*, "the possibility of terrorist attack is too far removed from the natural or expected consequences of agency action"; and in any event "the risk of a terrorist attack cannot be determined," such that "the analysis is likely to be meaningless."[54]

But on a petition for review, the Ninth Circuit vacated the NRC's decision, holding that the NRC violated NEPA by categorically refusing to include in its NEPA review the threat of terrorist attack. Importantly, the Ninth Circuit held the NRC could not ignore the threat of terrorist attack simply because, as NRC concluded, "the possibility of a terrorist attack . . . is speculative and simply too far removed from the natural or expected consequences of [NRC's] action . . . ."[55] That argument, the Ninth Circuit concluded, was premised upon a misreading of the Supreme Court's decision in *Metropolitan Edison*.

According to the Ninth Circuit, "[t]he events at issue here, as well as in *Metropolitan Edison* . . . form a chain of three events: (1) a major federal action; (2) a change in the physical environment; and (3) an effect."[56] The chain may have been the same in both *Metropolitan Edison* and the present case, but the two cases involved different *connections* in the chain: *Metropolitan Edison* "was concerned with the relationship between events 2 and 3 (the change in the physical environment, or increased risk of accident resulting from the renewed operation of a nuclear reactor, and the effect, or the decline in the psychological health of the human population)."[57]

"In the present case," by contrast, "the disputed relationship is between events 1 and 2 (the federal act, or the licensing of the Storage Installation, and the change in the physical environment, or the terrorist attack)."[58] Those connections in the chain, according to the Ninth Circuit, presented precisely the type of case that the Supreme Court stated it had *not* decided in *Metropolitan Edison*: namely, one in which "an agency is asked to consider effects that will occur if a risk is realized, for example, if an accident occurs . . . ."[59]

Thus, the Ninth Circuit concluded, where (as in the present case) the question is what change in the physical environment could result from the federal agency's action, "[t]he appropriate inquiry is . . . whether such attacks are so 'remote and highly speculative' that NEPA" does not require their consideration.[60] The court determined that the NRC's categorical refusal to consider the possibility of terrorist attack was unreasonable.[61] And the court further held that the NRC could not avoid NEPA's requirement by concluding that the risk of terrorist attack was "unquantifiable."[62]

The Ninth Circuit's analysis appears correct in at least one respect: in *Metropolitan Edison*, the Supreme Court specifically stated that it was "considering effects caused by the risk of an accident," and not the "effects that will occur if a risk [of accident] is realized."[63]

But the Ninth Circuit erred in attempting to equate its case and *Metropolitan Edison* as simply involving two separate segments in the same three-point chain. The Supreme Court made clear that *Metropolitan Edison*'s "chain" involved not three linked points, but *four*: (1) the agency's action, (2) the resulting increased risk of nuclear incident, (3) the project opponents' *perception* of that risk, and (4) the physical or physiological effects resulting from their perceiving that risk.[64] The Court held that NEPA required no analysis of that final link because *both* the second *and* the third links, together, stretched the causal chain too far.[65]

And *Metropolitan Edison*'s causal chain was not the only one that the Ninth Circuit misperceived; it also misperceived the causal chain at issue in the case at bar, which also involved not three links but at least *five*: (1) the agency's action, (2) the resulting risk (i.e., the increased opportunity for terrorist attack), (3) terrorists' identification of that opportunity, (4) terrorists' action to seize that opportunity, and (5) the results of the terrorists' action. This causal chain is *at least* as lengthy and uncertain as the chain in *Metropolitan Edison*.

And that uncertainty undermines the remainder of the Ninth Circuit's analysis. Even assuming that *San Luis Obispo Mothers for Peace* involved a case in which the NRC was only being asked "to consider the effects that will occur if the risk is realized," there remain several "links" of uncertainty. The "risk" is "realized" when terrorists identify their new opportunity for attack, but there remain the uncertainties of whether the terrorists will decide to seize that opportunity, whether the terrorists effect that decision by conducting a successful attack, and what the effects of that attack would be.

Third Circuit: Seeing Each Link in the Chain

While *San Luis Obispo Mothers for Peace* was pending in the Ninth Circuit, another company, AmerGen Energy, applied to NRC to renew its operating license for the Oyster Creek Nuclear Generating Station.[66] New Jersey regulators urged that NRC's NEPA review required the consideration of a threat of airborne attack on the nuclear plant.[67] In this case, as with Pacific Gas & Electric's aforementioned application, NRC concluded NEPA "imposes no legal duty on the NRC to consider intentional malevolent acts" because such acts are "too far removed from the natural or expected consequences of agency action."[68]

The Third Circuit affirmed NRC's analysis. Repeating *Metropolitan Edison*'s instruction that a mere risk of effect is not itself an effect, the Third Circuit traced the links in the causal chain, beginning with the NRC's approval of the project and ending with the hypothetical terrorist attack.[69] And in so doing, the Third Circuit identified some of the links that the Ninth Circuit had not considered. To analyze the threat of a terrorist attack would necessarily require the analysis of the nuclear facility's "status as a particularly vulnerable terrorist target"[70]—or, in terms of the five-link chain discussed above, the question of both how the terrorists would perceive the target, and whether their action on such a perceived opportunity would bear fruit. And with respect to the actual effect that such a terrorist attack would cause, that in turn depends on "[t]he government agencies specifically charged with preventing an airborne terrorist attack," which "also serve as intervening

forces."[71] Or, stated differently, "an aircraft attack on [the nuclear facility] requires at least two intervening events: (1) the act of a third-party criminal and (2) the failure of all government agencies specifically charged with preventing terrorist attacks." In light of that lengthy causal chain, the Third Circuit "conclude[d] that this causation chain is too attenuated to require NEPA review."[72]

The Third Circuit recognized that its analysis effectively created a split with the Ninth Circuit.[73] But it found that split to be no cause for abandoning its analysis and endorsing the Ninth Circuit's.

* * *

The petitioners challenging NRC's approval of the Oyster Creek relicense did not seek further review at the Supreme Court. Thus, for the time being, the Court will have no opportunity to resolve the circuit split. Nevertheless, in light of the heightened public awareness of the unexpected threats to nuclear facilities and other infrastructure that has followed the Gulf oil spill and Japanese tsunami, we can expect critics of energy infrastructure projects to cite the possibility of "Black Swan" events—terrorism, natural disasters, or otherwise—as necessary considerations in NEPA analyses. And the fundamental uncertainty of such hypothetical threats, if they are required to be considered in NEPA analyses, will only lengthen the agencies' process for reviewing infrastructure proposals, and offer litigants greater opportunity to overturn federal approvals of challenged projects.

The answer is neither to require agencies to attempt to predict and analyze the truly unpredictable, nor to give agencies *carte blanche* to ignore the possibility of such events. Rather, it is incumbent upon agencies to do their best, in good faith, to separate the reasonably foreseeable from the truly unknowable; to analyze the former as rigorously as possible; and to act prudently in light of the latter. We cannot pretend to be perfectly protected against Black Swan events, nor can we allow an exaggerated application of the Precautionary Principle to paralyze federal agencies and the development of energy infrastructure.

Endnotes

1 42 U.S.C. §§ 4321-4347.

2 U.S. Geological Survey, *Magnitude 9.0—Near the East Coast of Honshu, Japan*, *2011 March 11 05:46:23 UTC*, *available at* http://earthquake.usgs.gov/earthquakes/eqinthenews/2011/usc0001xgp/ (last visited Apr. 15, 2011).

3 Peter Behr, *Desperate Attempt to Save 3 Fukushima Reactors from Meltdown*, NYTimes.com, Mar. 14, 2011, http://www.nytimes.com/cwire/2011/03/14/14climatewire-desperate-attempts-to-save-3-fukushima-react-84017.html.

4 Keith Bradsher et al., *Japanese Officials on Defensive as Nuclear Alert Level Rises*, N.Y. Times, Apr. 13, 2011, at A5.

5 Hiroko Tabuchi, *Company Believes 3 Reactors Melted Down in Japan*, NYTimes.com, May 24, 2011, http://www.nytimes.com/2011/05/25/world/asia/25nuclear.html.

6 Andrew Morse & Mitsuru Obe, *At Plant, a Choice Between Bad, Worse*, WSJ.com, Mar. 30, 2011, http://online.wsj.com/article/SB10001424052748704471904576229854179642220.html.

7 Gautam Naik, *Earth's Energy Unleashed as Tectonic Plates Shift*, WSJ.com, Mar. 12, 2011, http://http://online.wsj.com/article/SB10001424052748703327404576194831527140212.html.

8 Phred Dvorak & Peter Landers, *Japanese Plant Had Barebones Risk Plan*, WSJ.com, Mar. 31, 2011, http://online.wsj.com/article/SB10001424052748703712504576232961004646464.html.

9 *Id.*

10 *Id.* (emphasis added).

11 *See* Press Release, U.S. Dep't of the Interior, U.S. Scientific Teams Refine Estimates of Oil Flow from BP's Well Prior to Capping (Aug. 2, 2010), *available at* http://www.doi.gov/news/pressreleases/US-Scientific-Teams-Refine-Estimates-of-Oil-Flow-from-BP-Well-Prior-to-Capping.cfm.

12 BP, Deepwater Horizon Accident Investigation Report 11 (2010).

13 *See* Marian Wang, *Despite Previous Equipment Failure, BP Says Spill 'Seemed Inconceivable'*, ProPublica, May 11, 2010, http://www.propublica.org/blog/item/despite-known-instances-of-equipment-failure-bp-says-spill-seemed-inconceiv.

14 Matthew L. Wald, *T.V.A. Considers Improvements for 6 U.S. Nuclear Reactors*, NYTimes.com, Apr. 14, 2011, http://www.nytimes.com/2011/04/15/science/earth/15nuclear.html; *see also* Matthew L. Wald, *Risk from Spent Nuclear Reactor Fuel Is Greater in U.S. than in Japan, Study Says*, NYTimes.com, May 24, 2011, http://www.nytimes.com/2011/05/25/business/energy-environment/25nuke.html.

15 *See generally* Nassim Taleb, The Black Swan (2007).

16 *Id.* at xvii.

17 *Id.* at xvii-xviii. For Taleb's more succinct explanation of his argument, *see* Stephanie Baker-Said, *Flight of the Black Swan*, Bloomberg Markets, May 2008, at 38-50, *available at* http://www.fooledbyrandomness.com/bloombergProfile.pdf.

18 *See*, *e.g.*, Shannon Harrington et al., *The Rush to Hedge Against Black Swan Events*, Businessweek, July 26, 2010; Adam Piore, *Planning for the Black Swan*, Scientific Am., June 2011, at 49-53. *See generally* Symposium, *The Age of Uncertainty*, 21 Critical Rev. 401 (2009) (debating Taleb's "Black Swan" theory).

19 *See*, *e.g.*, Richard Posner, Catastrophe: Risk and Response (2004); Cass Sunstein, Worst-Case Scenarios (2007); Francis Fukuyama et al., Blindside: How to Anticipate Forcing Events and Wild Cards in Global Politics (2007).

20 Posner, *supra* note 19, at 140-48, 171-87.

21 *See*, *e.g.*, Edward A. Parson, *The Big One: A Review of Richard Posner's* Catastrophe: Risk and Response, 45 J. Econ. Lit. 147 (2007).

22 Cass R. Sunstein, *Beyond the Precautionary Principle*, 151 U. Pa. L. Rev. 1003, 1003-04 (2003) (citing, e.g., Interpreting the Precautionary Principle (Timothy O'Riordan & James Cameron eds., 1994)).

23 Ron Suskind, The One Percent Doctrine 62 (2006).

24 Thomas L. Friedman, *Going Cheney on Climate Change*, N.Y. Times, Dec. 9, 2009, at A43. Friedman is far from the only public intellectual to draw connections between terrorism and climate change, with respect to the Precautionary Principle.

25 Sunstein, *Beyond the Precautionary Principle*, *supra* note 22, at 1020.

26 *Id.* at 1020-21.

27 Cass R. Sunstein, *On the Divergent American Reactions to Terrorism and Climate Change*, 107 Colum. L. Rev. 503 (2007); Cass R. Sunstein, *Irreversible and Catastrophic: Global Warming, Terrorism, and Other Problems*, 23 Pace Envtl. L. Rev. 3 (2005).

28 42 U.S.C. §§ 4321-4347.

29 40 C.F.R. §1500.1(a).

30 42 U.S.C. § 4331.

31 Vt. Yankee Nuclear Power Corp. v. NRDC, 435 U.S. 519, 558 (1978).

32 *See, e.g.*, 33 U.S.C. § 1341 (Clean Water Act); 16 U.S.C. § 1278 (Wild & Scenic Rivers Act); 15 U.S.C. § 717f (Natural Gas Act); 16 U.S.C. § 808

(Federal Power Act); 42 U.S.C. § 2133 (Atomic Energy Act).

33 42 U.S.C. § 4332(2)(C).

34 *See*, *e.g.*, Marsh v. Ore. Natural Res. Council, 490 U.S. 360, 373-74 (1989) ("hard look"); Calvert Cliffs' Coordinating Comm., Inc. v. AEC, 449 F.2d 1109, 1114 (D.C. Cir. 1971) (describing NEPA's requirements generally).

35 *Calvert Cliffs*, 449 F.2d at 1112. NEPA created the Council on Environmental Quality ("CEQ"), 42 U.S.C. § 4342, which in turn promulgated general regulations guiding the other agencies that apply NEPA in their own respective regulatory spheres. *See* 40 C.F.R. ch. V.

36 Envtl. Def. Fund v. Massey, 986 F.2d 528, 532 (D.C. Cir. 1993) (quoting Jones v. D.C. Redevelopment Land Agency, 499 F.2d 502, 512 (D.C. Cir. 1974)).

37 Robertson v. Methow Valley Citizens Council, 490 U.S. 332, 349 (1989).

38 *See* 42 U.S.C. § 4332(2)(C).

39 *Id.* That said, before an EIS is prepared, the relevant federal agency must first ascertain whether the EIS requirement applies at all—i.e., whether the agency proceedings involve a "major Federal action" with "significant" environmental effects. *Id.* This threshold determination may be accomplished through use of a less formal "environmental assessment." *See* 40 C.F.R. §§ 1501.4, 1508.9.

40 *See*, *e.g.*, Citizens Against Burlington, Inc. v. Busey, 933 F.2d 190, 199 (D.C. Cir. 1991).

41 Scientists' Inst. for Pub. Info. v. AEC, 481 F.2d 1079, 1092 (D.C. Cir. 1973).

42 Alaska v. Andrus, 580 F.2d 465, 473 (D.C. Cir. 1978).

43 Sierra Club v. Marsh, 976 F.2d 763, 767 (1st Cir. 1992); *see also* 40 C.F.R. § 1508.8(b) (defining a federal action's "indirect effects" as those effects that are "caused by the action and are later in time or farther removed in distance, but are still reasonably foreseeable").

44 Ethyl Corp. v. EPA, 541 F.2d 1, 18 (D.C. Cir. 1976).

45 *See* Metro. Edison Co. v. People Against Nuclear Energy, 460 U.S. 766, 774-75 (1983).

46 *Id.* at 775 (emphasis in original).

47 *Id.*

48 *Id.* at 774 (quotation marks omitted).

49 *Id.* at 774-76.

50 *Id.* at 775 n.9.

51 Final Report of the National Commission on Terrorist Attacks Upon the United States 245, 530 n.148 (2004).

52 San Luis Obispo Mothers for Peace v. NRC, 449 F.3d 1016, 1021 (9th Cir. 2006), *cert. denied*, 549 U.S. 1166 (2007).

53 *Id.* at 1022.

54 *Id.*

55 *Id.* at 1030.

56 *Id.* at 1029.

57 *Id.*

58 *Id.* at 1030.

59 *Id.* at 1029 (quoting *Metro. Edison*, 460 U.S. at 775 n.9)).

60 *Id.* at 1030.

61 *Id.*

62 *Id.* at 1031-32. In any event, the court held, the NRC had not actually supported its conclusion that the threat of terrorist was as a matter of fact unquantifiable. *Id.* at 1032.

63 460 U.S. at 775 n.9.

64 *Id.* at 775 ("[T]he element of risk *and its perception* by [observers] are necessary middle *links*." (emphasis added)).

65 *Id.*

66 N.J. Dep't of Envtl. Protection v. NRC, 561 F.3d 132, 135 (3d Cir. 2009).

67 *Id.*

68 *See id.* The NRC further concluded that a terrorism review would be "redundant" because "the core damage and radiological release from [terrorist] acts would be no worse than the damage and release to be expected from" other technical failures at the plant. *Id.* at 135-36.

69 *Id.* at 137-42.

70 *Id.* at 141.

71 *Id.*

72 *Id.* at 140.

73 *Id.* at 142.

Civil Rights

Affirmative Action for Men? Strange Silences and Strange Bedfellows in the Public Debate over Discrimination Against Women in College Admissions

*By Gail Heriot and Alison Somin**

While some news reports indicate that discrimination against women on the basis of sex[1] in college[2] admissions is increasingly common, there has been relatively little public discussion about it—especially compared to the much more heated public debate concerning race-based affirmative action. Not surprisingly, therefore, there have been few attempts to study the extent of the problem systematically. One such attempt with which we are both familiar—a study by the U.S. Commission on Civil Rights of sex discrimination at nineteen colleges and universities in the mid-Atlantic states—was unfortunately abandoned for what appear to be political rather than substantive reasons.[3] Although the fate of the Commission's probe may in part be explained by the Commission's institutional quirks, the muddled politics surrounding the attempted probe may reflect in microcosm the muddled politics of the broader national debate. In this article, we discuss those politics and suggest that the lack of attention the issue has received to date may be unfair.

The scope of the gender discrimination problem: Multiple news reports indicate that some colleges and universities, both public and private, have what they regard as "too many" women applicants and are therefore discriminating in favor of men—largely because more women than men apply to college and their academic credentials are in some ways better.[4] Several colleges have more or less openly admitted to discriminating against women – including the University of Richmond[5] (a private institution) and the College of William and Mary (a public institution).[6] Others—including Southwestern University (Texas),[7] Knox College (Illinois),[8] Brandeis University (Massachusetts),[9] Boston University (also Massachusetts),[10] and Pomona College (California)[11]—shy away from admitting directly that they are discriminating, but admit that maintaining an optimal gender balance by non-discriminatory means is difficult. Trustees at the University of North Carolina have proposed instituting affirmative action for men, but ultimately decided against doing so.[12]

Sex discrimination in admissions at public universities is illegal under Title IX of the Education Amendments of 1972.[13] But under federal law, it is perfectly legal for private institutions to engage in sex discrimination in admissions—though once both sexes are admitted, neither may be discriminated against. There has been relatively little litigation regarding these preferences. We are aware of just one case—in 2000, several plaintiffs brought suit against the University of Georgia challenging the legality of preferences for men along with preferences for under-represented racial minorities.[14] The district court found that the preference for men was indeed illegal under Title IX and the Equal Protection Clause of the Fourteenth Amendment, but the university opted not to appeal the gender issue to the Eleventh Circuit.[15]

Perhaps the most attention-getting piece on this topic was a 2006 *New York Times* op-ed by Jennifer Delahunty Britz, an admissions officer at Kenyon College, in which she admitted that her office often gave preferential treatment to men. Some admissions insiders wrote in response to Delahunty Britz's piece that these preferences were quite common—what was shocking was only Delahunty Britz's candor in airing this information publicly.[16] *Inside Higher Ed* noted that "[w]hile few admissions officers wanted to talk publicly about the column, the private reaction was a mix of 'of course male applicants get some help' along with 'did she have to share that information with the world?'" Several years later, after the wave of chatter over Delahunty Britz's piece had died down, Columbia University law professor Ted Shaw referred to such discrimination as an "open secret."[17]

Ultimately, while Delahunty Britz's piece did touch off a wave of chatter elsewhere in the media and in blogs, it did not lead immediately to any attempts to study the problem systematically.[18] One article in *U.S. News and World Report* did make some attempt to quantify the problem by printing a table of eighteen schools that have particularly pronounced differences in admission rates between men and women. At Wheaton College, the school with the largest gender gap of the schools examined, the male admissions rate was twenty-one percentage points higher than the rate for females; at Grove City College, the next school on the list, it was fifteen percentage points higher.[19] While these numbers are indeed eye-catching, there is a possible non-discriminatory explanation for them: the male applicant pool at these particular colleges may simply be stronger. For whatever reason, women who are marginal candidates may simply be more likely to send in applications to these particular institutions. Without studies controlling for relevant credentials, we cannot know for sure. In addition, even if the applicant pools turn out to be precisely identical, these figures do not show how large the preferences for male applicants are measured in SAT score points or in GPA.

The U.S. Commission on Civil Rights made perhaps the best publicized attempt of which we are aware to look at

** Gail Heriot is a law professor at the University of San Diego and one of eight members of the U.S. Commission of Civil Rights ("the Commission"), a bipartisan, federal agency that makes "appraisals of the laws and policies of the Federal Government . . . with respect to discrimination or denials of equal protection of the laws under the Constitution of the United States because of color, race, religion, sex, age, disability, or national origin, or in the administration of justice." 42 U.S.C. 1975(a). Alison Somin is Gail Heriot's special assistant and counsel at the Commission. The views expressed in this article are those of the authors alone and are not necessarily those of the Commission.*

sex discrimination in admissions.[20] In September 2009, the Commission voted to examine gender discrimination at a number of colleges and universities in the mid-Atlantic states. The late Dr. Robert Lerner, the head of the Commission's Office for Civil Rights Evaluation and Research, designed a study that would look at admissions decisions at nineteen different colleges and universities in that region. The schools were chosen so as to ensure a diverse sample in terms of size, selectivity, religious affiliation, and status as a historically black institution. The Commission's researchers would control for entering credentials such as high-school grade point average and SAT score to see if colleges appeared to be granting preferential treatment to one sex. During an eighteen-month long investigation, the Commission obtained data from fifteen colleges and universities— Lincoln University of Pennsylvania; University of Maryland-Eastern Shore; Virginia Union University; Howard University; Catholic University of America; Loyola College in Maryland; University of Richmond; York College, Pennsylvania, Goucher College; Goldey-Beacom College; Washington College; Shepherd University; Shippensburg University of Pennsylvania; the University of Delaware; and the University of Maryland at Baltimore County. After much resistance, three others—Georgetown University, Johns Hopkins University, and Gettysburg College—agreed to crunch their own numbers following the Commission's research protocols.[21]

Yet in a surprising March 2011 vote, a majority of members of the Commission voted to cancel the long-established study just as it was coming to fruition. No notice had been given to the study's supporters on the Commission that such a vote would be taken. Eighteen months' work was all gone in an instant.

Members of the Commission who voted to terminate the study claimed that the limited geographic scope of the project and shortcomings in the data of the project motivated their vote. But both were makeweight arguments. The argument that it was necessary for the Commission to undertake the expense of a national study rather than a study of schools in the mid-Atlantic region in order to better understand sex discrimination is decidedly unpersuasive. There is no evidence at this point that gender discrimination is a more severe problem in some regions than in others. As noted earlier, colleges in locations as diverse as Ohio, Georgia, Virginia, Texas, Massachusetts, Illinois and California have all admitted to struggles with gender imbalance. The Commission chose to study schools in the mid-Atlantic region because there could be no argument that the reach of its subpoenas issued by its national office extended at least that far. If, after completing the study, anyone was concerned that results from the mid-Atlantic states were not definitive, the Commission could have chosen to view this original project as a pilot study and expanded it accordingly. But to undertake a study of the entire country right off the bat would have been a waste of Commission resources.

As for the notion that there would have been shortcomings in the data, anyone who has ever done large-scale research on admissions policies (or any other large-scale social science research) knows there are always shortcomings in the data. There will be missing values because somebody forgot to check a box indicating whether a "Stacey Smith" or "Leslie Jones" is a he or a she. Some schools will accept students with either SAT or ACT scores, thus making it harder to compare one student's treatment to another's. In the end, some results will be statistically significant and some won't be. But there was nothing wrong with the data in this study that would remotely justify tossing the whole study out. There were fears that data from one of the schools in the study would turn out to be unusable. But given that eighteen schools were selected in the hope that the data would pan out on fifteen or so, the results had actually exceeded the expectations of staff researchers, as well as the commissioners who had voted in favor of the study.

It was obvious something else was going on. One might suspect that there had been a change in the Commission's membership. And one might suspect that the new members did not want a study on discrimination against women in higher education. But who? Was it the conservatives, who have been accused of not caring about women?

Well . . . no. There had indeed been a change in the Commission's makeup. In September 2009, when the project was undertaken, it was supported by five of six Republican-appointed members and neither of the two Democratic-appointed members of the eight-member Commission. By March 2011, when the project was abandoned, two of the Bush appointees had been replaced by Obama appointees.[22] As Andy Ferguson wrote in *The Weekly Standard*, "[T]he politics are very odd. [Republican appointees with generally right-of-center views] might be thought by the usual ideological taxonomy to be reluctant to press an investigation into wholesale discrimination against girls. On the other hand, the project should have been meat-and-taters to the Democrats—a chance to expose a concerted effort by large, wealthy, unaccountable institutions to deny an education to qualified women purely on the basis of their sex."[23] Indeed, Ferguson is not the only commentator to have noted the odd political valence of the issue. Richard Whitmire, an education reporter for *USA Today*, titled an article about the project's cancellation "The Muddled Politics of Male Gender Preferences,"[24] and conservative writer Charlotte Allen—in a piece that was somewhat critical of the Commission's project just as Ferguson's and Whitmire's were sympathetic—described the coalition that has coalesced around the issue as a group of "strange bedfellows."[25]

While Allen is correct to characterize the coalition forming in support of male preferences as "strange bedfellows," she is incorrect to describe the coalition *opposing* these preferences as comprised of "opponents of affirmative action for any group" and "hard-line feminists." And therein lies one of the more remarkable facts about the debate over affirmative action for men: contrary to what one might expect, most feminists, hard-line or otherwise, have said little or nothing at all about the issue. The lone prominent feminist whom Allen quotes as having spoken out about Jennifer Delahunty Britz's column—Katha Pollitt—never wrote or spoke publicly about the Commission's investigation, to our knowledge. In one *Inside Higher Ed* article about the Commission project, all of the representatives from feminist groups quoted were actually opposed to it.[26]

What could cause even the most hard-line feminists to turn a blind eye to what is probably among the more blatant forms of sex discrimination one can find domestically these

days? While there may be several plausible reasons to want to maintain the ability of private schools to discriminate on the basis of sex, many feminists have been eager to denounce anything that seems to even superficially work to women's detriment.[27] Is there some special reason that sex discrimination in college admissions—even at public institutions where it is illegal—does not bother them?

We can only speculate as to the reasons for their strange silence, although commentators have offered up a few intriguing and plausible-sounding possibilities. Some have claimed that women's groups have been reluctant to speak out about preferences for men because they realize that preferences exist because men are falling behind in K-12 education. If men are falling behind, then women have "won" the gender wars, and feminist organizations are no longer necessary—or so the argument goes.[28] Others claim that feminist groups are reluctant to speak out on this issue for fear that undermining affirmative action for men will lead to the undermining of affirmative action for racial and ethnic minorities, and these groups are part of the broad coalition of left-leaning activists first and advocates for women second.[29] We note only that the website for the Feminist Majority Foundation devotes far more space to a "Say No to Pesticides!" campaign than it does to sex discrimination in admissions.[30]

We find a third possible reason for feminists' silence to be especially intriguing—although we do not claim it fully explains the attitudes of feminist organizations. After years of failure to comment on sex discrimination in admissions, feminist opposition to the Commission study focused on a single line in the project proposal to justify their opposition to the study. That line read, "A small but significant problem may lie in the enforcement policies of the Department of Education" with regard to Title IX.[31]

The reference in the project proposal is to the method by which the Department of Education ensures compliance with Title IX in the area of athletics. In the almost forty years since Title IX's passage, a disproportionate share of the attention of enforcement officials has been focused on sex discrimination in athletics—so much so that non-experts are sometimes under the mistaken impression that Title IX is a law that forbids sex discrimination in athletics rather than sex discrimination more generally. What was originally a relatively obscure enforcement policy—the Department of Education's *A Policy Interpretation: Title IX and Intercollegiate Athletics*—may be having perverse and unintended effects on college and university admissions.

Unlike other college and university programs, most athletic programs are sex-segregated. Given that it is not self-evident that if given the choice men and women would choose the same kinds or the same level of athletic activity, athletics tends to raise the thorniest issues under Title IX. In chemistry classes, achieving equality is relatively easy: both women and men should be permitted to enroll, there should be no distinction made between men and women in the use of laboratory resources, and both men and women should be graded on the same scale. In athletics, it is not so easy. Since schools ordinarily prohibit men from joining the women's field hockey team or women from joining the men's wrestling team, some other way of establishing equality must be devised. Should schools have to spend the same total number of dollars on each sex? What if there are more male than female students at a given school? What if there are more male than female students interested in athletics? How should schools determine whether a student is interested in athletics or not? What if more women students prefer that resources be put in other extracurricular activities, such a chorus or theater—activities that men can participate in too, but sometimes choose not to?[32]

There is plenty of evidence that at this particular point in history, female students are somewhat less likely to be interested in sports than male students.[33] Nevertheless, the Department of Education has devised guidelines which make it extremely difficult for any school to do anything but presume that men and women are equally interested in athletics and spend accordingly. Before an extra slot can be created on a men's team, schools must show that women's interest has been fully met, and the level of proof demanded is essentially prohibitive.[34] Smart schools resign themselves to "substantial proportionality." In other words, a university that is 60% female has little choice but to offer 60% of its slots on athletic teams to women—even if there are not enough women interested in playing sports and willing to take these slots and even if it means cutting back on athletic opportunities for men.

Such an approach leaves colleges and universities that fear "gender imbalance" in a bind. Offering male students the opportunity to engage in or watch athletic competitions is a time-honored method for recruiting them as students. In recent years, it has been a particularly useful strategy for some small liberal arts schools that view themselves as lacking a sufficient number of men.[35] Because these schools ordinarily do not compete in Division I of the NCAA, they can offer some students, who would almost certainly be overlooked by the sports powerhouse schools, a chance to compete. But the Department of Education makes this strategy difficult.

The problem with the Department of Education's "substantial proportionality" approach is that in its zeal to prevent what is likely phantom sex discrimination in athletics, it can end up encouraging real sex discrimination in admissions. Because it is perfectly legal to discriminate against women in admissions, as discussed *supra*, the substantial proportionality approach can have the unintended consequence of encouraging universities to discriminate in admissions rather than attempting to attract male students by offering them more athletics, knowing that, if they do, they will also have to shell out for female sports teams for which there is little or no interest.

Several feminists have taken the public position that they oppose the Commission's study because it may touch on this issue. This seems odd to us on many levels. For one thing, as is clear from the full text of the Commission's proposal, the Commission's first priority was to get hard facts indicating whether discrimination was occurring at all. One would think that would be a priority for feminist organizations too. Discussion of the appropriate policy response is something that would come later if the study had confirmed that a substantial number of schools were in fact engaged in discrimination against female applicants.

Of course, it may be that feminists simply misunderstood the concerns about Title IX expressed in the proposal. One

remarked, for example, that "[t]he goal of this [i.e. the Commission's proposed] approach would be to stop schools from discriminating in admissions by permitting them to discriminate in athletics."[36] The proposal does not state—and neither of us thinks—that Title IX permits or should permit universities that receive federal funds to discriminate against women in offering athletic opportunities. The problem, rather, lies with the "substantial proportionality" approach that the Department of Education has adopted in enforcing Title IX. Under current law, women students are presumed to be too burdened by cultural stereotypes to express a desire to play varsity sports even when directly asked by the school officials via internet survey.[37] In lieu of the survey method, university officials are instead supposed to engage in a range of different activities to determine whether and to what extent there is unmet female interest. One Department of Education document states that to determine female interest in sports, colleges should among other things look at "participation rates in sports in high schools, amateur athletic associations, and community sports leagues that operate in areas from which the institution draws its students in order to ascertain likely interest and ability of its students and admitted students in particular sport(s)."[38] So national universities like our respective alma maters are perhaps supposed to look at female participation rates in sports at reliable feeders like Phillips Exeter Academy or Stuyvesant High School in New York City. Or maybe they are supposed to look at patterns of female athletic interest at secondary schools that do send them students occasionally if more rarely, such as Raffles Junior College in Singapore or Palmer High School in Palmer, Alaska (population 8,201), the high school alma maters of one author's college classmates. The Department of Education offers colleges no guidelines on what they are supposed to do if amateur athletics associations in Singapore compete in different sports than community sports leagues in Palmer, Alaska or if members of both such organizations have totally different preferences than Exeter and Stuyvesant students do. It is not surprising that most universities conclude that they have no choice except to throw up their hands regarding measuring possible unmet female interest and instead just to opt for quotas.

We advocate the simpler approach of sparing universities the trouble of beating the bushes of Singapore looking for female athletes and of showing some respect for college women's choices. If women want more opportunities for athletics, they can be expected to answer survey questions forthrightly.

Our fear is that some of these commentators do understand our concerns about the unintended consequences of current Title IX enforcement, but that they and the organizations that employ them are committed to an athletic-centric view of Title IX. They have built their reputations on this issue and cannot retreat, no matter what the unintended consequences turn out to be. In any event, it seems unlikely to us that many women, given the choice, would find being flat-out denied admission to a particular college preferable to not having an opportunity to participate in competitive ice fishing because that college failed to make a sufficiently careful study of trends in Alaskan community athletics association membership.[39]

Then there is the right half of the political spectrum. Unlike those on the left, conservatives appear to have controversy within their ranks regarding affirmative action for men—even a wee bit of acrimony. A few months after the Commission undertook its project on sex discrimination, conservative columnist Mona Charen, published a column entitled "Civil Rights Commission Blunders Again." In it she argued that the Commission "is about to subtract from national wisdom about college admissions by focusing on exactly the wrong problem." Charen argued her point this way:

> [W]e seem to have a boy problem here. For every 100 women who earn a college degree, only 73 men do. These statistics practically shout 'boy crisis.' Yet the Civil Rights Commission apparently sees the problem as one of discrimination.[40]
>
> . . .
>
> . . . The Civil Rights Commission can do us all a favor by going away.

In her essay, Ms. Charen argues that boys are shortchanged at K through 12 levels by "feminized school environments; that they are also disadvantaged by their family environments. This somehow justifies preferences at the college level. This argument parallels one of the more common arguments in favor of preferential treatment for under-represented minorities—that African Americans, Hispanics, and American Indians should receive preferential treatment at the college level, because they are shortchanged in K through 12 and/or in their family environment. (We note that on other occasions, Ms. Charen has supported California's Proposition 209, which bans both race and sex discrimination in, among other things, admissions to state colleges and universities.)[41] We would respond to those arguments the same way as we have to the argument in the race context.

We are perfectly willing to entertain the possibility boys are being shortchanged in K through 12 and/or in their family environments. (This point can be overstated, however, since part of the reason men are relatively scarce in higher education is that they are more likely to prefer to enter the military or the building trades, both honorable paths in life.) But if it is true that boys are being shortchanged, the problem needs to be corrected early on, while they are still in K through 12. It will do no good to paper over it at the college level, and it may well do harm.

There is now abundant evidence that accepting an affirmative action leg-up hurts a minority student's chances of becoming a doctor, scientist, or engineer.[42] A better strategy is to attend a school at which one's entering academic credentials roughly match the median student's. It is entirely possible that affirmative action is similarly backfiring for male students accepting a preference—and that preferences are thus leading to fewer rather than more male students fulfilling their ambition to become a physician, scientist, or engineer.[43] That is one among many reasons the current state of affairs is worth investigating.

Moreover, if the source of boys' difficulty in K through 12 is "feminized school environments" such that boys lose interest in school, Ms. Charen should have been positively

enthusiastic about the Commission's project and especially about examining the Department of Education's substantial proportionality rule.[44] Discrimination in admissions allows colleges and universities to perpetuate any such feminization. By enforcing the law against sex discrimination in admissions that cover public colleges and by promulgating similar laws that would cover private colleges, the federal government would be strongly encouraging schools to adopt other methods of attracting male students—like offering them more athletics, more programs in engineering and physical science, and perhaps even a few frogs and snails and puppy dog tails or whatever it is that is supposed to please male students.

We note for example that Brandeis University recently tried the "puppy dog tail" approach with a promotion that gave free baseball caps to the first 500 men who applied there—a gimmick that we believe is unlikely to have much impact.[45] Other institutions have engaged in creating academic or vocational programs designed to appeal to men, such as institutes for building and construction and a "motor sports program," which one university official described as an engineering program in disguise.[46] All of this is exactly what Charen should wish to encourage in her quest to counter any feminization of colleges and universities.

Some social conservatives dubbing themselves "biological realists" have claimed that preferences are necessary because women want or need to date men in college so that they can marry and have families within a narrow biological window of opportunity.[47] We certainly do not reject this argument out of hand. But we have several comments: First,we should note that nobody seems to have asked actual college-age women about this. The proposal for the Commission study suggested making efforts to get actual data on this question.[48] Second, sex discrimination in admissions probably does not increase the total number of male students attending college; it simply re-arranges them. From the standpoint of an individual school attempting to please its female "customers," sex discrimination in admissions may seem like a crowd pleaser, but from the standpoint of the system as a whole, there is no net benefit and hence it is in some ways gratuitous. Third, even if women generally prefer being discriminated against at the admissions level to attending schools where women dominate, that is not the whole of the issue. We very much doubt they prefer the Department of Education's misguided Title IX enforcement policies relating to athletics to sex-blind admissions policies. Put differently, if allowing colleges to create more opportunities for men to play sports—so long as they can show with reasonable evidence, like internet surveys, that they have already met the demand for opportunities in women's athletics—would attract more qualified men and hence obviate the need to discriminate in favor of less-qualified men, we strongly suspect that most women would prefer it.

Finally, there is another possible "biological reality" that deserves some thought: It has been our experience at least that many women want men who are at least as smart as they are and perhaps smarter. Without data, of course, it is difficult to do anything but speculate. But consider this: One of the consequences of widespread *race*-preferential admissions policies is that talented African American students end up distributed among colleges and universities in very different patterns from those of their white and Asian counterparts. When the schools that are highest on the academic ladder relax their admissions policies in order to admit more under-represented minority students, schools one rung down must do likewise. Otherwise, they will have far fewer minority students than they would have had under a general color-blind admissions policy. The problem is thus passed on to the schools another rung down, which respond similarly. As a result, students from under-represented minorities today are overwhelmingly at the bottom of the distribution of entering academic credentials at most selective colleges and universities. The problem with sex-preferential admissions is almost certainly not as great. But we are concerned that it may be greater than many realize. Is this really something "biological realists" or conservatives concerned about the "feminization" of schools would want to ignore?

This isn't an easy issue. We don't pretend to have all the answers. All we can offer at this point is food for thought. The one thing we are quite sure of is this: Despite the fact that neither those on the left nor those on the right want to "own" this issue, it cannot be wisely ignored. We very much regret that the Commission's study was cancelled after eighteen months of work.

Endnotes

1 Although we are aware that there is active and rancorous debate within the academic and policy worlds regarding the appropriate use of "sex" and "gender," the two terms are used interchangeably throughout this article.

2 The terms "college" and "university" are also used interchangeably throughout this piece.

3 Transcript of Business Meeting, U.S. Commission on Civil Rights, Mar. 11, 2011 at 125; Daniel de Vise, *Federal Panel Ends Probe of College Gender Bias*, Wash. Post, Mar. 16, 2011, *available at* http://www.washingtonpost.com/local/.../feds.../ABoIQ5g_story.html.

4 These days many indicators suggest that there are simply more qualified women than men applying to college. Although women continue to receive lower average scores than men on the math section of the SAT, women outscore men on the writing section of the test. Interestingly, although women receive slightly lower average scores on the "Critical Reading" section, there are more women than men scoring above 700 on this test section in part because more women take the test. A higher percentage of the women taking the test reported to the College Board that they had either an A range or B range grade point average. Similarly, higher percentages of women than men report being ranked in either the top tenth or second tenth of their high school classes. By contrast, a greater percentage of the men taking the test reported that they are in the bottom three-fifths of their high-school classes. Higher percentages of women than men also reported taking honors or Advanced Placement classes in English, mathematics, the natural sciences, history, and foreign languages. The College Bd., 2010 College-Bound Seniors: Total Group Profile Report, 1-8 (2010), *available at* http://professionals.collegeboard.com/profdownload/2010-total-group-profile-report-cbs.pdf (last accessed August 10, 2011).

5 Alex Kingsbury, *Many Colleges Reject Women at Higher Rates than for Men*, *U.S. News & World Rep.*, June 16, 2007, *available at* http://www.usnews.com/usnews/edu/articles/070617/25gender.htm.

6 Henry Broaddus, the director of admissions there, told *U.S. News and World Report* that "[e]ven women who enroll . . . expect to see men on campus. It's not the College of Mary and Mary; it's the College of William and Mary." *Quoted in id.* The College of William and Mary was named for a husband and wife pair of British monarchs who ruled jointly from 1689 to 1694. They are the only two monarchs to have ruled jointly in the millennium-long history

of the British monarchy. It is perhaps fortunate for Virginian women that the British Jacobites were not successful at their project of restoring James II to the throne in the 1690s. Otherwise, the College of William and Mary might well have been named the College of King James II instead, and by Broaddus's logic, women apparently should then never have been admitted there at all. But we digress.

7 Mark Clayton, *Overview: The Gender Equation*, Christian Sci. Monitor, May 22, 2001, *available at* http://www.csmonitor.com/2001/0522/p11s2.html.

8 *Id.*

9 Melana Zyla Vickers, *Where the Boys Aren't*, Wkly. Standard, Jan. 2, 2006.

10 *Id.*

11 Lauri Valerio, *Federal Government Investigates Gender Discrimination at Liberal Arts Colleges*, Student Life, Nov. 18, 2009, *available at* http://tsl.pomona.edu/new/index.php?option=com_content&view=article&id=577&Itemid=67.

12 Scott Jaschik, *Gender Gap at Flagships*, Inside Higher Ed, June 3, 2005, *available at* http://www.insidehighered.com/news/2005/06/03/gender.

13 *See* 20 U.S.C. 1681(a)(1), which reads in relevant part:

> (a) Prohibition against discrimination; exceptions. No person in the United States shall, on the basis of sex, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any education program or activity receiving Federal financial assistance, except that:
>
> > (1) Classes of educational institutions subject to prohibition. in regard to admissions to educational institutions, this section shall apply only to institutions of vocational education, professional education, and graduate higher education, and to public institutions of undergraduate higher education.

14 Johnson v. Bd. of Regents, 106 F. Supp. 2d 1362, 1376 (S.D. Ga. 2000). Notably, the case came down before the prominent pair of Supreme Court cases addressing race-based affirmative action in higher education—*Gratz v. Bollinger*, 539 U.S. 244 (2003), and *Grutter v. Bollinger*, 539 U.S. 306 (2003). The system that the University of Georgia was using more closely resembles the point-based system at issue in *Gratz* than the holistic system used in *Grutter*. So it might be that the Supreme Court would, post-*Gratz*, also find point-based systems unconstitutional in the gender context. But because gender classifications receive less scrutiny than do racial classifications, the Court might well decide that gender-based point systems are constitutionally permissible even if race-based point systems are not. It would indeed be interesting to find out if Georgia returned to using the point-based system for men after *Grutter*. That said, as many commentators have pointed out, there is little, if any, difference between how "holistic" systems of review work in practice compared to point-based ones.

15 *Id.*

16 17 Scott Jaschik, *Affirmative Action for Men*, Inside Higher Ed, Mar. 27, 2006, *available at* http://www.insidehighered.com/news/2006/03/27/admit. Elsewhere, Swarthmore College history professor Timothy Burke commented on his blog, "I have been a bit surprised about how surprised some observers are about Jennifer Delahunty Britz's op-ed piece in the New York Times regarding the role of gender in the admissions process at many selective private colleges and universities. . . . Unfortunately for anyone who objects to this approach [i.e. admissions preferences for men], it's actually pretty common." Kenyon's Confession, Posting of Timothy Burke to Easily Distracted, http://weblogs.swarthmore.edu/burke/2006/03/28/kenyons-confession/ (Mar. 28, 2006).

17 *See* Theodore M. Shaw, Professor, Columbia Law School, Remarks at the Heritage Foundation in Panel Discussion Sponsored by the Federalist Society's Civil Rights Practice Group and the Heritage Foundation on "Civil Rights in the Age of Obama" (May 13, 2009), *available at* http://www.fed-soc.org/publications/pubID.1416/pub_detail.asp.

18 One odd angle of this story is that Delahunty Britz seems to have publicly retreated from the position that she took in her *New York Times* op-ed. In an essay posted on Kenyon's website, she writes, "Some critics accuse Kenyon of favoring girls; others say we're favoring boys. The fact is we're favoring neither." Jennifer Delahunty Britz, A Personal Statement from Jennifer Delahunty Britz, http://www.kenyon.edu/x31612.xml (last visited Apr. 22, 2011). And in an NPR radio interview in November 2009, Delahunty Britz said, "I don[']t see anybody who has a policy that says, you know, that we[']re going to choose less qualified boys over more qualified girls. No admissions office would make those kinds of decisions." Yet these statements seem to directly contradict her account of one particularly difficult decision she and her colleagues in the Kenyon admissions office had to make: "Few of us sitting around the table were as talented and as directed at age 17 as this young woman. Unfortunately, her test scores and grade point average placed her in the middle of our pool. We had to have a debate before we decided to swallow the middling scores and write 'admit' next to her name. Had she been a male applicant, there would have been little, if any, hesitation to admit." Delahunty Britz indicated that this instance was far from atypical: "The reality is that because young men are rarer, they're more valued applicants."

We cannot say for sure what motivated the change in Delahunty Britz's public stance. Did her perception of Kenyon's admissions policies change? Was she afraid that Kenyon would retaliate against her for publishing a controversial newspaper column? In any case, her about-face may have had a small negative impact on efforts to stimulate a national conversation on gender discrimination in admissions—which is unfortunate.

19 Alex Kingsbury, *Education: Many College Reject Women at Higher Rates Than for Men*, U.S. News & World Rep., June 17, 2007. Some have speculated that sex discrimination in admissions is most pronounced at mid-level private colleges. While some schools fitting this description show up on Kingsbury's list—e.g., Skidmore, Richmond, Providence, and Wheaton—it may be important to note that no particular type of school dominates it. Regarding the "mid-level" claim, Kingsbury claims that the most selective universities in the country—the Ivy League schools and a handful of other institutions that are equally selective—generally have no difficulty attracting highly qualified men. Indeed, no Ivy League colleges show up on his list. But some colleges that do appear—including Swarthmore, Pomona, Vassar, and William and Mary—are equally or only slightly less selective. Several large public universities—Georgia State University, the University of Texas at Dallas, and Rutgers-Newark—also make appearances. The campus political climates of the universities on the list are equally varied. Grove City is generally considered among the most conservative universities in this country, whereas Vassar students are usually regarded as leaning politically far in the opposite direction. The diversity among institutions showing up on this list only underscores the need for broad probes like the Commission's.

20 See authors' byline for a brief description of the Commission's history, functions, and duties and the authors' connections to the Commission.

21 Although the Commission has the authority to issue subpoenas and it did so in this case, to enforce those subpoenas it depends on the Department of Justice. The Commission never requested the Department of Justice to step in.

22 Five commission members—Chairman Gerald Reynolds, Peter Kirsanow, Ashley Taylor, Gail Heriot, and Todd Gaziano—voted to approve the sex discrimination project in September 2009. Reynolds, Kirsanow, and Taylor were all appointed by President Bush. Heriot was appointed by President pro Tempore of the Senate on the recommendation of Minority Leader Mitch McConnell in 2007. Gaziano was appointed by the Speaker of the House upon recommendation by then-Minority Leader John Boehner 2008.

Three commission members were not present when the vote took place, having left the meeting earlier in an effort to defeat quorum for tactical reasons. These included Abigail Thernstrom, who was appointed to the Commission by President George W. Bush, but who has caucused with Democratic appointees rather than Republican appointees for the past two years; Michael Yaki, who was appointed on the recommendation of Minority Leader Nancy Pelosi; and Arlan Melendez, who was appointed on the recommendation of Senate Minority Leader Harry Reid.

In December 2010, Reynolds's and Taylor's terms expired. In January 2011, President Obama nominated their replacements, Martin Castro (who was later chosen as Chair) and Roberta Achtenberg. Arlan Melendez's term expired in December 2010 as well. Harry Reid appointed Dina Titus, a former Democratic congresswoman from Nevada, as his replacement in December 2010. Titus has since resigned from the Commission to pursue a run for congressional office. Commissioner Michael Yaki's term also expired in December. He was reappointed in April 2011, a month after the Commission's sex discrimination in admissions project was cancelled.

23 Andy Ferguson, The Quotas Everyone Ignores, Wkly. Standard, Mar. 28, 2011, *available at* http://www.weeklystandard.com/articles/quotas-everyone-ignores_554831.html.

24 Richard Whitmire, *The Muddled Politics of Male Gender Preferences*, Educ. Wk., Mar. 18, 2011, *available at* http://blogs.edweek.org/edweek/whyboysfail/2011/03/the_muddled_politics_of_male_gender_preferences.html.

25 Charlotte Allen, *The Quiet Preference for Men in Admissions*, Minding the Campus, June 7, 2010, *available at* http://www.mindingthecampus.com/originals/2010/06/the_quiet_preference_for_men_i.html.

26 Scott Jaschik, *Title IX Trojan Horse?*, Inside Higher Ed, Nov. 3, 2009, *available at* http://www.insidehighered.com/news/2009/11/03/titleix.

27 *See* Int'l Union, United Auto Workers v. Johnson Controls, 499 U.S. 187 (1991).

28 *See, e.g.*, Richard Whitmire, *Missed Opportunity*, Inside Higher Ed, Apr. 22, 2011, *available at* http://www.insidehighered.com/views/2011/04/22/essay_criticizing_end_of_inquiry_into_possible_bias_against_female_applicants_to_colleges. We do not believe it logically follows from the fact women do better than men in some walks of life that they are doing better in all walks. But it might take the punch out of some of the more extreme feminist rhetoric to have to concede how much better girls are doing than boys in high schools today.

29 *See, e.g.*, Richard Whitmire, *The Muddled Politics of Male Gender Preferences*, Educ. Wk., Mar. 18, 2011, *available at* http://blogs.edweek.org/edweek/whyboysfail/2011/03/the_muddled_politics_of_male_gender_preferences.html. ("The groups you would expect to complain the loudest about discrimination against women—national feminist groups and the American Association of University Women—are mute. Why are they okay with discriminating against females? After watching the issue for years I've come up with three explanations. First, liberals want to safeguard the freedoms granted to admissions officials to pick the freshman class they want, which includes putting the thumb down to favor minorities. They have a good point. Male preferences, for example, are no different from favoring a potential star quarterback, cello player or legacy admit whose parents are likely to finance a new dorm.")

Indeed, the debate over affirmative action for men throws into relief the tension between two of the more commonly offered justifications for affirmative action in the race context—the "compensatory justice" and "diversity" rationales.

The compensatory justice rationale for affirmative action holds that preferences are permissible when they are used to compensate a group for past discrimination or mistreatment. The diversity argument for preferences, by contrast, holds that preferences ought to be permissible when they help ensure that a broad variety of viewpoints will be represented on campus. As some commentators have hinted, it is difficult to justify preferences for men under the compensatory justice rationale for affirmative action. Men are historically thought to have been the privileged sex relative to women; there is no apparent need for remedial discrimination to help them. *See, e.g.*, Liz Dwyer, *Rejected from College: If You're a Woman, A Less Qualified Man Probably Took Your Spot*, Good, Apr. 25, 2011, *available at* http://www.good.is/post/rejected-from-college-if-you-re-a-woman-a-less-qualified-man-probably-took-your-spot/.

In some ways, it is easier to justify affirmative action for men under a diversity rationale for affirmative action. Yet ultimately, the argument even there is problematic. *Grutter v. Bollinger*, 539 U.S. 306 (2003), the leading court case spelling out the diversity rationale for affirmative action, emphasized that colleges and universities may use racial preferences to ensure a "critical mass" of students from a particular group on campus—e.g., that there are enough black students present to ensure that other black applicants do not decline to apply out of concern about racial isolation. *Id.* at 330. Yet the same argument makes less sense in the sex context; that is, it is hard to argue that somehow a university that is 40% male lacks a "critical mass" of men, but that one that is 50% male has reached said "critical mass." Similarly, attending a university that is 40% male can almost surely convey to a female student the range of male perspectives on a subject and can lead her to understand that not all men think the same way. Ultimately, the only real way that one can defend preferences for men under the "diversity" rationale is by equating "diversity" with "proportional representation"—which advocates of the diversity model have repeatedly and vocally claimed is not their goal.

For further elaboration of the tension between these two justifications for affirmative action in the race context, see Asian-American Applicants and Competing Rationales for Affirmative Action in Higher Education, Posting of Ilya Somin to The Volokh Conspiracy, http://volokh.com/2009/10/17/asian-american-applicants-and-competing-rationales-for-affirmative-action-in-higher-education/ (Oct. 17, 2009, 01:54 EDT); Immigrant Students and the Tension Between Two Rationales for Affirmative Action, Posting of Ilya Somin to The Volokh Conspiracy http://volokh.com/archives/archive_2007_02_04-2007_02_10.shtml#1170741033 (Feb. 5, 2007, 23:50 EST). In those two posts, Somin argues that the diversity rationale leads to perverse results in the cases of recent African immigrants—students who have black skin, but whose ancestors were never American slaves—and to discrimination against Asian-Americans—who, like women, have historically been victims of discrimination but who are currently over-represented in higher education. Although the problem of discrimination against Asian-Americans is of course different from that of discrimination against women, the two groups are similarly situated in that they are similarly harmed by the diversity model of affirmative action.

Some may argue that young men deserve preferences in college admission because they have been unfairly victimized by the excesses of the feminist movement. In some sense, this is a compensatory justice argument. We agree that some of the practices complained of are ridiculous; we discuss some of them *infra* at 17-18. At the same time, the severity of the alleged injuries to young men appear slight compared to the harms suffered by groups on whose behalf the compensatory justice arguments have more commonly been advanced, e.g., African-Americans. In any case, we worry that using affirmative action in this way will only distract attention from real problems in the education of boys in grades K-12. See 17-18, *infra*.

30 *See* Say NO to Pesticides?, Choices Campus Blog, http://feministcampus.blogspot.com/2011/04/say-no-to-pesticides.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+FeministCampusBlog+%28Choices+Feminist+Campus+Blog%29 (Apr. 19, 2011, 14:39). The National Women's Law Center's website contains no issue papers, press releases, blog posts, or other materials that we could find about affirmative action for men in university admissions. At the National Organization for Women's website, there is a section devoted to "Affirmative Action," http://www.now.org/issues/affirm/ (last visited Apr. 25, 2010), that solely contains papers about the need to protect affirmative action on behalf of racial and ethnic minorities. A page titled "Education and Title IX" is equally silent about admissions preferences for men.

31 A copy of the project proposal is available online at http://chronicle.com/article/Full-Text-The-Proposal-That/49012/.

32 This list is by no means exhaustive. What if market forces beyond any school's control have driven up the salaries of men's football coaches but not women's field hockey coaches? Is it okay to allocate more money to men's sports under those circumstances? Or should men be forced to tighten their belts in other ways to cover the costs of the coach? What if men's sports attract more spectators and hence more revenue?

33 *See* Jessica Gavora, Tilting the Playing Field: Schools, Sports, and Title IX (2002). *The New York Times* has reported on the lengths to which colleges and universities go to pad their women's sports team rosters—evidently in order to avoid having to cut back on men's sports teams. At Quinnipiac University, for example, women members of the cross-country team were *required* to join the indoor and outdoor track teams, so they could be counted three times. At Marshall University the women's tennis coach was evidently reduced to bargaining with students with no particular talent or time for tennis. "They could come to practice when they liked," the coach was said to have told these women, "and would not have to travel with the team." At the University of South Florida, half the women on the cross country team failed to show up for a race. When asked about it, a few laughed and said they had not realized they were on the team. South Florida had evidently been including its track athletes on the cross country rosters. *See* Katie Thomas, *College Teams, Relying on Deception, Undermine Gender Equity*, N.Y. Times, Apr. 25, 2011, *available at* http://www.nytimes.com/2011/04/26/sports/26titleix.html?pagewanted=all.

34 For fuller discussions of the problems with current enforcement of Title IX with regard to athletics, see Jessica Gavora, Tilting the Playing Field: Schools, Sports and Title IX (2002); Alison Somin, The Obama Administration: Changing the Rules of the Title IX Game?, Engage, Dec. 2010, at 26; U.S. Comm'n on Civil Rights, Title IX Athletics: Accommodating Interests and Abilities (2010), *available at* http://www.usccr.gov/pubs/TitleIX-2010-rev100610.pdf.

35 *See, e.g.*, David Moltz, *Looking to Sports to Bolster Enrollments*, Inside Higher Ed, Jan. 21, 2009, *available at* http://www.insidehighered.com/news/2009/01/21/ncaa (describing the success of LaGrange College in rural western Georgia in using football teams to recruit more men).

36 See the comments of Leslie Brucker, an attorney for Public Justice, quoted in Jaschik, *supra* note 16.

Our answers to three of the other critics quoted in the *Inside Higher Ed* article are similar. Donna Lopiano, president of Sports Management Resources, comments that "[c]ivil rights laws are not utilitarian actions meant to recalibrate student or social populations. . . . They are statements of human-rights principles, that if followed, will allow student or social populations to evolve without unfairness. To suggest that civil rights should be suspended for any reasons is an indication of someone missing the point by a long shot." Inasmuch as "suspending civil rights" here should be taken to mean "letting universities decide to fund teams on the basis of the sex of the athletes involved," we are in agreement with Ms. Lopiano that civil rights should not be suspended. Our disagreement with her is over the "substantial proportionality" approach, which in effect compels universities to fund women's teams in circumstances in which they likely would not fund a men's team because they are concerned about being able to show proportional numbers to the Department of Education. Likewise, inasmuch as universities are suspending women's civil rights by denying them admission based on their sex, this is an issue that we are very concerned about. We do not understand why this latter issue seems not to worry Ms. Lopiano.

Fatima Goss Graves of the National Women's Law Center comments that "[t]he hypothesis is flawed from the start. It presumes that only men are interested in sports, and that all men are interested in sports, so right there you are operating on flawed presumptions about the interests and abilities of students." Neither of us have ever assumed or suggested that only men are interested in sports or that all men are interested in sports. Each of us has known women who are excellent athletes and quite interested in sports; we have also each had occasion to interact with plenty of male law professors, a group that is disproportionately uninterested in sports and actively disdainful of those who are. Our more modest claim is that there is some evidence that, at this particular moment in time, college-age men are somewhat more interested in sports than women are. Expanding the athletic opportunities available to men at some campuses might increase at the margin the number of men applying there. We do not suspect or claim that expanding athletic opportunity alone would dramatically raise the numbers of men applying to many colleges and universities. But difficult problems are more often solved by a number of incremental adjustments rather than by a single fix. If there were a single fix that would remedy the problem easily, somebody would have thought of it already. See also *infra* at 17 for a discussion of how boys' lagging performances in K-12 education may also be contributing to gender imbalances in higher education.

Erin Buzuvis, an associate professor at the Western New England College of Law, states that "[t]here's always been resistance to Title IX," and notes that this is a new strategy, but with the same goal (apparently of undermining Title IX). She predicted that it would fail because "in the end most people believe in equal opportunities." We are both believers in equal opportunities, and this is why we are concerned about the perceived lack of equal opportunity for women in admissions at some schools. Our concern about Title IX is that the substantial proportionality approach represents a move away from emphasizing equality of opportunity and instead toward the creation of de facto quotas in athletics. This would be troubling even if it were not leading to actual discrimination against women in admissions. If it is, it should be a high priority to set things right.

37 To make it easier for some colleges to demonstrate that they were adequately meeting female interest in sports under Title IX, the Bush Administration propounded a Model Survey that schools could administer to determine if there was unmet female interest in some sports on their campuses. If a school was fully accommodating female interest in sports teams as shown by the survey results, the Department of Education would presume that the college was in compliance in Title IX. Feminist interest groups opposed the survey, and the Obama Administration rescinded the Bush policy a little more than a year ago.

We find it odd to suggest that a woman who is genuinely interested in playing a varsity sport would be unwilling to fill out a short internet survey conveying this preference to her university's administration. Playing a varsity sport generally requires at least ten to fifteen hours of practice and training time each week, sometimes more in the height of the season. Many college athletes have spent years competing in that same sport at the high school level and sometimes even for years before that. It is rare for a novice athlete to pick up a tennis racket or field hockey stick for the first time at nineteen and compete successfully for her college varsity team. In other words, playing on an intercollegiate athletic team is a huge commitment, especially in contrast to the modest time commitment required to fill out an internet survey. We find it hard to believe that there are many women sufficiently dedicated to a sport to make one commitment but not the other.

Although neither of us has experience recruiting potential college varsity athletes, one of us has experience recruiting college and law students to participate in non-athletic extracurricular endeavors such as newspaper writing and conservative/libertarian political clubs. It is our experience that it is extremely easy to convince students to check a box saying that yes, they are interested in joining the Federalist Society or in writing for the college paper. It is quite another to get many of these same students to commit to recruiting a speaker, writing an article, or even to attending a single lecture. We find it odd that opponents of surveying women about their athletic interests seem not to have noticed this common phenomenon.

38 U.S. Dep't of Educ., Clarification of Intercollegiate Athletics Policy Guidance: The Three Part Test (Jan. 16, 1996), *available at* http://ed.gov/about/offices/list/ocr/docs/clarific.html#two.

39 We hasten to add, however, that flat-out discrimination in allocation of athletic opportunities based on sex is illegal under Title IX and also wrong as a policy matter.

40 The Commission's jurisdiction extends to discrimination or denials of equal protection under the law (in this case on the basis of sex). It does not specifically include an examination of the status of one sex or the other.

41 Mona Charen, *Schwarzenegger Swaggers into the Race*, Townhall.com, Aug. 12, 2003, *available at* http://townhall.com/columnists/monacharen/2003/08/12/schwarzenegger_swaggers_into_the_race/page/full/ ("We [conservatives] were pleased by Proposition 209, which outlawed affirmative action . . . ."); Mona Charen, *The Fog of Affirmative Action*, Townhall.com, Jan. 21, 2003, *available at* http://townhall.com/columnists/monacharen/2003/01/21/the_fog_of_affirmative_action ("Racial preferences perpetuate the very worst stereotype about African-Americans—that they are not as smart as whites and Asians.").

42 Rogers Elliott, A. Christopher Strenta, Russell Adair, Michael Matier & Jannah Scott, *The Role of Ethnicity in Choosing and Leaving Science in Highly Selective Institutions*, 37 Res. Higher Educ. 681 (1996); Richard Sander & Roger Bolus, *Do Credentials Gaps in College Reduce the Number of Minority Science Graduates?* (Working Paper, Draft July 2009); Frederick L. Smyth & John McArdle, *Ethnic and Gender Differences in Science Graduation at Selective Colleges with Implications for Admissions Policy and College Choice*, 45 Res Higher Educ. 353 (2004); U.S. Comm'n on Civil Rights, Encouraging Minority Students to Pursue Science, Technology, Engineering and Math Careers (2010).

43 Although statistics generally indicate that men outnumber women in these fields, it might be that still more men would go into these fields if not for preferences for men in college admissions. Also, if affirmative action for men is a recent phenomenon, it may be that the trend will reverse at some point.

44 In her 2000 book *The War Against Boys*, Christina Hoff Sommers laid much of the blame for boys' lagging scholastic achievement at the feet of feminist interest groups. Christina Hoff Sommers, The War Against Boys: How Misguided Feminism Is Harming Our Young Men (2000). According to Sommers, these groups' success at popularizing flawed studies about how schools shortchanged girls distracted educators from looking at the problems of boys. She describes also how efforts at raising students' feminist consciousness in the K-12 grades may have inadvertently alienated boys. For example, there is a discussion of a guide called "Quit It!" designed to help little boys be less aggressive, in part by coming up with less aggressive variations on the game of tag. *Id.* at 52. Elsewhere, we learn about government-sponsored day care guides that encourage teachers to occasionally reverse the order of Jack and Jill's names in the traditional nursery rhyme, apparently in the name of gender equity. *Id.* at 77. It is unclear how common such extremes are, but it is not impossible what Charen calls "feminized school environment" has an adverse effect on male students.

Richard Whitmire rejects most of Sommers's claims. He argues instead in his provocatively titled *Why Boys Fail* that "[i]f forced to come up with a single sentence summarizing what I learned researching this book, it would be

this: The world has gotten more verbal; boys haven't." Richard Whitmire, Why Boys Fail: Saving Our Sons from an Educational System that Leaving Them Behind 28 (2010). Whitmire claims that teaching reading at ever younger ages and emphasizing whole language instruction over phonetics has disproportionately benefited girls over boys in the early grades, and that many boys never make up these early deficits. His evidence is intriguing, but neither of us is an expert in the teaching of reading and cannot comment fully on these issues here. Rather, we mention this only because the issue seems worthy of more attention and research than it has received.

45 Melana Zyla Vickers, *Where the Boys Aren't*, *Wkly. Standard*, Jan. 2, 2006.

46 Whitmire, *supra* note 44, at 199.

47 *See* Allen, *supra* note 25. Ms. Charen also alludes to this argument by suggesting that women will be "happier" on campuses where there are more men.

48 *See A Professor Proposes To Examine Gender Bias in College Admissions*, Chron. Higher Educ., Oct. 31, 2009, *available at* http://chronicle.com/article/Full-Text-The-Proposal-That/49012/. We note that asking students what they prefer is not always the best way to get at the truth. When asked whether they support racial diversity on campus, most students respond positively (though they do not respond positively to preferential treatment in admissions). On the other hand, Stanley Rothman and his colleagues have found that students express less satisfaction with their university experience as diversity increased (controlling for a large number of other factors, such as overall ranking of the college or university). Stanley Rothman, Seymour Martin Lipset & Neil Nevitte, *Does Enrollment Diversity Improve University Education?*, 15 Int'l J. Pub. Opinion Res. 8 (2003).

Criminal Law & Procedure

A Comprehensive Strategy Targeting Recidivist Criminals with Continuous Real-Time GPS Monitoring: Is Reverse Engineering Crime Control Possible?

*By Peter M. Thomson**

One of the most pressing criminal justice challenges facing the nation is reducing the incidence and prevalence of violent crime, the costs of which are incomprehensibly tragic, destructive, and far-reaching. Over the years, electronic monitoring has been a rising star in the criminal justice system. For nearly three decades, in fact, the technology has been used by correctional departments for supervising criminal offenders in a wide variety of settings.

Electronic monitoring's most traditional application, however, has been to assure that provisionally released offenders comply with judicially-imposed conditions, such as confining a defendant to his residence during a specified period of time. Recently, Global Positioning System ("GPS") technology has been employed to track domestic abusers and sex offenders, primarily to help ensure the safety of the community and/or the safety of former victims through enforcement of "exclusion zones," areas within which the offender's physical presence is prohibited. Even more recently, however, some jurisdictions have instigated pilot programs designed to combat recidivism in sex offenders.

According to the Bureau of Justice Statistics, a significant percentage of the general prison population is recidivist offenders. In fact, recidivists commit the majority of violent crime in the United States. Some studies of metropolitan areas suggest that as much as 70% of reported crimes are committed by felons with prior records.[1] Against this unsettling backdrop, could GPS monitoring be expanded to cover a far broader scope of criminal recidivists to strategically reduce crime in our nation?

This article examines whether it might be possible to craft a comprehensive strategy designed to dramatically reduce crime by using advances in GPS technology to effectively eliminate the recidivist criminal's ability to relapse into prior criminal conduct. Such a long-term strategic approach would implicate a number of constitutional and legal issues. However, if the legal hurdles can be overcome, such an innovative crime-reduction strategy might well be successful, particularly if it could integrate a number of other time-tested crime reduction strategies that criminal justice advocates have successfully employed. These strategies would support long-term, active GPS monitoring, and would include: crime scene correlation, active supervision, and community-oriented behavioral modification techniques such as restorative justice, a powerful program requiring criminals to interact with their victims and immediate social communities.

I. Conventional Methods to Control Crime Have Failed

Anyone who pays even the slightest attention to the evening news is aware that our criminal justice system is broken. The age-old solution of building more prisons and incarcerating more offenders with stiffer penalties has failed to stop a segment of our society from engaging in repetitive criminal behavior. Ironically, it seems that everyone understands the root causes of crime, yet, at the same time, we seem unable to do anything about it. The failure of our public schools to educate and the breakdown of the moral and family structure of society have combined to create cultural breeding grounds for crime. The results have been staggering. Pervasive cultures of crime now exist in many areas of the country which, once limited to the inner cities, have evolved and spread into other communities like a cancer.

Accordingly, it should come as no surprise that the United States has one of the highest per capita incarceration rates in the world.[2] As reported by *The New York Times* in 2008, America has less than 5% of the world's population but almost a quarter of the world's prisoners, which equates to over two million people behind bars, more than any other nation on earth.[3] In fact, the U.S. incarceration rate has almost doubled in each decade since 1970, increasing, for example, from 135 per 100,000 residents in 1978, to 244 in 1988, to 460 in 2003.[4]

Based on these sobering figures, one would think that the crime rate would have fallen proportionally. However, in reality, violent crime in the United States has soared over the long-term. Since 1964, the nation's crime rate has increased by as much as 350%. In 2007 alone, for example, over 11 million crimes were reported.[5] At the end of 2009, approximately 7.2 million people in the United States were on probation, in prison, or on parole.[6] As a consequence of this meteoric rise in criminal offenders, state correctional facilities are bursting at the seams, which has even resulted in the release of prisoners back onto the streets.[7] Recently, the Supreme Court, in *Brown v. Plata*,[8] ordered California to release 46,000 convicted criminals based on the prison system's inability to provide adequate physical health treatment. Today, approximately one in every 32 Americans is subject to the criminal justice system.[9]

To put all this in relevant perspective, a recent Justice

** Peter M. Thomson is a Partner with the Fowler Rodriguez Valdes-Fauli law firm in New Orleans, Louisiana, where he leads the firm's government investigations and maritime criminal defense practices. Prior to joining the firm in April, 2011, he served for twenty-three years as a federal prosecutor with the U.S. Department of Justice. In 2010, while participating in the Justice Department's Leadership Excellence and Achievement Program (LEAP), he worked with a small team to develop a concept he had years earlier for reducing crime through neutralizing recidivists with GPS monitoring technology. His team selected the GPS recidivist concept for their internal presentation (as DOJ employees) to the Justice Department. He would like to thank Tia Farmer, Sophia Kil, Jason Jaye, Lisa Levine, Robert Troxell, and Renee Williams for the contributions they made in helping craft the comprehensive crime control strategy presented in his article.*

Department study conducted over ten years in the most populous counties discovered that the majority of those committing violent felonies had multiple prior arrests and at least one prior felony arrest.[10] More specifically, the report found that an estimated 70% of violent felons in the 75 largest counties previously had been arrested, and 57% had at least one prior arrest for a felony.[11] Similarly, another recent Justice Department report found that nearly half of the inmates in local jails were on probation or parole at the time of their arrest; 40% had served three or more prior sentences of incarceration or probation; and 40% had a current or prior violent felony offense.[12] More than half of these inmates reported having a pre-existing criminal justice status at the time of arrest.[13]

Hence, the data have been mounting for quite some time that recidivist offenders are responsible for the majority of criminal acts in America. In a study by the Justice Department's Bureau of Justice Statistics, which tracked recidivism of released prisoners throughout the United States, approximately seven out of ten released inmates committed at least one serious new crime within the following three years.[14] Within those same three years, 52% of the former inmates were back in prison either because of a new offense or because of a violation of release conditions.[15] Among those with at least three prior arrests, 55% were re-arrested. Among the more serious repeat offenders, i.e., those having at least fifteen prior arrests, a whopping 82% were re-arrested within the same three-year period,[16] and this figure does not even take into account any crimes the former inmates committed for which they were never caught.

Clearly, the nation's criminal justice system was in a state of crisis *before* the economic downturn and skyrocketing unemployment witnessed in 2011. Police departments and prisons have been overwhelmed. In fact, the prisons *themselves* have become learning centers for better crime, not to mention the hardening of criminals that undeniably occurs. A new program, requiring extended GPS monitoring of the very group of criminals statistically responsible for the majority of crime, might offer the long-awaited solution. What if cutting the crime rate in half was merely as simple as wrapping GPS bracelets on the ankles of all the recidivist criminals? Would the device alter their behavior? Would it stop them from harming more victims?

II. The Current State of Electronic Offender Monitoring

Electronic monitoring is employed by courts and corrections departments to track the movement and location of criminal offenders for a wide variety of purposes. In its infancy, electronic monitoring, at least by today's standards, was fairly simple—a monitoring device was connected to a telephone line which alerted officials if the offender left his home. This form of supervising technology was used first in Florida during the mid-1980s as a part of a house arrest program.[17] Specifically, a device worn by the offender emitted a coded signal to a monitoring unit in the residence, which could dial officers over the phone line. In most instances, the device worn by the offender was an ankle bracelet transmitter. First-generation units relied on a radio-frequency ("RF") transmitter that sent the signal to the receiving unit. If an offender moved outside of a permissible range, the authorities would be notified. By 1990, some sort of home confinement with electronic monitoring was in place in all fifty states.[18]

While first-generation devices relied on RF technology, second-generation devices began employing GPS technology. Unlike the earlier systems, GPS devices are capable of tracking *all* of an offender's movements, both in and outside the home. In fact, commercial GPS technology is so accurate that systems can locate an offender anywhere on land or sea within a margin of error of about six feet. As the technology continues to advance, the transmitters undoubtedly will become even smaller and easier to manage and use.

GPS monitoring may be either passive or active. In a passive system, the transmitter communicates with a series of global positioning satellites (hence "GPS") to map out the person's movements. Then, once or twice a day, for example, the information is transferred into a computer where the offender's tracking data is stored for later use.[19] In an active system, however, the offender's location is monitored in real time through a continuous signal, and the transmission rate can be adjusted by seconds or minutes. If an offender were to enter or leave a restricted zone, for example, the monitoring station would be instantly notified of his then-current location.

Today, more than 120 federal, state, county, and local law enforcement organizations have implemented GPS systems (many pilot programs) to track offenders short-term who are on pretrial supervision, serving terms of probation or parole, or as an alternative to jail.[20] The technology is being used to track gang members on supervised release, parolees convicted of certain kinds of burglaries, immigrants who are awaiting hearings,[21] and persons convicted of domestic violence. Accordingly, jurisdictions have found ways to leverage GPS technology to create a wide range of structured offender monitoring programs for various types of needs. As one example, a Massachusetts law grants authority to judges to require domestic abusers to wear GPS transmitters where they have violated restraining orders and been determined to be dangerous after undergoing an assessment.[22] The movements of these offenders are monitored by several centers. If an offender crosses an "exclusion zone" mapped digitally around the victim or her children, the police are instantly notified.[23] In fact, the use of GPS to reign in domestic abusers appears to be catching on. Twelve other states have passed similar legislation, and, as a result, about 5000 domestic abuse offenders are being tracked nationwide.[24]

However, "[t]he greatest use by corrections agencies of the [GPS] technology has been in tracking sex offenders, mostly to keep these offenders away from areas such as schools or near the homes of a previous victims."[25] At least seventeen states have laws enabling some form of electronic tracking for sex offenders on supervised release,[26] and some states impose a lifetime monitoring requirement on certain types of these offenders.[27] Hence, while the majority of other programs generally employ monitoring for limited durations of no more than a few months, far longer terms are being imposed in sex offender cases by legislatures and judges. For example, a Florida statute entitled *Jessica's Act* requires that persons convicted of sexual offenses

against children under the age of twelve be subject to lifetime electronic monitoring, and Pennsylvania and California have followed suit with similar laws.[28]

Reducing Recidivism

Several years ago, a pilot program was initiated in Bergen, Illinois, whereby sex offenders were placed on continuous GPS monitoring. Significantly, only one offender was charged with a new sex crime while under electronic supervision,[29] and less than ten percent were caught committing other non-sex related crimes such as tampering with the equipment or violating the GPS statute. A report issued by the program found that "GPS monitoring appears to encourage these high-risk offenders to control their behavior, and avoid situations that would inspire new crimes."[30]

A shining example of the GPS monitoring platform in the high-risk category can be found in Washington, D.C., with the Court Services and Offenders Supervision Agency ("CSOSA"), which was created by Congress in 1997 to improve public safety in the district through effective community supervision of criminal offenders.[31] Over the past decade, CSOSA has been using GPS devices to monitor largely high-risk offenders on a selective basis. Offenders are assessed with risk-screening tools that effectively predict recidivism and the probability of negative outcomes for those seeking admission into the monitoring program, an alternative to incarceration. For example, as part of the screening process, CSOSA collects information on criminal history, drug use, mental status, and instances of past violence. The stated objectives of the CSOSA program include: (1) ensuring that offenders meet appointments and otherwise comply with program requirements, (2) enforcing domestic violence exclusion areas, (3) restricting the movement of offenders within or without certain geographic locations, and (4) assisting law enforcement officials in solving crimes. With regard to solving crimes, CSOSA cooperates with the police to coordinate data with crime-scene locations for the purpose of developing leads and apprehending those responsible. CSOSA does not monitor offenders for extended periods of time; most are monitored for no more than several months. Even with such short durations, CSOSA has witnessed an increased rate of success in offender reentry with fewer instances of recidivism.

Recently, the use of GPS has been expanded to targeting violent and repeat offenders, largely for the purpose of monitoring them to ensure they follow conditions of release before trial. In 2007, the Charlotte-Mecklenburg Police Department began operating a GPS-based monitoring program targeting certain offenders, requiring them to wear GPS transmitters as a condition of pre-trial release for certain categories of serious criminal offenses.[32] The program also employed a crime-scene correlation feature which allowed officers to quickly identify re-offenders and return them to jail, which reduced the likelihood of others engaging in repetitive criminal behavior. The program's results were stunning: 88% of the offenders in the program did not commit a crime or have their bond revoked while being monitored.[33] Monitoring these serious criminals with GPS locators effectively neutralized nearly nine out of ten of them.

In fact, there is now a consensus among professionals that electronic monitoring is not only effective in reducing the costs of incarceration, but also in reducing recidivism among the offenders released.[34] In the juvenile setting, for example, "one study revealed a 3% recidivism rate for electronically monitored home detention cases."[35] In September 2011, the Justice Department weighed in, concluding that electronic monitoring reduces recidivism.[36]

III. Extended GPS Monitoring of Recidivists Combined with Crime-Scene Correlation and Community-Oriented Behavioral Modification Techniques

None of the GPS monitoring programs in use today appears to have been engineered, from the ground up, for the central purpose of intentionally reducing victim crime as a whole. Why not take a quantum leap forward and launch such a GPS-based model that is strategically and comprehensively designed to neutralize the population of recidivists who are responsible for the majority of crime? If recidivist offenders as an entire class were technologically "blocked" from engaging in criminal behavior, would not crime rates plummet correspondingly? What would such a program look like? What would it require? Would it be feasible? Would it be successful? Would its application be legal? Constitutional?

A threshold question would be how to define the term "recidivist" for purposes of the program. While recognizing that no single definition in the criminal justice system exists, the likely answer would be in the broadest sense possible, consistent with constitutional principles of due process and established definitions adopted by the Department of Justice and state corrections agencies. Of course, the broader the term, the more sweeping the net, and, for the program to reach its greatest potential, that net must capture the largest number of recidivists as possible who are statistically responsible for most crime. Otherwise, the program's effect would have less impact on the incidence and prevalence of crime. Whatever the definition's final scope, inclusion of qualifying offenders would be based on objective criteria following a personalized assessment process comparable to the one currently employed by CSOSA in Washington, D.C.

This personalized assessment process likely would include screening for mental function and anti-social behavior, and it might also examine the individual's needs, strengths, health, and disabilities, as well as an assessment of anticipated risk through other screening tools capable of predicting future criminal behavior. Based on the assessments and the offender's prior criminal record, a determination would be made by professionals administering this new program as to whether to classify the offender as a recidivist or not. If designated as a recidivist, the offender would be required to wear a GPS transmitter, either through judicial order, voluntary agreement, or legislative mandate, when not incarcerated for reasons of (or as a condition of) bond, probation, supervision, or prison overcrowding. The GPS device would monitor the recidivist's precise location in real time, record the data for future use as evidence in court, and would alert the police instantly if an exclusion or "hot" zone (if appropriate) were violated or if the recidivist committed a violation or crime (via the enhanced crime-scene correlation infrastructure).

The program would require an offender classified as a recidivist to be monitored with GPS for an extended period of time, possibly up to three to five years, or longer, depending on his underlying criminal history, the recidivist's personalized assessment, and/or the post-release conditions imposed by a judge. Therefore, the monitoring under this new program would occur during the pre-trial, post-trial, *and* post-sentence release settings, including during terms of probation and supervised release. In fact, the enabling legislation could even limit or entirely remove discretion from the judges so that GPS monitoring would be mandated. The required monitoring in this model would not serve, in most instances, as a substitute for prison. GPS tracking still would be required for recidivists notwithstanding the length of the prison sentence they might actually serve. For example, a repeat felon might serve a sentence of seven years in prison and then be placed on GPS monitoring for several years after his release.

Additionally, state-of-the-art crime-scene correlation technology would be employed, setting new standards for its infrastructure, technologies, and methods by: (1) introducing regional or central monitoring, (2) implementing real time links between GPS monitoring systems and crime-scene time and location data, and (3) infusing the local law enforcement authorities into the GPS monitoring process. Under such a comprehensive GPS/crime-scene correlation model, the local police would work side by side with the GPS monitoring vendors to achieve greater crime-scene correlation impact. Hence, computer cross-referencing would be nearly instantaneous, which would allow law enforcement to quickly identify or eliminate suspects and increase the likelihood of apprehending the offender.

Community-oriented behavioral modification techniques, including the "restorative justice" method, would be integrated into the new GPS recidivist monitoring scheme as well. Thus, the recidivist also would participate in a mandated restorative justice program, which permits the victim and the community to have an active role in addressing the recidivist with the goal of having him or her accept responsibility and see and feel the personal impact that his or her criminal conduct has had on the life of the respective victim. The restorative justice aspect of the new program would focus on the needs of victims and offenders, rather than the need to satisfy abstract principles of law or exact punishment.

Many jurisdictions are facing budget cuts requiring the release of prisoners as prison populations continue to increase, resulting in compromised public safety. Further, corrections agencies are required to supervise more and more offenders with dwindling resources. This new model employing active GPS monitoring provides help by employing cost-effective technologies that will allow supervising officers to monitor more people than they would be capable of monitoring otherwise due to cuts in budgets and resources. For example, the Napa County Board of Corrections recently adopted a GPS program, noting that the program cost only $15 a day per offender compared to the $109 a day cost to keep an offender in jail.[37] Hence, electronic monitoring, when used in lieu of incarceration, is substantially less expensive—about six times less than imprisonment.[38]

Of course, GPS systems do have some noteworthy disadvantages. The devices are not infallible, and they may not always work when the offender is inside a structure or in an area surrounded by tall trees or buildings. The devices are expensive to replace if broken or lost. A percentage of those being monitored also will tamper with or remove their devices, forcing manufacturers to find ways to defeat tampering intended to interfere with the satellite signal. Finally, there are looming legal concerns. Undoubtedly, constitutional challenges will be mounted against such a comprehensive strategy if launched as a pilot program in a city near you.

IV. The Legal Landscape

Electronic monitoring technologies have advanced so rapidly over the decades that the courts have had a difficult time keeping pace. Electronic GPS tracking and its evolving applications, particularly the long-term monitoring of certain kinds of criminals (for example, sex offenders and recidivists) raise a number of constitutional issues and concerns, including issues under the Fourth Amendment, which some have claimed prevents the government from invading an individual's privacy with today's high-tech GPS surveillance. [39] Just recently, in fact, the Supreme Court scheduled arguments in *United States v. Jones*[40] before year's end on the question of whether police need a search warrant to track a suspect with GPS in public spaces over an extended period of time. Although the facts of *Jones* relate to warrantless police surveillance, as opposed to a legislatively-created offender monitoring program, the potential nevertheless exists that the decision will impact statutorily-mandated tracking of offenders with GPS devices, particularly for lengthy periods of time or life terms, as is required for many sex offenders.

Evolution of the Law: The Fourth Amendment and Privacy Rights

Long before states or the federal government ever thought of requiring sex offenders, much less recidivists, to wear GPS-like devices for electronic monitoring purposes, the Supreme Court examined the constitutionality of warrantless tracking of suspects with "beepers," a pre-GPS device using ground-based location-finding technology. The Supreme Court in *United States v. Knotts*[41] held that the warrantless use of beepers to track a suspect over public roads is not a search within the meaning of the Fourth Amendment and, therefore, did not implicate a fundamental right. The Court reasoned that the defendant had no reasonable expectation of privacy on public highways. The Supreme Court in *United States v. Karo*[42] further held, contemporaneously with *Knotts*, that the constitutionality of a warrantless beeper, however, depends on the *kind* of information the device reveals. Surveillance that reveals the beeper's location in a public place does not require a warrant; however, surveillance that reveals the person's location inside private property does.[43]

To reiterate, both *Knotts* and *Karo* focused on the nature of the information collected rather than the kind of technology employed. This approach has proven so far to be adaptable to the rapid evolution in electronic tracking technologies. Based on *Knotts* and *Karo,* whether tracking data is collected by a

beeper or a GPS device should not affect the constitutionality of the surveillance. "[I]t is the exploitation [through the nature of the information collected] of technological advances that implicates the Fourth Amendment, not their mere existence."[44] Notwithstanding this statement, Justice Rehnquist, who wrote the majority opinion in *Knotts*, was insightful enough to leave the door open to the possibility of having to curtail more intrusive surveillance methods in the future as technologies advance. Thus, the Court expressed reservation with using beeper technology to conduct what might amount to "dragnet-type law enforcement."[45]

Since *Karo* and *Knotts*, courts generally have upheld warrantless GPS monitoring of a vehicle moving in public space.[46] However, a question remaining is whether long-term surveillance is permissible, particularly with the capability of GPS systems to continuously map a person's movements. Whether this intrusive ability implicates *Knotts* "dragnet-type" concerns is an exceedingly legitimate question as GPS surveillance is certainly more intrusive than older technologies because it can reveal far more intimate details concerning a person's life: the doctors they visit, how often they gamble, which political rallies they prefer to attend, etc. Moreover, the longer the term of monitoring, the more details and patterns in a person's life begin to emerge.

The constitutionality of long-term GPS tracking is the precise subject matter under review by the Supreme Court in *Jones*, albeit in a warrantless and non-statutorily enabled program context. The case, due to be decided in early 2012, stems from the district court's decision in *United States v. Maynard*,[47] which held that individuals have a reasonable expectation of privacy in the totality of their movements over a lengthy period of time, even if they don't have a reasonable expectation of privacy in a single movement from one point to another.[48] On appeal, Jones argued that the GPS device violated his reasonable expectation of privacy per *Katz v. United States*,[49] and was thus a search within the meaning of the Fourth Amendment. The D.C. Circuit split with other circuits[50] that had approved long-term GPS surveillance per *Knotts*. The D.C. Circuit held that *Knotts* does not govern and that the use of the GPS device constituted "a search because it defeated Jones' reasonable expectation of privacy."[51] In distinguishing *Knotts*, the D.C. Circuit concluded that round-the-clock use of a GPS transmitter without a warrant violated the Fourth Amendment. The Court reasoned that prolonged monitoring was not covered by *Knotts* because of the intrusive nature, specifically the enhanced ability to learn private details about an individual's life.

In *Katz*,[52] the Supreme Court established that an individual has a right against unreasonable search and seizure in areas where he has "exhibited an actual (subjective) expectation of privacy," and that expectation is "one that society is prepared to recognize as reasonable."[53] "The Court also has held that the government violates that right when, without a warrant, it uses various kinds of technology to gain information about acts within such a constitutionally protected space."[54]

Although the Supreme Court's ruling in *Jones* likely will be limited to warrantless tracking in public spaces of police suspects, it might nonetheless impact statutorily-created GPS programs under the Fourth Amendment because of the requirement for offenders to wear the device continuously, including in protected areas. Will the Supreme Court draw a line of distinction between warrantless police tracking and GPS tracking statutes?

In addition to the Fourth Amendment, statutorily-created GPS programs for the purpose of tracking criminally-charged and/or convicted offenders raise a number of other constitutional issues, including the Ex Post Facto Clause together with procedural and substantive due process rights affecting cognizable fundamental rights or liberty interests. The most prevalent GPS statutory regimes, which are those involving sex offenders, have been challenged in the courts based on a variety of grounds, including the Fourth Amendment, double jeopardy, cruel and unusual punishment, invasion of privacy, the Ex Post Facto Clause, and equal protection, as well as substantive and procedural due process.

Procedural Due Process

Arguably, the most likely procedural vehicle for challenging a *technological* restraint, such as continuous GPS monitoring, falls within the constitutional rubric of procedural due process,[55] which must be afforded to persons throughout the criminal process, including during the conviction and sentencing phases.[56] No state shall "deprive any person of life, liberty, or property, without due process of law . . . ."[57] In order to implicate due process, however, the state must act in some manner that deprives a person of a "liberty" or a "property interest."[58] The Supreme Court has acknowledged that the concept of "liberty" is not limited to mere freedom from bodily restraint,[59] although physical restraint (e.g., prison) undeniably is the most elemental of all restraints on liberty interests. However, beyond that, the concept of liberty rights remains somewhat foggy.

In the context of GPS monitoring, courts generally have looked to whether the monitoring serves to deprive an offender of a protected liberty interest, such as the freedom to move.[60] In a recent North Carolina case,[61] a state court of appeal concluded that the state's requirement that offenders enroll in a GPS monitoring program violated due process by unconstitutionally infringing on the offender's protected liberty interest in having freedom to move. The state GPS program required constant, continuous surveillance through a permanently-installed GPS device that tracked offenders in real time.

Notwithstanding the North Carolina decision, GPS programs largely have withstood most procedural due process challenges.[62] The arguments in favor of due process compliance are varied. Obviously, a small GPS device strapped on an ankle presents no significant deprivation of liberty in and of itself and, as discomfiting as one may be in the early stages, over time, the device will feel less intrusive to the person wearing it. A GPS device imposes no physical harm to speak of, and, therefore, is unlikely to be evaluated as punitive. Wearing a GPS device does not impede travel in the slightest degree. In fact, GPS tracking is just a more cost-effective and efficient means of doing something already held to be constitutional—the imposition of lengthy terms of supervision and probation. Accordingly, to the extent that prolonged GPS tracking of recidivist offenders likely does not violate a cognizable "liberty interest" and, thus,

implicate a procedural due process violation, there is even less of a probability of violating a substantive due process interest.

Substantive Due Process

Substantive due process, according to the Supreme Court, is implicated when the government acts in a manner that affects an individual's fundamental right,[63] such as the right to have children, the right to travel, and the right to be free of physical restraint.[64] If a law encroaches on a fundamental liberty interest, the courts apply a "strict scrutiny" analysis that seeks to determine whether the intrusion is "narrowly tailored to serve a compelling state interest."[65] Hence, strict scrutiny looks to whether the legislation (1) serves a compelling state interest, and (2) is narrowly tailored to achieve that interest.[66]

In the event that a fundamental right is not found to be implicated by the new GPS program, the test is then whether the statutory scheme is "rationally related to legitimate government interests."[67] In fact, this lesser standard would probably be applied because GPS tracking of recidivist "street" and violent criminals likely does not implicate a fundamental right. "The rational basis standard is 'highly deferential' and [courts] hold legislative acts unconstitutional under a rational basis standard in only the most exceptional circumstances."[68] GPS tracking of the very subset of criminals responsible for the majority of crime would likely pass the rational-basis test because the monitoring would be rationally related to the legitimate government interest of protecting the public from their continued criminal acts.

However, as stated above, in the event that a fundamental right were determined to be implicated, the constitutionality of the legislation would be evaluated under "strict scrutiny." Under the first prong of the test, it could be argued that protecting the public from high-risk criminal offenders who are likely to re-offend is undeniably a compelling state interest. Protecting the public is one of the state's highest orders of duty. It has long been a compelling interest of the state to protect its citizens from criminal activity. Under the second prong, GPS tracking of recidivist criminals could be narrowly tailored to achieve the state's interest in protecting the public from further violence. Legislation implementing such a GPS program would have to be narrowly tailored to track only high-risk recidivists who have been properly screened based on an objectively-determined likelihood that the subjects will re-offend.

When the state's interest in preventing crime is weighed with the new law's anticipated effectiveness against the resulting intrusion into the recidivist criminal's privacy rights, the GPS program likely will be upheld.[69] In fact, a number of federal circuit courts have ruled that GPS tracking does not violate a defendant's privacy.[70] States clearly have a substantially strong interest in stopping violent crime and ending recidivism, and, based on the prior studies and GPS programs, there is clear evidence that subjecting the recidivist criminal population to long-term GPS monitoring will substantially reduce crime.

Ex Post Facto Clause

A GPS program of this magnitude and scope might also implicate the Ex Post Facto Clause, assuming the mandatory monitoring provision of the statute can be triggered by an offender's prior conduct, such as arrests and convictions, or any other factors pre-dating the effective date of the enabling legislation. Following this line of inquiry, it is first critical to determine whether the legislation is intended to be criminal or civil (i.e., regulatory) in nature. However, even if a state labels a law as being a regulatory one, the Supreme Court can override that label if the monitoring scheme is "so punitive either in purpose or effect as to negate" any intent of creating a civil penalty.[71]

Courts will treat a law as "criminal" if it serves the purpose of either retribution or deterrence, both of which are primary objectives of punishment.[72] Such a law would be ex post facto if it retroactively alters the punishment that the offender received at sentencing.[73] If the legislature's intent was for the GPS bracelet to be a civil or "regulatory" restraint, then courts examine whether the law is so punitive in effect that it is more properly characterized as punishment.[74] In deciding whether a statute is "punitive" in effect and thus violative, a number of factors per *Mendoza-Martinez*[75] are considered, including (1) whether the statute imposes "an affirmative disability or restraint" on the offender, (2) whether the statute seeks to promote retribution or deterrence, and (3) whether "it has historically been regarded as a punishment," together with a number of other relevant factors.[76] The "ultimate question always remains whether the punitive effects of the law are so severe as to constitute the 'clearest proof' that a statute intended by the legislature to be nonpunitive and regulatory should nonetheless be deemed to impose ex post facto punishment."[77]

When analyzing whether an electronic monitoring statute violates the Ex Post Facto Clause, courts often look to the Supreme Court's decision in *Smith v. Doe*,[78] which involved a sex offender registration act that required offenders to register based on offenses committed prior to the law's effective date. Most courts analyzing continuous GPS monitoring of sex offenders under the Ex Post Facto Clause have found no violation,[79] regardless of the fact that the monitoring laws did not exist at the time of the predicate sex crimes. There was no finding that the monitoring law increased the punishment received. However, the issue remains somewhat unsettled. In the recent case *Commonwealth v. Cory*, for example, the Massachusetts Supreme Court distinguished *Smith*, finding that it only applies to registration laws, not to a statute requiring constant electronic GPS monitoring of sex offenders, and found the Massachusetts statute to be punitive in effect.[80] The *Cory* court stressed that continuous GPS surveillance during probation or supervised release imposed more restraints and burden on sex offenders than registration alone.

In 2007, the Sixth Circuit took up the constitutionality of a sex offender monitoring law intended to reduce recidivism in Tennessee. The program required offenders to wear a GPS bracelet twenty-four hours a day and be subjected to continuous surveillance. Similar to other states enacting measures to rein in sex offenders, the law applied even with regard to sentences the offenders received prior to the effective date of the law, opening the door to an ex post facto challenge. However, the Sixth Circuit found no constitutional infirmities and held that the Act did not violate the Ex Post Facto Clause.[81] In reaching this conclusion, the court followed *Smith* and found that the legislature intended the law to be civil, concluding

that monitoring sex offenders with GPS locating devices is not punitive. Applying the *Mendoza-Martinez* factors, the court reasoned that the monitoring restrictions did not impose an affirmative disability and did not result in additional incarceration (and hence was nonpunitive). According to the court, the deferential intent of the Act alone was not enough to overcome the other factors weighing in favor of the state, and the legislature's goals were rationally related to a nonpunitive purpose.[82]

The controversy continues, however, as only this year a New Jersey appellate court ruled that requiring a sex offender, convicted two decades earlier, to wear a GPS device pursuant to a new monitoring law violated the Ex Post Facto Clause. Even though it determined that the legislature's intent was civil, the court concluded that the law was so punitive in effect that it violated the Constitution.[83]

Notwithstanding the New Jersey case, a comprehensive GPS program targeting recidivists likely will survive challenges under the Ex Post Facto Clause. Although permanently wearing a GPS ankle bracelet might be cumbersome and, arguably, an affirmative disability, it also can be considered to be a mere nuisance that is relatively minor and employed not for the purposes of punishment. The GPS device would in no way serve to physically restrain a recidivist in the program, who, except for selected exclusion zones that might be appropriate on a case-by-case basis, would be free to travel anywhere. Other than exclusion zones designated to protect victims, the only physical restraint presented by the GPS device would be that of not committing other crimes. Hence, the monitoring of recidivists would effectively promote deterrence, and any burden on the offender would be far outweighed by society's interests in not only preventing crime, but dramatically reducing the incidence and prevalence of both violent and non-violent victim crimes. Where, as here, there would be a rational connection to a non-punitive purpose, and, in most instances, lifetime monitoring probably would not be required, numerous courts have already upheld similar laws in the context of sex offender statutes.

Endnotes

1 Brian A. Reaves, U.S. Dep't of Justice, Bureau of Justice Statistics, Violent Felons in Large Urban Counties (2006), *available at* http://www.bjs.gov/content/pub/pdf/vfluc.pdf.

2 GPS Tracking Can Reduce Prison Costs, RMT Daily GPS News, http://www.rmtracking.com/blog/2010/10/08/gps-tracking-can-reduce-prison-costs/ (Oct. 8, 2010).

3 Adam Liptak, *Inmate Count in U.S. Dwarfs Other Nations'*, N.Y. Times, Apr. 23, 2008, *available at* http://www.nytimes.com/2008/04/23/us/23prison.html?hp.

4 Sara Steen & Rachel Bandy, *When the Policy Becomes the Problem: Criminal Justice in the New Millennium*, Punishment & Soc'y, Jan. 2007, at 5.

5 *See* Crime in the United States,http://en.wikipedia.org/wiki/Crime_in_the_United_States (last visited Oct. 24, 2011).

6 *See* Lauren E. Glaze, U.S. Dep't of Justice, Bureau of Justice Statistics, Correctional Populations in the United States, 2009 (2010), *available at* http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2316.

7 Adam Liptak, *Justices, 5-4, Tell California to Cut Prisoner Population,* N.Y. Times, May 23, 2011, *available at* http://www.nytimes.com/2011/05/24/us/24scotus.html?pagewanted=all.

8 131 S. Ct. 1910 (2011).

9 Associated Press, *Report: 7 Million Americans in Justice System*, msnbc.com, Nov. 30, 2006, http://www.msnbc.msn.com/id/15960666/ns/us_news-crime_and_courts/t/report-million-americans-justice-system/.

10 *See* Reaves, *supra* note 1.

11 *See id.*

12 Doris J. James, U.S. Dep't of Justice, Bureau of Justice Statistics, Profile of Jail Inmates, 2002 (2004).

13 *Id.*

14 *Id.*

15 *Id.*

16 *Id.*

17 S. Mainprize, *Electronic Monitoring in Corrections: Assessing Cost Effectiveness and the Potential for Widening the Net of Social Control*, 34 Canadian J. Criminology 161 (1992).

18 *See* Kathy S. Padgett et al., *Under Surveillance: An Empirical Test of the Effectiveness and Consequences of Electronic Monitoring,* 5 Criminology & Pub. Pol'y 61 (2006).

19 For example, an offender might wear a GPS device during the day and in the evening connect it to another device to upload his tracking data into a computer system; or, as another example, an offender might be "required to answer a telephone call from a case officer or insert a transmitter into a home device to verify his presence." Stacey L. Sklaver, *The Pros and Cons of Using Electronic Monitoring Programs in Juvenile Cases*, Juv. Just. Committee Newsl., July 2010.

20 Jonathan J. Wroblewski, *RESTART, GPS, Offender Reentry, and a New Paradigm for Determinate Sentencing*, 20 Fed. Sent'g Rep. 314, 315 (2008) (citations omitted).

21 Press Release, Cal. Dep't of Corr. & Rehab., California Department of Corrections and Rehabilitation Announces GPS Partnership with the City of San Bernardino to Monitor High-Risk Gang Activity (Mar. 15, 2006), *available at* http://www.cdcr.ca.gov/News/Press_Release_Archive/2006_Press_Releases/press20060314.html; Associated Press, *Paroled Burglars to Be Fitted with GPS Tracking Devices*, Boston.com, Aug. 12, 2007; *see* Wroblewski, *supra* note 20.

22 Zoila Hinson, *GPS Monitoring and Constitutional Rights*, 43 Harvard C.R.-C.L. L. Rev. 285 (2008).

23 S. 1351, 185th Gen Ct., Reg. Sess., § 7 (Mass. 2007).

24 Ariana Green, *More States Use GPS to Track Abusers,* N.Y. Times, May 8, 2009, *available at* http://www.nytimes.com/2009/05/09/us/09gps.html.

25 *See* Wroblewski, *supra* note 20.

26 Alabama, Arizona, Arkansas, Colorado, Florida, Georgia, Illinois, Indiana, Iowa, Kansas, Michigan, Missouri, Oklahoma, South Carolina, Virginia, Washington, and Wisconsin. *See* Ala. Code § 15-20-26.1 (LexisNexis 2007); Ariz. Rev. Stat. Ann. § 13-902(G) (2007); Ark. Code Ann. § 12-12-923 (2007); Colo. Rev. Stat. § 18-1.3-1005 et seq. (2007); Fl. Stat. Ann. § 947.1405 (West 2008); Ga. Code Ann. § 42-1-14 (2007); 730 Ill. Comp. Stat. 5/5-8A-6 (2007); Ind. Code Ann. § 35-38-2.5-3 (LexisNexis 2007); Iowa Code Ann. § 692A.4A (West 2007); Kan. Stat. Ann. § 74-9101(15) (2006); Mich. Comp. Laws Ann. § 791.285 (West 2007); Mo. Rev. Stat. § 217.735 (West 2007); Okla. Stat. tit. 57, § 510.10 (2008); S.C. Code Ann. § 23-3-540 (2007); VA. Code Ann. § 37.2-908(E) (2007); Wash. Rev. Code Ann. § 9.95.435 (West 2007); Wis. Stat. Ann. § 301.135 (West 2007).

27 *See, e.g.*, Ga. Code Ann. § 42-1-14 (2007); Mich. Comp. Laws Ann. § 791.285 (West 2007); Mo. Ann. Stat. § 217.735 (West 2007).

28 Hinson, *supra* note 22.

29 N.J. State Parole Bd., Report on New Jersey's GPS Monitoring of Sex Offenders (2007), *available at* http://www.state.nj.us/parole/docs/reports/gps.pdf.

30 *Id.*

31 *See* Wroblewski, *supra* note 20.

32 Charlotte-Mecklenburg Police Department Electronic Monitoring Program, http://www.popcenter.org/library/awards/goldstein/2009/09-44.pdf (last visited Oct. 24, 2011).

33 *Id.*

34 *See* Wroblewski, *supra* note 20.

35 OJJDP Model Programs Guide, Home Confinement/Electronic Monitoring, http://www.ojjdp.gov/mpg/progTypesHomeConfinement.aspx (last visited Oct. 24, 2011).

36 U.S. Dep't of Justice, Nat'l Inst. of Justice, Electronic Monitoring Reduces Recidivism (2011), *available at* http://nij.gov/pubs-sum/234460.htm.

37 Katrina Sifford, Changing the Criminal Character: Nanotechnology and Criminal Punishment (June 17, 2009) (unpublished manuscript), *available at* http://www.elmhurst.edu/~phl/pdf/socialscale.pdf.

38 *See* Electronic Monitoring, *supra* note 36, at n.1

39 *See*, e.g., U.S. v. Cuevas-Perez, 2011 U.S. App. Lexis 8675 (7th Cir. Apr. 28, 2011).

40 United States v. Jones, No. 10-1259 (U.S. cert. granted June 27, 2011).

41 460 U.S. 276 (1983).

42 468 U.S. 705 (1984).

43 *Id*.

44 *Id.* at 712.

45 *Knotts*, 460 U.S. at 284.

46 *See, e.g.*, United States v. Sparks, 750 F. Supp. 2d 384 (D. Mass. 2010); United States v. Coombs, 2009 U.S. Dist. Lexis 105547 (D. Ariz. Nov. 12, 2009).

47 615 F.3d 544 (D.C. Cir. 2010).

48 *Id.* (Agents had placed a GPS tracking unit on Antoine Jones' vehicle without a valid warrant for nearly a month. Using the tracking data and other information gleaned from cell phone records and a wiretap, they conducted a search of a house and recovered close to 100 kilograms of cocaine. The district admitted the GPS data excluding only the data when it was located on private residential grounds.).

49 389 U.S. 347 (1967).

50 *See, e.g.*, United States v.Pinedo-Moreno*,* 591 F.3d 121 (9th Cir. 2010).

51 *Id.* at 16 (internal citation omitted).

52 389 U.S. 347 (1967).

53 *Id.* at 361.

54 Hinson, *supra* note 22, at 287 (citing Kyllo v. United States, 533 U.S. 27 (2001)).

55 *See* Erin Murphy, *Paradigms of Restraint*, 57 Duke L.J. 1321, 1351 (2008).

56 *See* Hamdi v. Rumsfeld, 542 U.S. 507, 533 (2004).

57 U.S. Const. amend. XIV, § 1.

58 *See, e.g.,* Bd. Of Regents v. Roth, 408 U.S. 564, 570-71 (1972).

59 Bolling v. Sharpe, 347 U.S. 497, 499 (1954).

60 *See, e.g.,* United States v. Smedley, 611 F. Supp. 2d 971, 975 (E.D. Mo. 2009).

61 State v. Stines, 683 S.E. 2d 411 (N.C. Ct. App. 2009)

62 *See*, *e.g.*, Stop Child Predators, The Constitutionality of GPS Tracking of Sex Offenders (2007), *available at* http://www.stopchildpredators.org/pdf/SCP_PR041707.pdf (cases cited therein). *See generally* Murphy, *supra* note 55, at 1351-1358.

63 *See* Reno v. Flores, 507 U.S. 292, 301-01 (1993).

64 *See* Foucha v. Louisiana, 504 U.S. 71, 80 (1992).

65 *See, e.g.*, *Reno*, 507 U.S. at 302.

66 *See id.*

67 Washington v. Glucksberg, 521 U.S. 702, 728 (1997).

68 Doe v. Moore, 410 F.3d 1337, 1345 (11th Cir. 2005).

69 *See* Brown v Texas*,* 443 U.S. 47 (1979).

70 United States v. Cuevas-Perez, 2011 U.S. App. Lexis 8675 (7th Cir. Apr. 28, 2011); United States v. Pinedo-Moreno, 591 F.3d 121 (9th Cir. 2010); United States v. Marquez, 605 F.3d 604 (8th Cir. 2010).

71 United States v. Ward, 448 U.S. 242, 248-49 (1980).

72 *See* Kansas v. Hendricks, 521 U.S. 346, 361-62 (1997).

73 *See* Lindsey v. Washington, 301 U.S. 397, 401 (1937).

74 *Ward,* 448 U.S. at 248.

75 Kennedy v. Mendoza-Martinez, 372 U.S. 144 (1963).

76 *Id.* at 168.

77 Megan Janicki, *Better Seen than Herded: Residency Restrictions and Global Positioning System Tracking Laws for Sex Offenders*, 16 Pub. Int. L.J. 299 (2007) (citing Doe v. Miller, 405 F.3d 700, 719 (8th Cir. 2005)).

78 Smith v. Doe*,* 538 U.S. 84 (2003).

79 *See* United States v. Morris, No. 08-0142, 2008 U.S. Dist. Lexis 104029, at *2 (W.D. La. Nov. 14, 2008) (dismissing ex post facto challenge against Louisiana's sex offender monitoring law); Uresti v. Collier, No. H-04-3094, 2005 U.S. Dist. LEXIS 34292, at *35 (S.D. Tex. June 23, 2005) (dismissing ex post facto challenge to Texas's sex offender monitoring law); State v. Bare, 677 S.E.2d, 518, 531 (N.C. Ct. App. 2009) (finding North Carolina's satellite-based monitoring constitutional).

80 Commonwealth v. Cory, 911 N.E.2d 187 (Mass. 2009).

81 Doe v. Bredesen, 507 F.3d 998 (6th Cir. 2007).

82 *See id*. at 1005-1006.

83 Riley v. N.J. State Parole Bd., No. A-1004-09T1 (N.J. Super. Ct. App. Div. Sept. 22, 2011).

Honest Services Fraud After *Skilling v. United States*

*By Steven Wisotsky**

The mail fraud statute of 1872 may be regarded as the progenitor of what we now call white collar crimes. Originating with the Postmaster General's concern[1] that the mail system was being used to facilitate fraudulent schemes, the mail fraud statute has evolved into a powerful prosecutorial weapon. The core prohibition in the statute, first amended in 1909,[2] punishes "any scheme or artifice to defraud, or for obtaining money or property by means of false or fraudulent pretenses, representations, or promises."[3] Not only does the statute reach far and wide in its own right, it is also a predicate crime for RICO[4] and money laundering prosecutions.[5]

The classic violation of the statute is a case in which A defrauds B of money or property, as in a Ponzi scheme when it fails. But what if there is no victim in the traditional sense or the losses are abstract? Beginning in the 1940s, the courts developed a theory embracing generalized intangible losses. According to the Supreme Court in *Skilling v. United States*,[6] *Shushan v. United States*[7] originated the doctrine by affirming the mail fraud prosecution of a public official who allegedly accepted bribes from entrepreneurs in exchange for urging city action beneficial to the bribe payers. *Shushan* thus established the theory that the mail fraud statute covers a scheme to defraud the public, although the "loss" sustained by the public is not monetary but rather an intangible right to the honest administration of government. "A scheme to get a public contract on more favorable terms than would likely be got otherwise by bribing a public official would not only be a plan to commit the crime of bribery, but would also be a scheme to defraud the public."[8]

The court in *Shushan* further stated that no trustee has more sacred duties than a public official, and any scheme to obtain an advantage by corrupting such official must in the federal law be considered a scheme to defraud.[9] Nevertheless, subsequent cases extended the principle of fiduciary duty to a private employer-employee relationship. An employee who bought from vendors at reasonable rates but took a kickback from the vendor committed honest services fraud by using the mails in furtherance of the scheme.

Surveying public and private honest services fraud the Supreme Court in *Skilling* stated:

> "Most often these cases . . . involved bribery of public officials," *United States v. Bohonus*, 628 F.2d 1167, 1171 (C.A.9 1980), but courts also recognized private-sector honest-services fraud. In perhaps the earliest application of the theory to private actors, a District Court, reviewing a bribery scheme, explained:
>
> "When one tampers with [the employer-employee] relationship for the purpose of causing the employee to breach his duty [to his employer,] he in effect is defrauding the employer of a lawful right. The actual deception that is practised is in the continued representation of the employee to the employer that he is honest and loyal to the employer's interests." *United States v. Proctor & Gamble Co.*, 47 F. Supp. 676, 678 (Mass. 1942).
>
> Over time, "[a]n increasing number of courts" recognized that "a recreant employee"—public or private—"c[ould] be prosecuted under [the mail-fraud statute] if he breache[d] his allegiance to his employer by accepting bribes or kickbacks in the course of his employment," *United States v. McNeive*, 536 F.2d 1245, 1249 (CA8 1976); by 1982, all Courts of Appeals had embraced the honest-services theory of fraud, Hurson, Limiting the Federal Mail Fraud Statute—A Legislative Approach, 20 Am. Crim. L. Rev. 423, 456 (1983).[10]

In the 1970's prosecutors increasingly began to use the mail and wire fraud statutes in cases brought against political officials for failure to provide honest services.[11] In *United States v. States*,[12] the Eight Circuit affirmed the district court's ruling that the two phrases of section 1341 should be read separately and independently, thus proscribing two distinct offenses: 1) schemes to defraud and 2) schemes for obtaining money or property by means of fraudulent pretenses.[13] Without the requirement of a money or property scheme honest services became the "Stradivarius, the Colt 45, the Louisville Slugger[,] . . . and the true love of federal prosecutors."[14]

McNally

In 1987, the Supreme Court put an end to the expansion of the mail fraud statute in the case of *McNally v. United States*.[15] *McNally* involved the conviction of Gray, a former Kentucky public official, and McNally, a private individual, for participation in a self-dealing patronage and kickback scheme. The violation asserted by the Government was nondisclosure: the failure to disclose the defendants' sharing of insurance commissions, thus depriving the people of Kentucky of their right to have the commonwealth's affairs conducted honestly.[16] Footnote nine of *McNally* shows the broad reach that intangible rights doctrine had attained:

> [I]t was not charged that requiring the Wombell agency to share commissions violated state law. We should assume that it did not. For the same reason we should assume that it was not illegal under state law for Hunt and Gray to own one of the agencies sharing in the commissions and hence to profit from the arrangement, whether or not they disclosed it to others in the state government. It is worth observing as well that it was not alleged that the mail fraud statute would have been violated, had Hunt and Gray reported to state officials the fact of their

---

** Steven Wisotsky is a tenured professor of law at Nova Southeastern University Law Center, where he teaches white collar crimes and appellate practice. He is co-chair of the ABA Criminal Justice Section Appellate and Habeas Committee and author of Professional Judgment on Appeal (2009).* *He thanks Dane Leitner, NSU Law Center 2011, for his valuable assistance in the preparation of this article.*

financial gain. *The violation asserted is the failure to disclose their financial interest, even if state law did not require it, to other persons in the state government whose actions could have been affected by the disclosure.* It was in this way that the indictment charged that the people of Kentucky had been deprived of their right to have the Commonwealth's affairs conducted honestly.

> It may well be that congress could criminalize using the mails to further a state officer's efforts to profit from governmental decisions he is empowered to make or over which he has some supervisory authority, even if there is no state law proscribing his profiteering or even if state law expressly authorized it. But if state law expressly permitted or did not forbid a state officer such as Gray to have an ownership interest in an insurance agency handling the State's insurance, it would take a much clearer indication than the mail fraud statute evidences to convince us that having and concealing such an interest defrauds the State and is forbidden under federal law.[17]

The Court in *McNally* held that the scheme did not qualify as mail fraud. "Rather than construing the statute in a manner that leaves its outer boundaries ambiguous and involves the Federal Government in setting standards of disclosure and good government for local and state officials, we read the statute as limited in scope to the protection of property rights."[18] The Court relied in part upon the rule of lenity in reaching its conclusion, and stated that "if Congress desires to go further, it must speak more clearly than it has."[19]

Congressional Reaction to *McNally*

It took Congress only a year to "speak more clearly." In 1988, Congress enacted 18 U.S.C. §1346 as part of the Drug Abuse Act. Some have called section 1346 the "stealth bill" because it was never referred to any committee of either the House or the Senate, and was never the subject of any floor debate.[20] The statute itself is only twenty-eight words long, has virtually no legislative history,[21] and fails to define the term honest services. Section 1346 reads: "For the purpose of this chapter, the term 'scheme or artifice to defraud' includes a scheme or artifice to deprive another of the intangible right of honest services." It is clear that Congress disapproved of *McNally* because of either its result or its rationale.

Skilling

It was not until *Skilling* in 2010 that the Supreme Court attempted to resolve the issues surrounding the honest services doctrine. Petitioner Jeffrey K. Skilling was the former chief executive officer of Enron, then an energy and commodities company, from February until August 2001. Less than four months later, Enron, the seventh-highest revenue-grossing company in America at the time, crashed into bankruptcy, and its stock plummeted in value.

Skilling was charged with engaging in a scheme to deceive investors about Enron's true finances by manipulating its publicly reported financial results and making false and misleading statements about Enron's financial performance. Count One of the indictment charged Skilling with conspiracy to commit honest services wire fraud by depriving Enron and its shareholders of the intangible right of his honest services.[22]

After a four-month trial, the jury found Skilling guilty of nineteen counts, including the honest services fraud conspiracy charge, and not guilty of nine insider trading counts. Skilling appealed, claiming his conviction was premised on an improper theory of honest services wire fraud.[23] Skilling maintained that section 1346 was unconstitutionally vague, and alternatively, that his conduct did not fall within the statute. The Fifth Circuit affirmed Skilling's convictions, and the Supreme Court granted certiorari.

The Court acknowledged that Skilling's vagueness challenge had force because honest services decisions preceding *McNally* were not models of clarity or consistency, but held that there is a definable core of honest services cases. Rather than invalidating section 1346, the Court in *Skilling* determined that the statute should be construed and pared down to the "core" of pre-*McNally* case law concerning honest services. The Court explained that its longstanding practice is to consider a limiting construction before striking a federal statute as impermissibly vague.[24] "Although some applications of the pre-*McNally* honest services doctrine occasioned disagreement among the Courts of Appeals, these cases do not cloud the doctrine's solid core: The vast majority of the honest services cases involved offenders who, in violation of a fiduciary duty, participated in bribery or kickback schemes."[25]

In view of the history of honest services cases, the Court determined that Congress intended section 1346 to reach "*at least*" bribes and kickbacks. The Court stated that the *McNally* case itself, which spurred Congress to enact section 1346, presented a paradigmatic kickback fact pattern. The Court acknowledged that reading the statute to proscribe a wider range of offensive conduct would raise the due process concerns underlying the vagueness doctrine.

The Government urged the Court in *Skilling* to include conflict-of interest cases, i.e., undisclosed self-dealing by a public official or private employee within the scope of section 1346.[26] The Government asserted that while pre-*McNally* cases involving undisclosed self-dealing were not as numerous as the bribery and kickback cases, they were abundant. The government argued that the theory of liability in *McNally* itself was nondisclosure, which is an accurate statement of the case in the lower courts. Although a kickback scheme was in fact involved in *McNally,* "the violation asserted is the failure to disclose their financial interest . . . ."[27] The Court in *Skilling* found that "[g]iven the relative infrequency of [conflict of interest] prosecutions [in comparison to bribery and kickback charges] and the intercircuit inconsistencies they produced, . . . that a reasonable limiting construction of §1346 must exclude this amorphous category of cases."[28] Echoing *McNally*, the Court in *Skilling* stated that "if Congress desires to go further, it must speak more clearly than it has." So far, Congress has not done so.

Regarding the due process issue, the Court concluded that vagueness is not a problem when section 1346 is limited to bribery and kickback schemes because there is fair notice of what the statute prohibits. "'[W]hatever the school of thought

concerning the scope and meaning of' §1346, it has always been 'as plain as pikestaff that' bribes and kickbacks constitute honest-services fraud."[29] The Court did not perceive a significant risk of arbitrary prosecutions because section 1346 draws content on bribes and kickbacks from pre-*McNally* case law and from federal statues proscribing and defining similar crimes.[30] In short, a criminal defendant who participates in a bribery or kickback scheme cannot plausibly complain about prosecution on vagueness grounds.[31]

As to Skilling's conduct, the Court noted that the Government did not, at any time, allege that Skilling solicited, accepted, or offered payments to or from a third party in exchange for making these misrepresentations. Therefore, he did not commit honest services fraud.[32] But because the indictment alleged three objects of the conspiracy—honest services wire fraud, money or property fraud, and securities fraud—the Court remanded to determine if there was error and if it required reversal of the conspiracy conviction. On remand, the Fifth Circuit held that the error was harmless beyond a reasonable doubt and affirmed the conviction on all counts.[33]

Justices Scalia and Thomas concurred with the majority's judgment but not its honest services rationale. Scalia argued that the doctrine of honest services includes more than bribes and kickbacks. "Among all the pre-*McNally* smorgasbord offerings of varieties of honest services fraud, not one is limited to bribery and kickbacks. That is a dish the Court has cooked up all on its own."[34] Additionally, Scalia argued, there remains uncertainty as to when the fiduciary obligation arises and if it comes from state or federal law. The Court in *Skilling* briefly addressed this concern in a footnote stating that debates were rare regarding the source and scope of fiduciary duties in bribe and kickback cases. "The existence of a fiduciary relationship, under any definition of that term, was usually beyond dispute . . . ." Examples include public official-public, employee-employer, and union official-union members.[35] The Court then cited *Chiarella v. United States*[36] for the "established doctrine that [a fiduciary] duty arises from a specific relationship between two parties."[37]

In sum, a charge under section 1346 now requires that the defendant have a fiduciary duty, and the scheme to defraud of honest services is limited to bribery and kickback schemes.

Analysis of Post-*Skilling* Cases

A. Fiduciary Duty

The First Circuit analyzed fiduciary duty in an honest services charge after *Skilling* in *United States v. Urciuoli*.[38] Urciuoli, the chief executive officer of a medical center in Rhode Island, employed John Celona, then a Rhode Island state senator, to market a nursing home Urciuoli's medical center owned. But, in substance, the Senator was employed to use his office on behalf of the medical center to support or oppose bills and influence major insurance companies.[39] Urciuoli was convicted of multiple honest services fraud counts, including undisclosed conflict of interest and bribery. He appealed, and a new trial was ordered. On retrial, the government excluded any mention of conflict of interest, and he was again convicted.[40] Uriciuoli appealed again and, relying on language from *Skilling*, argued that as a private citizen he owed no fiduciary duty to the public and therefore the honest services charge could not stand.

The court held that nothing in *Skilling's* language or context suggests that the Court was distinguishing between the fiduciary who received the bribe and the non-fiduciary who gave it.[41] Indeed, the court stated that of the nine circuit court cases that *Skilling* cited as examples of core honest service fraud cases, two involved convictions of individuals who bribed another to violate his fiduciary duties.[42] The First Circuit affirmed, and held that this case is the core bribery offense preserved by *Skilling*.[43]

B. The Agreement Required for a Bribe or Kickback Scheme

In *United States v. Siegelman*,[44] then-Alabama Governor Don Siegelman was convicted of federal funds bribery, honest services fraud, conspiracy, and obstruction of justice. Siegelman's honest services mail fraud convictions were based on allegations that Richard Scrushy, founder and former chief executive officer of HealthSouth Corporation, made and executed a corrupt agreement with Siegelman. Scrushy allegedly gave Siegelman $500,000 in exchange for Siegelman appointing him to Alabama's Certificate of Need Review Board (the "CON Board"), and Scrushy used the CON Board seat to further HealthSouth's interests.[45]

Siegelman argued that counts six and seven of the honest services charges were not proper because there was no express quid pro quo for Siegelman's appointment of Scrushy to the CON Board. The district court stated that an express agreement or words of promise are not needed for honest services fraud; "the government need only show that a public official has obtained a payment to which he was not entitled, knowing that the payment was made in return for official acts."[46] Nor is there any requirement that an agreement be memorialized in writing. "Since the agreement is for some specific action or inaction, the agreement must be explicit, but there is no requirement that it be express."[47] The court found that testimony at trial was sufficient to prove that Siegelman and Scrushy had agreed to a deal in which Scrushy's donation would be rewarded with a seat on the CON Board.[48] This corrupt quid pro quo thereby proved counts six and seven. Siegelman petitioned for a writ of certiorari, and the Supreme Court vacated judgment and remanded the case for further consideration in light of *Skilling*.[49]

On January 19, 2011, the Eleventh Circuit heard oral arguments concerning Siegelman's case. Siegelman's attorney argued that the government did not prove an explicit quid pro quo bribe or kickback scheme to warrant a conviction on honest services fraud. The government took the position that it had proven Siegelman and Scrushy did have a quid pro quo arrangement and that the *Skilling* case had no bearing on their convictions, as the honest services charge involved bribery. The Eleventh Circuit has not ruled as of the date of this publication.

Unresolved Issues

The cycle of the Court imposing limits on the mail fraud statute and inviting Congress to speak "more clearly"

may continue. On September 28, 2010, the Honest Services Restoration Act was introduced in the Senate by Senator Patrick J. Leahy.[50] The bill was sent to committee, but it was not passed before the session ended, thereby killing the bill. The bill included both undisclosed self-dealing for public officials and undisclosed private self-dealing for officers and directors.[51] The term "public official" was defined, but "officers" and "directors" were not. The bill also required a mens rea requirement; the individual must "knowingly falsify, conceal, cover up, or fail to disclose material information that is required to be disclosed regarding the financial interest in question by any Federal, State, or local statute, rule, regulation or charter applicable" to the individual.[52] It is unknown whether the bill will be re-introduced.

The *Skilling* decision did not address important questions regarding honest services and has also created some new ones. For example, it stated that the Court's definition of honest services fraud only reaches serious culpable conduct without defining the term "serious" or explaining whether "minor" frauds can be prosecuted under the honest services statute.[53] Other lacunae relate to the existence vel non of a fiduciary duty, public or private.

A recent article by Elizabeth R. Sheyn addresses these issues and provides recommendations for a new honest services statute.[54] Her major recommendations are to define or describe when a fiduciary duty arises, to add a mens rea requirement that a defendant act with specific intent to defraud, and to require proof of actual economic harm.[55]

Conclusion

To conclude, the mail fraud statute has undergone dramatic expansion over the years. The *Skilling* decision is the most recent, holding that honest services charges under section 1346 are limited to bribery or kickback schemes. The Court in *Skilling* also required that the scheme involve, as the defendant or an unindicted coconspirator, an individual with a fiduciary duty, even though it declined to define how or when the fiduciary duty arises. Post-*Skilling*, conflict of interest charges or undisclosed self dealing charges under section 1346 are not allowed. Perhaps in the near future Congress may "speak more clearly" in response to *Skilling* as it did after *McNally*.

Endnotes

1 Carrie A. Tendler, *An Indictment of Bright Line Tests for Honest Services Mail Fraud*, 72 Fordham L. Rev. 2729, 2731 (2004).

2 The Supreme Court first addressed the scope of the mail fraud statute in Durland v. United States, 161 U.S. 306, 314 (1896). The Court in *Durland* held that the mail fraud statute reached more broadly than the then-existing version of the common law crime of false pretenses. The statute "includes everything designed to defraud by representations as to the past or present, or suggestions and promises as to the future." *Id*. at 313. In 1909, Congress codified *Durland*.

3 18 U.S.C. §1341.

4 18 U.S.C. § 1961.

5 18 U.S.C. § 1956.

6 130 S. Ct. 2896 (2010).

7 117 F.2d 110, 113 (5th Cir. 1941).

8 *Id*. at 115.

9 *Id*.

10 *Skilling*, 130 S. Ct. at 2926-2927 (2010).

11 Jeffrey J. Dean & Doye E. Green, Jr., McNally v. United States *and Its Effect on the Federal Mail Fraud Statute: Will White Collar Criminals Get a Break?*, 39 Mercer L. Rev. 697, 703 (1988).

12 488 F.2d 761, 764 (1973).

13 *Id*.

14 Hon. Jed S. Rakoff, *The Federal Mail Fraud Statute* (Part 1), 18 Duq. L. Rev. 771, 771 (1980).

15 483 U.S. 350 (1987).

16 *Id*. at 361.

17 *Id*. (emphasis added).

18 *Id*. at 360.

19 *Id*.

20 The specific text of what has become 18 U.S.C. §1346 was inserted in the Omnibus Drug Bill for the first time on the very day that the Omnibus Drug Bill was finally passed by both the House and the Senate. The text of what is now §1346 was never included in any bill as filed in either the House of Representatives or the Senate. As a result, the text of §1346 was never referred to any committee of either the House or the Senate, was never the subject of any committee report from either the House or the Senate, and was never the subject of any floor debate reported in the Congressional Record. United States v. Brumley, 116 F.3d 728, 742 (5th Cir. 1997) (Jolly, J., dissenting).

21 A report submitted by the Senate Judiciary Committee after 18 U.S.C §1346 was enacted stated that the provision was designed to "overturn the decision in *McNally*" and "reinstate all of the pre-*McNally* case law pertaining to the mail and wire fraud statutes without change." Elizabeth R. Sheyn, *Criminalizing the Denial of Honest Services After Skilling*, 2011 Wisc. L. Rev. 27, 35 (2011) (citing 134 Cong. Rec. S17,360, S17,376 (daily ed. Nov. 10, 1988)).

22 Skilling was also charged with over twenty-five substantive counts of securities fraud, wire fraud, making false representations to Enron's auditors, and insider trading.

23 On appeal Skilling also contended that pretrial publicity and community prejudice prevented him from obtaining a fair trial. Both the Fifth Circuit and the Supreme Court concluded that Skilling's fair trial argument failed, and that he did not establish that a presumption of juror prejudice arose or that actual bias infected the jury that tried him. Skilling v. United States, 130 S. Ct. 2896, 2907 (2010).

24 *Id*. at 2930.

25 *Id*.

26 *Id*. at 2932.

27 McNally v. United States, 483 U.S. 350, 361 n.9 (1987).

28 *Skilling*, 130 S. Ct. at 2932.

29 *Id*. at 2933. The Court quoted its decision in *Broadrick v. Oklahoma*, 413 U.S. 601 (1973): "'Even if the outermost boundaries of a statute are imprecise, any such uncertainty has little relevance . . . where appellants' conduct falls squarely within the "hard core" of the statute's proscriptions.' *Skilling*, 130 S. Ct. at 2933 (quoting *Broadrick*, 413 U.S. at 608). And it stated, "Today's decision clarifies that no other misconduct falls within §1346's province." *Skilling*, 130 S. Ct. at 2933.

30 *Skilling* cites as examples 18 U.S.C. §§ 201(b), 666(a)(2), and 41 U.S.C. § 52(2). *Skilling*, 130 S. Ct. at 2933. A recent district court case, decided March 11, 2011, found that *Skilling* did not, as the defendant suggested, require simultaneous proof of § 201(b)(1) and (b)(2) in order to successfully prove honest services fraud. United States v. Ring, 2011 WL 833335 (D.D.C. 2011).

31 *Skilling*, 130 S. Ct. at 2934.

32 *Id.* at 2934.

33 United States v. Skilling, 2011 WL 1290805 (5th Cir. Apr. 6, 2011).

34 *Skilling*, 130 S. Ct. at 2939.

35 *Id.* at 2930 n.41.

36 445 U.S. 222 (1980).

37 *Skilling*, 130 S. Ct. at 2931 n.41 (quoting *Chiarella*, 445 U.S. at 233).

38 613 F.3d 11 (1st Cir. 2010).

39 *Id*. at 13.

40 The bribe to the Senator was not to press or oppose legislation directly through his votes; rather, the purchased use of official power was the implied threat of such action—and also the potential use of influence over legislation in committee—that the Senator conveyed largely by implication through the orchestrated meetings. But the distance between this and paying outright for legislative votes is not great: both involve the misuse of office. *Id.* at 17.

41 *Id*. at 18.

42 *Id*.

43 *Id*.

44 561 F.3d 1215 (11th Cir. 2009).

45 United States v. Siegelman, 561 F.3d 1215, 1219 (11th Cir. 2009).

46 *Id*. at 1225-1227, citing bribery decisions *Evans v. United States,* 112 S. Ct. 1881(1992) and *McCormick v. United States,* 111 S. Ct. 1807 (1991).

47 *Siegelman*, 561 F.3d at 1226.

48 Bailey, an associate of Siegelman, provided testimony that Siegelman showed him a check, said that it was from Scrushy, and that Scrushy was "halfway there." Bailey asked, "[W]hat in the world is he going to want for that?" Siegelman replied, "[T]he CON Board." Bailey then asked, "I wouldn't think that would be a problem, would it?" Siegelman responded, "I wouldn't think so." *Id.* at 1228.

49 Siegelman v. United States, 130 S. Ct. 3542 (2010).

50 Honest Services Restoration Act, S. 3854, 111th Cong. (2010).

51 *Id*.

52 *Id*.

53 *Id*. at 47.

54 Elizabeth R. Sheyn, *Criminalizing the Denial of Honest Services After* Skilling, 2011 Wisconsin L. Rev. 27, 35 (2011).

55 *Id*. at 55-58.

Environmental Law & Property Rights

EPA Regulation of Fuel Economy: Congressional Intent or Climate Coup?[1]

*By Marlo Lewis**

Note from the Editor:

This paper assesses EPA's rule setting standards for motor vehicle greenhouse gas emissions. As always, The Federalist Society takes no position on particular legal or public policy initiatives. Any expressions of opinion are those of the author. The Federalist Society seeks to foster further discussion and debate about the EPA's new regulation. To this end, we offer links below to different sides of this issue and invite responses from our audience. To join the debate, you can e-mail us at info@fed-soc.org.

Related Links:

- EPA, National Highway Traffic Safety Administration, Final Tailpipe Rule: http://www.globalwarming.org/wp-content/uploads/2011/08/Final-Tailpipe-Rule.pdf
- EPA, National Highway Traffic Safety Administration & California Air Resources Board, Interim Joint Technical Assessment Report: Light Duty Vehicle Greenhouse Gas Emission Standards and Fuel Economy Standards for Model Years 2017-2025: http://www.epa.gov/oms/climate/regulations/ldv-ghg-tar.pdf
- Letter from Rep. Darrell E. Issa to EPA Administrator Lisa Jackson (Sept. 30, 2011): http://www.washingtonpost.com/r/2010-2019/WashingtonPost/2011/09/30/Health-Environment-Science/Graphics/oversight930.pdf
- Letter from EPA Administrator Lisa Jackson to Sen. Jay Rockefeller (Feb. 22, 2010): http://epa.gov/oar/pdfs/lpj_letter.pdf
- Letter from Dave McCurdy, President and CEO, Alliance of Automobile Manufacturers, to Sens. Nancy Pelosi, Harry Reid, John Boehner, and Mitch McConnell (Mar. 17, 2010): http://www.openmarket.org/wp-content/uploads/2010/06/auto-alliance-letter-to-house-and-senate-leaders-march-17-2010.pdf
- Sen. Lisa Murkowski, opening statement, debate on S.J. Res 26: http://www.gpo.gov/fdsys/pkg/CREC-2010-06-10/pdf/CREC-2010-06-10-pt1-PgS4789.pdf#page=1

In May 2010, the Environmental Protection Agency (EPA) issued a rule setting standards for motor vehicle greenhouse gas emissions. By creating these standards, EPA is implicitly regulating fuel economy. Because the rule also obligates EPA to regulate greenhouse gases from stationary sources, the agency is now determining national policy on climate change. EPA has asserted that it is simply implementing the Clean Air Act. But the Clean Air Act was neither designed nor intended to regulate greenhouse gases, and it provides no authority to regulate fuel economy.

Last year, Congress declined to give EPA explicit authority to regulate greenhouse gases when Senate leaders abandoned cap-and-trade legislation. A key selling point for the Waxman-Markey cap-and-trade bill was that it would exempt greenhouse gases from regulation under several Clean Air Act programs.[2] If instead of introducing a cap-and-trade bill, Reps. Waxman and Markey had introduced legislation authorizing EPA to do exactly what it is doing now—regulate greenhouse gases through the Clean Air Act as it sees fit—the bill would have been rejected. The notion that Congress gave EPA such authority in 1970, almost two decades before global warming emerged as a public concern, and five years before Congress enacted the first fuel economy statute, defies common sense.

* *Marlo Lewis, Ph.D., is a senior fellow at the Competitive Enterprise Institute's Center for Energy and Environment.*

I. EPA Is Regulating Fuel Economy

Motor vehicle greenhouse gas emission standards implicitly regulate fuel economy. EPA and the National Highway Traffic Safety Administration (NHTSA) confirm this—albeit not in so many words—in their joint May 2010 greenhouse gas/fuel economy Tailpipe Rule. As the agencies acknowledge, no commercially-proven technologies exist to filter out or capture carbon dioxide (CO_2) emissions from fossil fuel-powered vehicles. Consequently, the only way to decrease grams of CO_2 per mile is to decrease fuel consumption per mile, i.e., increase fuel economy.

Although the Tailpipe Rule also targets other greenhouse gas emissions from new motor vehicles, such as hydrofluorocarbons (HFCs) from vehicle air conditioning systems, CO_2 constitutes 94.9% of vehicular greenhouse gas emissions, and "there is a single pool of technologies . . . that reduce fuel consumption *and thereby* reduce CO_2 emissions as well."[3]

That EPA is regulating fuel economy is also evident from EPA, NHTSA, and the California Air Resources Board's (CARB's) *Interim Joint Technical Assessment Report*, the framework document for the Administration's current plan to increase average fuel economy to 54.5 miles per gallon by 2025. The document proposed a range of fuel economy targets from 47 mpg to 62 mpg. The mpg targets are simple reciprocals of four CO_2 reduction scenarios: "Four scenarios of future stringency are analyzed for model years 2020 and 2025, starting

with a 250 grams/mile estimated fleet-wide level in MY 2016 and lowering CO_2 scenario targets at the rate of 3% per year, 4% per year, 5% per year, and 6% per year."[4]

The 54.5 mpg target represents a negotiated compromise between the 4% per year (51 mpg) and 5% per year (56 mpg) CO_2 reduction scenarios.[5]

II. Clean Air Act Does Not Provide the Authority to Regulate Fuel Economy

Does section 202 of the Clean Air Act, the provision through which EPA is promulgating motor vehicle greenhouse gas emission standards, say anything about fuel economy? It did not in 1970, but as amended in 1977, it does.

Section 202(b)(4)(C) authorizes EPA to grant an automaker a four-year waiver from nitrogen oxides (NO_X) emission control standards if the waiver is necessary to develop innovative power train or emission control systems that have "a potential for long-term air quality benefit or the potential to meet or exceed the average fuel economy standard applicable under the Energy Policy Conservation Act after the waiver expires." No waiver may apply to more than 5% of a manufacturer's production or more than 50,000 vehicles, or engines, whichever is greater.

So when Congress amended the Clean Air Act in 1977, it spoke directly to the issue of fuel economy in section 202, and what it granted EPA was a limited authority to grant temporary waivers from NO_X emission standards. Congress did not, in addition, authorize EPA to develop or adopt fuel economy standards.

Congress, through separate statutes—the 1975 Energy Policy Conservation Act (EPCA) and 2007 Energy Independence and Security Act (EISA)—gave NHTSA sole responsibility to prescribe fuel economy standards.[6] The Secretary of Transportation is to consult with the EPA Administrator before prescribing fuel economy standards,[7] and EPA is to calculate the fuel economy of vehicles and test automakers' compliance with fuel economy standards.[8] But prescribing fuel economy standards is NHTSA's responsibility, not EPA's.

III. The Administration's Greenhouse Protection Strategy

Because EPA regulation of fuel economy exceeds the statutory scheme Congress created, EPA's actions are vulnerable to both legal challenge and legislative repeal. But that is the case only if the auto industry has the will to fight. Therefore, obtaining industry buy-in has become a key objective of the Obama Administration. Using CARB as the heavy, EPA endangered the distressed auto industry's very survival. Then EPA offered to protect auto makers from this threat if—but only if—they pledged not to challenge EPA and CARB's new greenhouse gas/fuel economy regulations.

Although the details of the negotiations may never be known, the basic terms and outcome are clear enough. In February 2009, EPA Administrator Lisa Jackson commenced a rulemaking[9] to reconsider Bush EPA Administrator Stephen Johnson's denial[10] of California's request for a waiver to establish greenhouse gas motor vehicle emission standards. Because the waiver would also allow other states to adopt the California program, because states would be implicitly regulating fuel economy, and because automakers would have to reshuffle the mix of vehicles delivered for sale in each "California" state to achieve the same average fuel economy, Jackson's proceeding could have balkanized the U.S. auto market.

The National Automobile Dealers Association clearly explained this possibility in a January 2009 report titled *Patchwork Proven*.[11] Consumer preferences differ from state to state, so the same automaker typically sells a different mix of vehicles in each state. Only by sheer improbable accident would the average fuel economy (or grams CO_2-equivalent/mile) of an automaker's vehicles delivered for sale in one state be the same as that in another state. But if EPA granted the California waiver, each automaker would have to achieve the same average fuel economy (grams CO_2-equivalent/mile) in every state opting into the California program. If all fifty states were to adopt the program, each automaker would have to manage fifty separate fleets, reshuffling the mix in each state regardless of consumer preference. This would produce a chaotic situation.

This possibility gave EPA and CARB the advantage in closed-door negotiations with the auto industry over EPA's greenhouse gas/fuel economy regulations. As part of the "Historic Agreement"[12] brokered by Obama Environment Czar Carol Browner, California and other states agreed to consider compliance with EPA's greenhouse gas emission standards as compliance with their own.[13] But in return, auto manufacturers and their trade associations had to support both the Tailpipe Rule and the California waiver.[14] In a September 30, 2011 letter to EPA administrator Lisa Jackson, House Oversight and Government Reform Committee Chairman Darrell Issa summarized the terms for auto makers under the "Historic Agreement":

> In addition to the extreme secrecy, participating automobile manufacturers, as well as their representative trade associations, waived their legal rights to:
>
> 1. Pursue litigation challenging California's regulation of GHG emissions, including litigation concerning preemption under the Energy Policy and Conservation Act (EPCA);
>
> 2. Contest any final decision by EPA granting California's waiver request; and
>
> 3. Contest any final fuel economy regulations issued by either EPA or NHTSA.[15]

IV. The Disappearing, Reappearing Patchwork

In January 2010, Alaska Senator Lisa Murkowski sponsored a Congressional Review Act resolution of disapproval (S. J. Res. 26)[16] to nullify the legal force and effect of EPA's Endangerment Rule.[17] The Endangerment Rule is the trigger for the Tailpipe Rule and the prerequisite for all other EPA greenhouse gas regulations. Sen. Murkowski is neither a climate skeptic nor an opponent of greenhouse gas regulation per se. But in her view, "politically accountable members of the House and Senate, not unelected bureaucrats, must develop our nation's energy and climate policies."[18]

In a February 2010 letter to West Virginia Sen. Jay Rockefeller, EPA Administrator Lisa Jackson warned that enactment of S. J. Res. 26, by overturning the Endangerment Rule on which the Tailpipe Rule depends, would "undo" the

"Historic Agreement," leaving California and other states free to create a regulatory patchwork inimical to the health of the U.S. auto industry.[19] This was also the key talking point of the Alliance of Automobile Manufacturers, the main industry group opposing S. J. Res. 26.[20]

Neither Jackson nor the Alliance mentioned that the patchwork possibility exists because she granted the waiver[21] in the first place. Had Jackson reaffirmed Johnson's denial, there would have been no such possibility; hence no reason for the auto industry to defend EPA's new regulations.

The peril of a "regulatory patchwork" was one of EPA Administrator Johnson's reasons, in December 2007, for denying California's request for a waiver.[22] Proponents dismissed this concern out of hand.[23] They argued as follows: EPA has been granting California waivers for the past forty years. Waivers have never created a state-by-state regulatory "patchwork." In each case, the waiver created two easily-managed standards, federal and California. A waiver for greenhouse gas emission standards would also create just two standards, not a "patchwork."

Johnson's detractors overlook the fact that, unlike all previous California emission standards, greenhouse gas emission standards implicitly and unavoidably regulate fuel economy. Only after EPA finalized the Endangerment Rule did environmental advocates recognize the patchwork issue created by the California waiver. None mentioned they had changed their tune; none acknowledged that Administrator Johnson had been correct.

V. Case for Denying the Waiver

Administrator Jackson approved the California waiver in late June 2009.[24] Johnson's decision denying the waiver sets out a number of persuasive reasons why she should not have done so.

Johnson's argument[25] may be summarized as follows. Section 209(b)(1)(B) of the Clean Air Act says that "[n]o such waiver shall be granted if the Administrator finds that . . . such state does not need such standards to meet extraordinary and compelling conditions." California's "compelling and extraordinary conditions" are the state's geography, meteorology, and large number of vehicles, which cause severe "local and regional air pollution." Those California-specific conditions have no "close causal ties" to the "global air pollution" linked to climate change:

1. Greenhouse gas concentrations are essentially uniform throughout the globe, and are not affected by California's geography and meteorology.

2. California's vehicles emit greenhouse gases, but so do mobile and stationary sources throughout the world. The resulting "global pool" of greenhouse gas emissions is not any more concentrated in California than anywhere else.

3. Even if one assumes that "extraordinary and compelling" refers not to the "global air pollution" itself but its potential impacts, such as heat waves, drought, and sea-level rise, California's vulnerability is not "sufficiently different" from the rest of the nation to justify waiving federal preemption of state motor vehicle emission standards.

As my colleague CEI General Counsel Sam Kazman quipped approvingly, "They call it global warming, not California warming."

One might restate Johnson's argument as follows. Given California's unusual geography, meteorology, and number of vehicles, the state cannot clean up its air, or attain federal air quality standards, unless it obtains waivers to adopt tougher-than-federal motor vehicle emission standards. This rationale for granting waivers does not apply to greenhouse gases, which are not *air quality* contaminants, and for which federal *air quality standards* do not exist.

In addition, EPA could not grant the waiver without authorizing California to do that which Congress has prohibited—regulate fuel economy. EPCA states:

> When an average fuel economy standard prescribed under this chapter is in effect, a State or a political subdivision of a State may not adopt or enforce a law or regulation related to fuel economy standards or average fuel economy standards for automobiles covered by an average fuel economy standard under this chapter.[26]

This is a very strong statement of preemption. States are prohibited from adopting laws or regulations "related to" fuel economy standards. This broad language bars the adoption of fuel economy standards packaged as something else or commingled with other measures. A balkanized auto market—the threat created by the California waiver—is what the EPCA preemption was designed to prevent.

VI. California Program Is Related to Fuel Economy Standards

That the California greenhouse gas motor vehicle emissions law, AB 1493, is highly "related to" fuel economy is obvious from CARB's 2004 Staff Report presenting the agency's "initial statement of reasons" for its regulatory proposal.[27] Nearly all of the Staff Report's recommended options for reducing greenhouse gas emissions were previously recommended as fuel-saving options in the National Research Council (NRC)'s 2002 fuel economy report.[28]

CARB proposes a few additional options not included in the NRC study, but each is a fuel-saving technology, not an emission-control technology.

The text of AB 1493 also implies that CARB is to regulate fuel economy. AB 1493 requires CARB to achieve "maximum feasible" greenhouse gas reductions that are also "cost-effective," defined as "[e]conomical to an owner or operator of a vehicle, taking into account the full life-cycle costs of the vehicle."[29] CARB interprets this to mean that the reduction in "operating expenses" over the average life of the vehicle must exceed the "expected increases in vehicle cost [purchase price] resulting from the technology improvements needed to meet the standards in the proposed regulation."[30] Virtually all of the "operating expenses" to be reduced are expenditures for fuel. The CARB program cannot be "cost-effective" unless CARB regulates fuel economy.

In a letter earlier this year to House Energy and Power Subcommittee Chairman Ed Whitfield, CARB Executive Officer James Goldstene explains why he believes EPCA does

not preempt California's greenhouse gas motor vehicle emission standards:

> CARB has never claimed that there is no relation between the pollution [CO_2] emitted by burning fossil fuels and the rate at which they are burned [gallons of fuel consumed per distance traveled, i.e. fuel economy]. CARB merely maintains the fact that pollution control and fuel economy are not identical—fuel economy and pollution control regulations have different policy objectives, utilize different incentive and flexibility features, and there are technologies that reduce pollution that are not counted under fuel economy measures, and some fuel economy improvements do not reduce emissions commensurately.[31]

There are several problems with this argument.

1. A greenhouse gas emission standard does not have to be "identical" to a fuel economy standard to be "related to" it. EPCA preempts state laws or regulations "related to" fuel economy.

2. CARB does not maintain that fuel economy and greenhouse gas standards "have different policy objectives." CARB's selling point (set out elsewhere in Goldstene's letter) is that combining EPA's greenhouse gas standards with NHTSA's corporate average fuel economy (CAFE) standards yields 33% more fuel savings.

3. The fact that EPA's greenhouse gas standards utilize "different incentives and flexibility features" is irrelevant. Neither greenhouse gas regulation nor fuel economy regulation is defined by those features and incentives. The CAFE program, for example, would still be a fuel economy program if it did not allow for payments of fines in lieu of compliance or award credits for flex-fuel vehicle sales.

4. Although some technologies—e.g., improved sealants for automobile air conditioning systems—"are not counted under fuel economy measures," such technologies address only 5.1% of motor vehicle greenhouse gas emissions.[32] The remaining 94.9% can only be addressed by fuel-saving technologies. For that share, fuel economy improvements do reduce greenhouse gas emissions "commensurately."

Being highly "related to" fuel economy, California's AB 1493 program violates EPCA's express prohibition.

VII. CARB: Fuel Economy Retro

Although not an issue Johnson considered when denying the California waiver, it is worth noting that the fuel economy program implicitly established by AB 1493 conflicted with fuel economy reforms Congress had enacted in the 2007 Energy Independence and Security Act (EISA). EISA replaced the "flat" standards of the original CAFE program, which applied to an automaker's entire fleet, with standards based on fuel efficiency-related vehicle "attributes." The "attribute-based" standards NHTSA developed vary according to a vehicle's "footprint"—the area formed by the wheel base multiplied by the track width. The flat, fleet-wide approach encouraged automakers to increase production and sale of smaller vehicles to offset the sale of larger, more profitable vehicles rather than improve fuel economy across all vehicle types. Congress switched to the attribute-based approach in hopes of encouraging compliance via technological innovation.[33]

Although California's greenhouse gas emission standards are calibrated in CO_2-equivalent grams per mile rather than miles per gallon, they are flat, not attribute-based. As in the pre-EISA federal program, there is one average standard for all light vehicles and one for all heavier vehicles. As CARB noted last year:

> The AB 1493 regulations set separate greenhouse gas emission standards for both passenger cars and light-duty trucks (PC/LTD1) and heavier light-duty trucks and medium-duty passenger vehicles (LDT2/MDPV). . . . Compliance is determined on a fleet-wide basis, meaning that while each individual model can be above or below the standard, the average of a manufacturer's fleet must meet the standard or else the manufacturer incurs debits that must be equalized within five years.[34]

Between the time California applied for a waiver and Johnson's denial in March 2008, AB 1493 had become a fuel economy anachronism, mandating a regulatory structure Congress had discarded. The Historic Agreement obscures the basic incompatibility between AB 1493 and EISA by aligning CARB's standards with NHTSA's.

VIII. The Process Behind the Agreement

The process by which the "Historic Agreement" was negotiated raises additional legal issues. The Presidential Records Act states:

> Through the implementation of records management controls and other necessary actions, the President shall take all such steps as may be necessary to assure that the activities, deliberations, decisions, and policies that reflect the performance of his constitutional, statutory, or other official or ceremonial duties are adequately documented and that such records are maintained as Presidential records pursuant to the requirements of this section and other provisions of law.[35]

Rather than documenting the negotiations producing the "Historic Agreement," White House Environment Czar Carol Browner required participants to observe a "vow of silence" and forbade them to take notes. "We put nothing in writing, ever," CARB Chairman Mary Nichols told *The New York Times*.[36]

In his September 30, 2011 letter to Administrator Jackson,[37] Chairman Issa notes three circumstances suggesting that the Obama Administration tied its offer of bailout money to automakers' participation in the agreement:

1. The Administration reached multi-billion dollar agreements to bail out GM and Chrysler three weeks after the "Historic Agreement" was struck.

2. Former EPA Associate Administrator Lisa Heinzerling served on "the Presidential Task Force charged with bailout negotiations and was also a primary negotiator of the 'Historic Agreement.'"

3. One domestic manufacturer received over $200 million in federal support for the development of electric vehicles—"two loans being authorized in the weeks leading up to the agreement, and one authorized on May 20, 2009, the day after the 'Historic Agreement' was announced. . . ."

A deal combining bailout money with protection from the patchwork possibility EPA created could be characterized as an offer the auto industry could not refuse.

IX. More on The Process

The more recent negotiations culminating in the EPA/NHTSA/CARB greenhouse gas/fuel economy standards for model years 2017-2025 also appear to be problematic.

Citing Jeremy Anwyl,[38] CEO of Edmunds.com, and Jack Nerad[39] of Kelley Blue Book, in an August 11, 2011 letter[40] to White House Counsel Kathryn Ruemmler, Chairman Issa contends that although the Administration conferred with environmentalists, automakers, and union labor, there was no one at the table representing "the very consumers who will be asked to buy a new generation" of higher-priced vehicles. The 54.5 mpg standard was the product of an "off-the-record political negotiation." From this point on, the rulemaking process will be a "mere formality"—a criticism also voiced by Amy Sinden of the pro-regulatory Center for Progressive Reform.[41]

The Administrative Procedure Act "does provide agencies with the option of conducting a negotiated rulemaking," notes Issa. However, "such a process is subject to additional transparency requirements, such as those required under FACA [Federal Advisory Committee Act]." FACA requires the head of the lead agency to (i) make an official determination that a negotiated rulemaking committee serves the public interest;[42] (ii) publish in the *Federal Register* a notice that lists the persons proposed to represent the affected interests, describes the agenda of the negotiation, and solicits public comment;[43] and (iii) keep minutes and records.[44] EPA and NHTSA, the lead federal agencies in the negotiation, did not take those steps.

X. Outside the Scope of Law?

Issa also contends that the Obama Administration's recent fuel economy deal is "outside the scope of law." EPA and NHTSA plan to establish fuel economy standards for model years 2017-2025—a nine-year period. But EPCA limits the setting of fuel economy standards to "not more than 5 model years."[45]

EPA and NHTSA address this issue in their November 2011 joint proposed rulemaking. Due to EPCA's five-year limitation, NHTSA's CAFE standards for MYs 2022-2025 are "conditional." In contrast, "EPA's standards for those model years will be legally binding when adopted in this round." NHTSA's MY 2022-2025 standards "will be determined with finality in a subsequent, de novo, notice and comment rulemaking" based on a "mid-term evaluation" to be completed no later than April 1, 2018. To maintain the "benefits" of "harmonization," NHTSA is proposing standards for all nine model years, "but the last 4 years of standards will not be legally binding as part of this rulemaking."[46]

Thus, to get around EPCA's five-year limit, NHTSA proposes only to *propose* but not finalize fuel economy standards for MYs 2022-2025. Yet automakers had better plan to comply with those standards anyway, because EPA's standards for MYs 2022-2025 are legally binding, the two sets of standards are "harmonized," and NHTSA will finalize its standards (or something similar) after a "mid-term evaluation." NHTSA's "conditional," "non-binding" MY 2022-2025 standards are not voluntary.

The agencies' joint proposed rule does not explain the legal basis for this plan. Nowhere does EPCA authorize NHTSA to propose "conditional" fuel economy standards, much less "conditional" standards that exceed the five-year limitation.

This nine-year plan also conflicts with another EPCA provision. EPCA obligates the Secretary of Transportation to consider "economic practicability" when setting fuel economy standards.[47] But, observes Issa, "At this time it is impossible for NHTSA to adequately consider economic practicality for fuel standards in MYs 2022-25, primarily because car manufacturers themselves do not have product plans for that year, and market conditions are unknown 14 years into the future."[48]

XI. Harmonized and Consistent?

In *Massachusetts v. EPA*, the Court rejected the argument that EPA "cannot regulate carbon dioxide emissions from motor vehicles because doing so would require it to tighten mileage standards, a job (according to EPA) that Congress has assigned to DOT [Department of Transportation]." The Court did not explain why it rejected that argument. It simply asserted: "The two obligations may overlap, but there is no reason to think the two agencies cannot both administer their obligations and yet avoid inconsistency."[49]

Recent history suggests the two agencies cannot avoid inconsistency. NHTSA's approval of a nine-year fuel economy standards program conflicts with EPCA's five-year limitation. NHTSA and EPA's off-the-record stakeholder negotiations conflict with FACA and the Presidential Records Act. NHTSA's support for the California waiver conflicts with EPCA's prohibition of state laws and regulations "related to" fuel economy.

Echoing the Court, the agencies claim that EPA and CARB's greenhouse gas standards are "harmonized and consistent" with NHTSA's fuel economy standards. Yet the same officials contend that if Congress were to overturn EPA's greenhouse gas component of the Tailpipe Rule, Americans would consume 25% more oil (an additional 19.1 billion gallons) over the lifetime of the same vehicles. How can that be?

CARB Executive Director David Goldstene addresses the issue in his aforementioned letter to Chairman Whitfield:

> That the National Program [NHTSA + EPA] achieves greater emissions reductions and fuel savings than the CAFE standards alone is a result of the different underlying statutory authority that results in different program components. The four key differences are: 1) unlike the Energy Policy Conservation Act (EPCA), the CAA [Clean Air Act] allows for the crediting of direct

emission reductions and indirect fuel economy benefits from improved air conditioners, allowing for greater compliance flexibility and lower costs; 2) EPCA allows Flexible Fuel Vehicle (FFV) credits through model year 2019, whereas the EPA standard requires demonstration of actual use of a low carbon fuel after model year 2015; 3) EPCA allows for the payment of fines in lieu of compliance but the CAA does not; and 4) treatment of intra firm trading of compliance credits between cars and light trucks categories.[50]

Difference 1) doesn't get us near 19.1 billion gallons in additional fuel savings. According to the Tailpipe Rule, CO_2 emissions due to air conditioner-related loads on automobile engines account for 3.9% of total passenger car greenhouse gas emissions, and various technologies could reduce air conditioner-related CO_2 emissions by 10% to 30%.[51] A 30% reduction of the 3.9% of motor vehicle emissions associated with air conditioner engine load would decrease fuel consumption by only 1.1%.

Differences 2) and 3) are likely the big factors. Per difference 2), automakers cannot comply with EPA's greenhouse gas standards by manufacturing flexible-fueled vehicles. And per difference 3), automakers cannot pay fines in lieu of compliance with EPA's greenhouse gas standards.

Because of differences 2) and 3), EPA will be able to mandate additional fuel savings beyond those required by the statutory scheme Congress created.

The National Program is "harmonized and consistent" only in the sense that EPA and CARB's standards trump NHTSA's standards when the two conflict. Yet, to repeat, Congress authorized NHTSA, not EPA, to prescribe fuel economy standards, and prohibited state agencies like CARB from doing so.

In a July 11, 2011 letter to Chairman Whitfield responding to questions from Energy and Commerce Committee members,[52] EPA Associate Administrator David McIntosh also vouched for the harmony and consistency of the National Program.

In his question to EPA, Rep. John Shimkus pointed out that EISA extended the CAFE credit granted to manufacturers of FFVs, phasing it out in 2020, whereas EPA's greenhouse gas regulations allow credits "only during the period from model years 2012 to 2015." After that, "EPA will only allow FFV credits based on a manufacturer's demonstration that the alternative fuel is actually being used in the vehicles." Shimkus asked: "How can this rule be characterized as 'harmonized and consistent' if the way EPA treats FFV [credits] is markedly different than the way Congress mandated FFV credits be treated under CAFE?" McIntosh replied:

> EPA treats FFVs for model years 2012-2016 the same as under EPCA [as amended by EISA]. Starting with model year 2016, EPA believes the appropriate approach is to ensure that FFV emissions are based on demonstrated emissions performance, which will correlate to actual usage of alternative fuels. This approach was supported by several public comments.

Thus, according to McIntosh, starting in 2016, EPA will not give an automaker a CAFE credit for building FFV vehicles unless the automaker demonstrates that its customers actually use alternative fuels—a requirement inconsistent with EISA. Several people submitting comments on EPA's greenhouse gas standards supported this approach. And that is the only justification needed to override the policy set forth in law.

In sum:

- In 2016-2019, NHTSA gives credits for building FFVs.
- In 2016-2019, EPA does not give credits for building FFVs.
- The two policies are harmonized and consistent.

McIntosh did not reply to another question from Shimkus: "Could the logical reason for Congress's silence on FFVs in section 202(a) be that Congress never envisioned the Clean Air Act would be used to regulate fuel economy?"

XII. Is California the Tail that Wags the Dog?

The "National Program" transfers power from NHTSA to EPA and CARB in a more fundamental way. EPA and CARB can compel NHTSA to "harmonize" its regulations with theirs just by proposing new, more stringent greenhouse gas emission standards. Since EPA attributes endangerment to the "elevated concentrations" of atmospheric greenhouse gases,[53] since even full implementation of the non-ratified Copenhagen climate treaty would only slow the growth of atmospheric concentrations,[54] and since even a "low probability" risk of a "high impact" event qualifies as endangerment,[55] EPA and CARB will always have reason to tighten emission standards.

Even so, the process moved faster than most outsiders expected. On May 21, 2010, President Obama issued a memorandum directing EPA and NHTSA to develop greenhouse gas/fuel economy standards for MYs 2017-2025,[56] fourteen days after publication of the agencies' Tailpipe Rule prescribing greenhouse gas/fuel economy standards for MYs 2012-2016.

Under EISA, NHTSA is not required to prescribe MY 2017 fuel economy standards until April 2015.[57] Yet the Administration initially planned to finalize fuel economy standards for MY 2017 and later by July 2012, "nearly three years before they are due."[58] What is the reason for such a speedy turnaround?

In a January 11, 2011 letter to Chairman Issa, the Alliance of Automobile Manufacturers reported that "CARB intends to pursue the development of its own separate rules for MY 2017-2025 light duty GHG emission regulations early this year—*more than a year ahead of the federal rule* [emphasis in original]." The Alliance letter complained that California's "rushed effort toward a state rulemaking is not in the spirit of a collaborative effort to develop a single national program for fuel economy/GHG standards."[59] By rushing, California recreated the possibility of a fuel-economy patchwork, necessitating a new round of stakeholder negotiations and a new "Historic Agreement."[60]

Two differences between the July 2011 "Historic Agreement" and the May 2009 "Historic Agreement" are worth

noting. First, whereas the May 2009 agreement set fuel economy standards only moderately more aggressive than those proposed in NHTSA's 2008 rulemaking to implement EISA,[61] the July 2011 agreement proposes fuel economy standards that are far more aggressive. It is doubtful that Congress would approve a 54.5 mpg standard if it were proposed in legislation and put to a vote. Second, the July 2011 agreement commits EPA to grant a waiver for California's MY 2017-2025 greenhouse gas emission standards *before* California requests it or finalizes the standards to which it would apply.[62]

Note that Obama's May 21, 2010 memorandum directs NHTSA and EPA to "produce joint federal standards that are harmonized with applicable State [i.e. California] standards." EPA and NHTSA's standards are to harmonize with CARB's standards, not the other way around.

The term "National Program" is misleading. Our current fuel economy regime is the California Program, not the statutory scheme Congress created through either EPCA or the Clean Air Act.

XIII. The Greenhouse Briar Patch

In addition to regulating fuel economy, EPA is applying Clean Air Act permitting requirements to large stationary sources of greenhouse gases: power plants, refineries, steel mills, pulp and paper factories, and cement production facilities.[63] EPA will soon establish greenhouse gas New Source Performance Standards (NSPS) for coal-fired power plants and petroleum refineries.[64] If these go unchallenged, it is likely that EPA will develop greenhouse gas performance standards for numerous other industrial source categories. We can also expect EPA to set quasi-fuel economy standards for aircraft, marine vessels, and non-road engines and vehicles,[65] even though no existing statute authorizes any agency to prescribe such standards. In short, EPA is legislating climate change policy.

EPA claims that its climate policy regulations follow inexorably, like a row of falling dominoes, from *Mass. v. EPA*. According to EPA, the Court left the agency no choice but to make an endangerment finding, the Endangerment Rule compelled EPA to establish motor vehicle greenhouse gas emission standards, the Tailpipe Rule automatically made greenhouse gases from "major" stationary sources "subject to regulation" under the Prevention of Significant Deterioration (PSD) pre-construction and Title V operating permits programs, and litigation pursuant to the Endangerment Rule compelled EPA to establish NSPS for coal-fired power plants and petroleum refineries.

EPA's reading of *Mass. v. EPA* will be tested in litigation before the D.C. Circuit Court of Appeals in *Coalition for Responsible Regulation, Inc. v. EPA*. Petitioners seek to overturn EPA's Endangerment, Tailpipe, Triggering, and Tailoring Rules.[66] Whatever the ruling in that case, Congress would still be free to overturn the agency's greenhouse gas regulations for either statutory or policy reasons.

A question seldom explored, however, is why, in *Mass. v. EPA*, counsel for EPA did not argue then, as EPA argues now, that regulating greenhouse gases via the Clean Air Act leads to "absurd results."

EPA's July 2008 Advance Notice of Proposed Rulemaking,[67] June 2010 greenhouse gas Tailoring Rule,[68] and September 2011 brief in *Coalition for Responsible Regulation v. EPA*[69] develop the argument that applying PSD and Title V permitting requirements to greenhouse gases produces regulations that conflict with and undermine congressional intent.

Whereas only large industrial facilities emit enough smog- and soot-forming air pollutants (100/250 tons per year) to meet the PSD/Title V major source applicability thresholds, millions of non-industrial facilities—big box stores, office buildings, churches, hospitals, schools, Dunkin Donut shops—emit enough carbon dioxide (CO_2) to meet the thresholds. Permitting agencies could not keep up with the volume of permit applications, and the ever-growing backlog would cripple both environmental enforcement and economic development. Annual PSD permit applications would jump from 280 to more than 81,000 per year, a 300-fold increase. Sources requiring operating permits would increase from 14,700 to 6.1 million, a 400-fold increase. A 40-fold increase in permit applications would extend processing time from 6-10 months to 10 years—greatly exceeding the maximum of 18 months allowed by the statute.[70] To avoid permit gridlock, EPA and its state counterparts would have to hire an estimated 230,000 additional staff at any annual cost to taxpayers of $21 billion.[71]

This assessment raises several questions. Why didn't counsel for EPA explain to the Supreme Court that an endangerment finding would lead, via a tailpipe rule, to absurd results? Why didn't EPA's counsel argue that the chain of causality from endangerment finding to absurd results is evidence Congress did not design or intend for the Clean Air Act to be a framework for greenhouse gas regulation?

To suggest that EPA had no grasp of the regulatory ramifications of an endangerment finding until after the Court decided *Mass. v. EPA* is not credible. It is tantamount to saying that the expert in the Clean Air Act did not understand how the statute works.

In June 1998, technology analyst Mark P. Mills published a report warning that a CO_2 endangerment finding could compel EPA to regulate over 1 million small- to mid-sized businesses.[72] The study was a response to EPA General Counsel Jonathan Z. Canon's April 1998 memorandum, which argued that several Clean Air Act regulatory provisions are "potentially applicable" to greenhouse gases.[73] Petitioners in *Mass. v. EPA* cited the Cannon memorandum in their initial petition for a rulemaking to establish greenhouse gas emission standards for new motor vehicles.[74] The Mills study was published by the Greening Earth Society, a project of the Western Fuels Association, one of EPA's stakeholders. The agency could not have been unaware of it.

In its brief in *Coalition for Responsible Regulation v. EPA*, EPA states that for more than thirty years, it has consistently taken the position that PSD applies to *any* regulated air pollutant. In the agency's words: "EPA expressly confirmed the applicability of PSD to *any* pollutant regulated under the Act, including specifically all non-NAAQS pollutants, in regulations issued in 1978, 1980, and 2002."[75] Greenhouse gases would become "regulated air pollutants" the moment any EPA regulation controlling greenhouse gas emissions from motor vehicles took effect.

In *Mass. v. EPA*, the Court based its decision partly on the view that an endangerment finding would not lead to "extreme measures," such as an outright ban on motor vehicle greenhouse gas emissions.[76] However, requiring tens of thousands of small sources to obtain PSD permits and 6.1 million to obtain Title V permits annually would be an extreme case. The Court might not have been so quick to dismiss the risk of "extreme measures" had it understood how a section 202 endangerment finding would affect EPA's obligations under other provisions of the Act.

EPA's counsel similarly made no attempt to challenge petitioners' argument that the case dealt solely with EPA's authority to regulate new motor vehicles under Title II of the Clean Air Act, which, they asserted, is "separate" from Title I, and "entirely separate" from the agency's Title I authority to promulgate national ambient air quality standards (NAAQS).[77] But the PSD program is a Title I program. Moreover, once EPA made an endangerment finding under section 202 and began regulating stationary sources under PSD, it was predictable that EPA would sooner rather than later develop greenhouse gas performance standards for industrial source categories under section 111, also a Title I authority.[78] Title I and Title II may be "separate" but they are not "entirely separate"; they are linked.

Moreover, EPA's Title II endangerment finding arguably creates a compelling precedent for NAAQS regulation of greenhouse gases.

As noted above, EPA attributes endangerment of public health and welfare to the "elevated concentrations" of atmospheric greenhouse gases. Having made this determination under section 202, EPA could not without self-contradiction fail to make the same finding in a section 108 endangerment proceeding. Section 108 requires EPA to initiate a NAAQS rulemaking for "air pollution" from "numerous or diverse mobile or stationary sources" if such pollution "may reasonably be determined to endanger public health or welfare." Carbon dioxide obviously comes from numerous *and* diverse mobile *and* stationary sources, and EPA has already determined that the associated "air pollution" — the "elevated concentration"— endangers public health and welfare. Logically, EPA must now establish NAAQS for greenhouse gases set below current atmospheric levels.

Environmental groups have picked up on this logic. In December 2009, the Center for Biological Diversity and 350. Org petitioned EPA to initiate rulemakings, under section 108, to establish NAAQS for CO_2 at 350 parts per million (about 40 parts per million below current concentrations) and for other greenhouse gases at preindustrial levels.[79] There is a large potential for "extreme measures." The Clean Air Act requires states to come into attainment with primary (health-based) NAAQS within five years, or no later than ten years if EPA determines additional time is required.[80] Yet not even a worldwide depression permanently reducing global economic output and emissions to, say, 1970 levels would stop greenhouse gas concentrations from rising.[81]

For perspective, the Waxman-Markey bill aimed to help achieve the Copenhagen climate treaty goal of stabilizing atmospheric CO_2-equivalent greenhouse gas concentrations at 450 parts per million by 2050.[82] A NAAQS requiring states to make a proportionate contribution[83] to CO_2 stabilization at 350 parts per million and other greenhouse gases at pre-industrial levels in five to ten years would cause the United States to become a single non-attainment area, and the Clean Air Act would function as a no-growth mandate, contradicting a core purpose of the Act: protecting the "productive capacity" of the population.[84]

It is tempting to dismiss NAAQS regulation of greenhouse gases as a conceit of leftwing extremism or rightwing paranoia. But EPA's July 2008 Advance Notice of Proposed Rulemaking contains a lengthy discussion of the subject,[85] and the Department of Justice's (DOJ) August 2010 brief before the Supreme Court in *American Electric Power v. Connecticut* treats NAAQS as a viable mechanism for regulating greenhouse gases.

DOJ's brief argues that EPA's current and future Clean Air Act regulations preempt federal common law litigation to control greenhouse gas emissions. The brief mentions section 202 (motor vehicle standards), section 165 (PSD permitting), section 111 (new source performance standards), and Title V (operating permits) as applicable Clean Air Act authorities. Then there is the following: "Section 108 of the CAA also provides EPA with a mechanism for listing pollutants that 'endanger public health or welfare' and meet certain other criteria. When an air pollutant is listed, the Act requires States to regulate emissions to prevent pollution from exceeding EPA standards."[86] EPA's Tailoring Rule, which seeks to avoid permit gridlock by exempting small greenhouse gas emitters from PSD and Title V, would not mitigate the economic fallout from NAAQS regulation of greenhouse gases.

The Tailoring Rule has legal problems of its own. Over the next five years, the Tailoring Rule replaces the statute's numerical definitions of "major emitting facility"—a potential to emit 250/100 tons per year—with new thresholds—a potential to emit 100,000/75,000 tons—that are orders of magnitude larger. "Tailoring" is the same as *amending* in this scenario. Administrative agencies have no authority to amend statutes.

Here, then, is the argument EPA's counsel did not make in *Mass. v. EPA*:

- EPA cannot regulate greenhouse gases from new motor vehicles under section 202 without regulating greenhouse gases *throughout the economy* under the Act *as a whole*, including PSD, Title V, NSPS, and, logically, NAAQS.
- There is no evidence, textual, historical, or otherwise, that when Congress enacted the Clean Air Act in 1970, or amended the Act in 1977 and 1990, it intended for EPA to implement an economy-wide climate change mitigation program.
- Indeed, regulating greenhouse gases through the Clean Air Act leads to "absurd results"—extreme measures that conflict with and undermine congressional intent.
- "Tailoring" (amending) the Act to avoid crashing the PSD and Title V programs would simply substitute one absurd result for another, because administrative agencies have no power to amend statutes.

• Absurd results are additional evidence that Congress did not design or intend the Clean Air Act to be used as a framework for regulating greenhouse gases.

It is unclear whether EPA's counsel did not make these arguments for this reason, but by losing the case, EPA gained the power to regulate CO_2, the most ubiquitous byproduct of industrial civilization.

Endnotes

1 The title of this article is inspired by the excellent collection of essays, *Climate Coup: Global Warming's Invasion of Our Government and Our Lives* (Patrick J. Michaels ed., 2011).

2 American Clean Energy and Security Act, H.R. 2454, 111th Cong. §§ 831-835 (2009), would bar regulation of greenhouse gases under the National Ambient Air Quality Standards, Hazardous Air Pollutants, New Source Review, and Title V programs. The text is available at http://www.gpo.gov/fdsys/pkg/BILLS-111hr2454pcs/pdf/BILLS-111hr2454pcs.pdf.

3 EPA, National Highway Traffic Safety Administration [hereinafter NHTSA], Final Tailpipe Rule, 75 Fed. Reg. 25,324, 25,424, 25,327 (May 7, 2010) (to be codified at 40 C.F.R. pts. 85, 86, and 600; 49 C.F.R. pts. 531, 533, 536, et al.) [hereinafter Tailpipe Rule], *available at* http://www.globalwarming.org/wp-content/uploads/2011/08/Final-Tailpipe-Rule.pdf.

4 U.S. Envtl. Prot. Agency, Nat'l Highway Traffic Safety Admin. & Cal. Air Res. Bd., Interim Joint Technical Assessment Report: Light Duty Vehicle Greenhouse Gas Emission Standards and Fuel Economy Standards for Model Years 2017-2025, at viii (2010) [hereinafter EPA, NHTSA, CARB, Interim Joint Technical Assessment Report], *available at* http://www.epa.gov/oms/climate/regulations/ldv-ghg-tar.pdf.

5 *Id.* at ix.

6 49 U.S.C. § 32902(a).

7 49 U.S.C. § 32902(b).

8 49 U.S.C. § 32904.

9 EPA, California Motor Vehicle Pollution Control Standards; Reconsideration of Previous Denial of Waiver of Preemption, 74 Fed. Reg. 7040 (Feb. 12, 2009), *available at* http://www.epa.gov/fedrgstr/EPA-AIR/2009/February/Day-12/a2913.pdf.

10 EPA, California State Motor Vehicle Pollution Control Standards; Notice of Decision Denying a Waiver of Clean Air Act Preemption for California's 2009 and Subsequent Model Year Greenhouse Gas Emission Standards for New Motor Vehicles, 73 Fed. Reg. 12,156 (Mar. 6, 2008) [hereinafter EPA, Notice of Decision Denying a Waiver], *available at* http://www.epa.gov/fedrgstr/EPA-AIR/2008/March/Day-06/a4350.pdf.

11 Nat'l Auto. Dealers Ass'n, Patchwork Proven: Why A Single National Fuel Economy Standard Is Better for America than A Patchwork of State Regulations (2009), *available at* http://www.nada.org/NR/rdonlyres/DBCC625E-2E8E-4291-8B23-B94C92AFF7C4/0/patchworkproven.pdf.

12 President Barack Obama, Remarks on National Fuel Economy Standards (May 19, 2009), transcript available at http://www.whitehouse.gov/the-press-office/remarks-president-national-fuel-efficiency-standards.

13 State of California Air Resources Board, Resolution 10-15, Feb. 25, 2010, *available at* http://www.arb.ca.gov/regact/2010/ghgpv10/res1015.pdf.

14 Letter from Edmund G. Brown, Jr., Attorney General, California, to Lisa Jackson, Administrator, EPA, and Raymond H. LaHood, Secretary, Department of Transportation at 2 (May 18, 2009), *available at* http://www.epa.gov/oms/climate/regulations/calif-atty-general.pdf.

15 Letter from Rep. Darrell E. Issa, U.S. House of Representatives, to Lisa Jackson, Administrator, EPA (Sept. 30, 2011), *available at* http://www.washingtonpost.com/r/2010-2019/WashingtonPost/2011/09/30/Health-Environment-Science/Graphics/oversight930.pdf.

16 The text of S.J. Res. 26 is available at http://www.gpo.gov/fdsys/pkg/BILLS-111sjres26pcs/pdf/BILLS-111sjres26pcs.pdf.

17 EPA, Endangerment and Cause or Contribute Findings for Greenhouse Gases Under Section 202(a) of the Clean Air Act, 74 Fed. Reg. 66496 (Dec. 15, 2009), *available at* http://epa.gov/climatechange/endangerment/downloads/Federal_Register-EPA-HQ-OAR-2009-0171-Dec.15-09.pdf.

18 Sen. Lisa Murkowski, opening statement, debate on S.J. Res 26, 156 Cong. Rec. S87, , S4789-4793 (2010), *available at* http://www.gpo.gov/fdsys/pkg/CREC-2010-06-10/pdf/CREC-2010-06-10-pt1-PgS4789.pdf#page=1.

19 Letter from Lisa Jackson, Administrator, EPA, to Sen. Jay Rockefeller (Feb. 22, 2010), *available at* http://epa.gov/oar/pdfs/lpj_letter.pdf.

20 Letter from Dave McCurdy, President and CEO, Alliance of Automobile Manufacturers, to the Honorable Nancy Pelosi, Harry Reid, John Boehner, and Mitch McConnell, U.S. Senate (Mar. 17, 2010), *available at* http://www.openmarket.org/wp-content/uploads/2010/06/auto-alliance-letter-to-house-and-senate-leaders-march-17-2010.pdf.

21 EPA, California State Motor Vehicle Pollution Control Standards; Notice of Decision Granting a Waiver of Clean Air Act Preemption for California's 2009 and Subsequent Model Year Greenhouse Gas Standards for New Motor Vehicles, 74 Fed. Reg. 32744 (July 8, 2009), *available at* http://www.gpo.gov/fdsys/pkg/FR-2009-07-08/pdf/E9-15943.pdf.

22 Letter from Stephen Johnson, Administrator, EPA, to Arnold Schwarzenegger, Governor, California (Dec. 19, 2007), *available at* http://www.epa.gov/otaq/climate/20071219-slj.pdf.

23 Letter from Arnold Schwarzenegger, Governor, California, and thirteen other Governors to Stephen Johnson, Administrator, EPA (Jan. 23, 2008), *available at* http://www.californiahydrogen.org/content/gov-schwarzenegger-joins-13-governors-letter-epa; *Oversight of EPA's Decision to Deny the California Waiver: Hearing Before the S. Comm. on Env't & Pub. Works*, 110th Cong. (2008) (statements of David Doniger, Natural Resources Defense Council, Connecticut Gov. M. Jodi Rell, Maryland Gov. Martin O'Malley, and Pennsylvania Gov. Edward G. Rendell), *available at* http://epw.senate.gov/public/index.cfm?FuseAction=Hearings.Hearing&Hearing_ID=8a7a0a80-802a-23ad-44b4-a1d4f1b4cc15.

24 EPA, California State Motor Vehicle Pollution Standards; Notice of Decision Granting a Waiver of Clean Air Act Prevention for California's 2009 and Subsequent Model Year Greenhouse Gas Emission Standards for New Motor Vehicles, 74 Fed. Reg. 32,744 (July 8, 2009), *available at* http://edocket.access.gpo.gov/2009/pdf/E9-15943.pdf.

25 EPA, Notice of Decision Denying a Waiver, *supra* note 10, at 12160-12164.

26 49 U.S.C. § 32919(a).

27 Cal. Air Res. Bd., Staff Report: Initial Statement of Reasons for Proposed Rulemaking, Public Hearing to Consider Adoption of Regulations to Control Greenhouse Gas Emissions from Motor Vehicles, Table 5.2-3 (2004), *available at* http://www.arb.ca.gov/regact/grnhsgas/isor.pdf.

28 Nat'l Research Council, Effectiveness and Impact of Corporate Average Fuel Economy (CAFE) Standards 31-43, especially Tables 3-1 & 3-2 (2002), *available at* http://www.nap.edu/openbook.php?isbn=0309076013.

29 The text of AB 1493 is available at http://en.wikisource.org/wiki/California_AB_1493.

30 Cal. Air Res. Bd., Staff Report, *supra* note 27, at 148.

31 Letter from James N. Goldstene, Executive Officer, California Air Resources Board, to Rep. Ed Whitfield (Mar. 22, 2011), *available at* http://www.globalwarming.org/wp-content/uploads/2011/06/CARB-QFR-Goldstene-EC-2011-02-09.pdf.

32 Tailpipe Rule, *supra* note 3, at 25424.

33 Brent D. Yacobucci & Robert Bamberger, Cong. Research Serv., Automobile and Light Truck Fuel Economy: The CAFE Standards (2010), *available at* http://nepinstitute.org/get/CRS_Reports/CRS_Energy/Energy_Efficiency_and_Conservation/CAFE_Standards_for_Light_Trucks_and_Autos.pdf.

34 Cal. Air Res. Bd., Title 13, Notice of Public Hearing to Consider Proposed Amendments to New Passenger Vehicle Greenhouse Gas Emission Standards for Model Years 2012-2016 to Permit Compliance Based on Federal Greenhouse Gas Emission Standards, Feb. 25, 2010, at 2, *available at* http://www.arb.ca.gov/regact/2010/ghgpv10/ghgpvnot.pdf.

35 44 U.S.C. § 2203(a).

36 Collin Sullivan, *Vow of Silence Key to White House-Calif. Fuel Economy Talks*, N.Y. Times, May 20, 2009, *available at* http://www.nytimes.com/gwire/2009/05/20/20greenwire-vow-of-silence-key-to-white-house-calif-fuel-e-12208.html.

37 Letter from Issa to Jackson, *supra* note 15, at 3-4.

38 Jeremy Anwyl, *A Letter to the EPA*, AutoObserver, July 28, 2011, http://www.autoobserver.com/2011/07/a-letter-to-the-epa.html.

39 *Top Auto Execs Gather with Obama in Washington to Boast About New Fuel Economy Requirements, But Is Everyone Really Happy?*, LeaseTrader.com, Aug. 1, 2011, http://news.leasetrader.com/archive/2011/08/01/Top-Auto-Execs-Gather-with-Obama-in-Washington-to-Boast.aspx.

40 Letter from Rep. Darrell Issa, U.S. House of Representatives, to Kathryn Reummler, Counsel to the President (Aug. 11, 2011), *available at* http://www.globalwarming.org/wp-content/uploads/2011/08/Darrel-Issa-letter-regarding-CAFE-deal-Aug-11-2011.pdf.

41 White House Flouts Agency Heads, Rolls out Backroom Deal on Fuel Economy Standard, Posting of Amy Sinden to CPR Blog, http://www.progressivereform.org/CPRBlog.cfm?idBlog=7426C8E2-CF0F-8446-72B4F05FF595E94B (July 29, 2011).

42 5 U.S.C. § 563, *available at* http://www.law.cornell.edu/uscode/5/usc_sec_05_00000563----000-.html.

43 5 U.S.C. § 564, *available at* http://www.law.cornell.edu/uscode/5/usc_sec_05_00000564----000-.html.

44 5 U.S.C. § 566, *available at* http://www.law.cornell.edu/uscode/5/usc_sec_05_00000566----000-.html.

45 49 U.S.C. §. 32902(b)(3)(B), *available at* http://www.law.cornell.edu/uscode/49/ usc_sec_49_00032902----000-.html.

46 EPA, NHTSA, 2017 and Later Model Year Light-Duty Vehicle Greenhouse Gas Emissions and Corporate Average Fuel Economy Standards*,* 76 Fed. Reg. 74,854 (proposed Dec. 1, 2011) (to be codified at 49 C.F.R. pts. 523, 531, 533, 536, and 537), *available at* http://www.nhtsa.gov/staticfiles/rulemaking/pdf/cafe/2017-25_CAFE_NPRM.pdf.

47 49 U.S.C. § 32902(f).

48 Letter from Issa to Ruemmler, *supra* note 40, at 4.

49 Massachusetts v. EPA, 549 U.S. 432 (2007).

50 Letter from Goldstene to Whitfield, *supra* note 31, at 2.

51 Tailpipe Rule, *supra* note 3, at25427-25428.

52 Letter from David McIntosh, Assistant Administrator, EPA, to Ed Whitfield, Chairman, House Energy and Power Subcommittee (July 11, 2011), *available at* http://www.globalwarming.org/wp-content/uploads/2011/06/EPA-response-to-harmonized-and-consistent-June-2011.pdf.

53 EPA, Endangerment Rule, *supra* note 17, at 66516.

54 The Copenhagen agreement aimed to stabilize CO_2-equivalent greenhouse gas concentrations (factoring in the estimated cooling effects of sulfate aerosols) at 450 parts per million by 2050. UNDP Env't & Energy Group, The Bali Road Map: Key Issues Under Negotiation 31 (2007), *available at* http://www.undp.se/assets/Ovriga-publikationer/Bali-road-map.pdf. In 2005, CO_2-equivalent concentrations were estimated at 311-435 parts per million. IPCC, Climate Change 2007: Mitigation 1, *available at* http://www.ipcc.ch/pdf/assessment-report/ar4/wg3/ar4-wg3-chapter1.pdf.

55 EPA, Endangerment Rule, *supra* note 17, at 66524.

56 Presidential Memorandum Regarding Fuel Efficiency Standards (May 21, 2010), *available at* http://www.whitehouse.gov/the-press-office/presidential-memorandum-regarding-fuel-efficiency-standards.

57 49 U.S.C. § 32902(a).

58 Letter from Issa to Jackson, *supra* note 15, at 10.

59 Letter from Shane Karr, Vice President, Federal Government Affairs, Alliance of Automobile Manufacturers, Honorable Darrell Issa, Chairman, House Committee on Oversight and Government Reform (Jan. 11, 2011), *available at* http://www.autonews.com/assets/PDF/CA72313112.PDF.

60 "President Obama today announced a historic agreement with thirteen major automakers to pursue the next phase in the Administration's national vehicle fuel program, increasing fuel economy to 54.5 miles per gallon for cars and light duty trucks by Model Year 2025." Press Release, The White House, President Obama Announces Historic 54.5 mpg Fuel Efficiency Standard (July 29, 2011), *available at* http://www.whitehouse.gov/the-press-office/2011/07/29/president-obama-announces-historic-545-mpg-fuel-efficiency-standard.

61 EISA, 49 U.S.C. § 32902(b)(2)(A), directs NHTSA to prescribe passenger car and light truck fuel economy standards that "achieve a combined fuel economy average for model year 2020 of at least 35 miles per gallon." NHTSA's 2008 rulemaking aimed to achieve a combined standard of 31.8 mpg for MY 2015 with average fuel economy increasing at the MY 2011-2015 rate (4.5% per year) during MYs 2016-2020. NHTSA, Average Fuel Economy Standards, Passenger Cars and Light Trucks, Model Years 2011-2015, Docket No. NHTSA-2008-0177, at 38, 700, *available at* http://keithhennessey.com/wp-content/uploads/2009/05/NHTSA_analysis.pdf. That works out to about 32.8 mpg in MY 2016 (my calculation). EPA and NHTSA's Tailpipe Rule, which implements the May 2009 "Historic Agreement," sets a combined average fuel economy standard of 34.1 mpg for MY 2016. Tailpipe Rule, *supra* note 3, at 25330. In short, the 2009 "Historic Agreement" is only moderately more aggressive than NHTSA's 2008 rulemaking pursuant to EISA.

62 Letter from Mary D. Nichols, Chairman, California Air Resources Board, to the Honorable Ray LaHood, Secretary, U.S. Department of Transportation, and the Honorable Lisa Jackson, Administrator, EPA 2 (July 28, 2011), *available at* http://www.epa.gov/otaq/climate/letters/carb-commitment-ltr.pdf.

63 U.S. Envtl. Prot. Agency, PSD and Title V Permitting Guidance for Greenhouse Gases (2011), *available at* http://www.epa.gov/nsr/ghgdocs/ghgpermittingguidance.pdf.

64 EPA, Proposed Settlement Agreement, Clean Air Act Citizen Suit, 75 Fed. Reg. 82,392 (Dec. 30, 2010), *available at* http://edocket.access.gpo.gov/2010/pdf/2010-32935.pdf; EPA, Proposed Settlement Agreement, 75 Fed. Reg. 82,390, (Dec. 30, 2010), *available at* http://edocket.access.gpo.gov/2010/pdf/2010-32929.pdf.

65 EPA, Regulating Greenhouse Gas Emissions under the Clean Air Act*,* Advance Notice of Proposed Rulemaking, 73 Fed. Reg. 44,354, 44,458-44,474 (July 30, 2008), *available at* http://edocket.access.gpo.gov/2008/pdf/E8-16432.pdf.

66 According to the Triggering Rule, PSD and Title V requirements apply to stationary sources of greenhouse gas emissions the moment the Tailpipe Rule takes effect (i.e. January 2, 2011). EPA, Reconsideration of Interpretation of Regulations That Determine Pollutants Covered by Clean Air Act Permitting Programs, 75 Fed. Reg. 17,004 (Apr. 2, 2010), *available at* http://www.gpo.gov/fdsys/pkg/FR-2010-04-02/pdf/2010-7536.pdf. According to the Tailoring Rule (see below), EPA has authority under the doctrines of "absurd results" and "administrative necessity" to exempt small emitters of greenhouse gases from PSD and Title V.

67 EPA, Regulating Greenhouse Gas Emissions under the Clean Air Act, *supra* note 65, at 44,420, 44,513.

68 EPA, Prevention of Significant Deterioration and Title V Greenhouse Gas Tailoring Rule, 75 Fed. Reg. 31,514, 31,515-31608 (June 3, 2010), *available at* http://www.cdphe.state.co.us/climate/FinalTailoringRule75FR31513.pdf.

69 EPA, Brief for Respondents at 26-27, Coal. for Responsible Regulation, Inc. v. U.S. Envtl. Prot. Agency (D.C. Cir. Sept. 16, 2011).

70 42 U.S.C. § 7661B(c).

71 EPA, Brief for Respondents at 26-27, Coalition for Responsible Regulation, Inc. v. U.S. Envtl. Prot. Agency (D.C. Cir. Sept. 16, 2011).

72 Mark P. Mills, *Carbon Dioxide a Pollutant? A Stunning Regulatory Burden*, Greening Earth Soc'y (1998), *available at* http://www.openmarket.org/wp-

content/uploads/2010/05/ges_report_mills_stunning-regulatory-burden1.pdf.

73 Memorandum from Jonathan Z. Cannon, General Counsel, EPA, to Carol M. Browner, Administrator, EPA, EPA's Power to Regulate Pollutants Emitted by Electric Power Generation Sources (Apr. 10, 1998), *available at* http://www.law.umaryland.edu/faculty/bpercival/casebook/documents/epaco2memo1.pdf.

74 International Center for Technology Assessment, Petition for Rulemaking and Collateral Relief Seeking the Regulation of Greenhouse Gas Emissions from New Motor Vehicles under Section 202 of the Clean Air Act at 10, (Oct. 20, 1999), *available at* http://www.icta.org/doc/ghgpet2.pdf.

75 EPA, Brief for Respondents at 35, Coal. for Responsible Regulation, Inc. v. U.S. Entl. Prot. Agency (D.C. Cir. Sept. 16, 2011).

76 Massachusetts v. EPA, 549 U.S. 531 (2007).

77 Brief for Petitioners at 28, Massachusetts v. EPA (U.S. Aug. 31, 2006), *available at* http://www.americanbar.org/content/dam/aba/publishing/preview/publiced_preview_briefs_pdfs_06_07_05_1120petitioners.authcheckdam.pdf.

78 Compare the similar language in the Title II and NSPS endangerment tests. Section 202(a):

> The Administrator shall by regulation prescribe . . . standards applicable to the emission of any air pollutant from any class or classes of new motor vehicles or new motor vehicle engines, which in his judgment cause, or contribute to, air pollution which may reasonably be anticipated to endanger public health or welfare.

Section 111(b): "He shall include a category of sources in such list if in his judgment it causes, or contributes significantly to, air pollution which may reasonably be anticipated to endanger public health or welfare."

79 Ctr. for Biological Diversity & 350.Org, Petition to Establish National Pollution Limits for Greenhouse Gases Under the Clean Air Act (2009), *available at* http://www.biologicaldiversity.org/programs/climate_law_institute/global_warming_litigation/clean_air_act/pdfs/Petition_GHG_pollution_cap_12-2-2009.pdf.

80 EPA, Regulating Greenhouse Gas Emissions under the Clean Air Act, *supra* note 65, at 44481.

81 Dialing in Your Own Climate, World Climate Report, http://www.worldclimatereport.com/index.php/2006/04/10/dialing-in-your-own-climate/ (Apr. 10, 2006).

82 American Clean Energy and Security Act, H.R. 2454, 111th Cong., at 692 (2009), *available at* http://www.gpo.gov/fdsys/pkg/BILLS-111hr2454pcs/pdf/BILLS-111hr2454pcs.pdf.

83 Under Clean Air Act section 179A, states are required to make such efforts as would be sufficient to attain or maintain a NAAQS "but for emissions emanating outside the United States." States therefore would not be responsible for offsetting, say, China's greenhouse gas emissions. Nonetheless, doing their "fair share" to reduce CO_2 concentrations to 350 parts per million within five years might be hard to distinguish from a deindustrialization program.

84 42 U.S.C. § 7401(b)(1).

85 EPA, Regulating Greenhouse Gas Emissions under the Clean Air Act, *supra* note 75, at 44477-44486.

86 Brief for Tennessee Valley Authority on Behalf of Petitioners at 2, Am. Elec. Power v. Connecticut (U.S. August 2010), *available at* http://www.openmarket.org/wp-content/uploads/2010/08/obama-brief-aep-v-connecticut-aug-2010.pdf.

Federalism & Separation of Powers

A Return to "the Heady Days"? The Supreme Court Addresses Whether the *Bivens* Doctrine Should Extend to Employees of Government Contractors in *Minneci v. Pollard*

By Robert T. Numbers, II and Lisa L. Dixon***

I. Introduction

On November 1, 2011, the Supreme Court heard oral arguments in *Minneci v. Pollard*, a case that will determine whether employees of government contractors can be held liable for damages for alleged constitutional violations under *Bivens v. Six Unknown Named Agents of Federal Bureau of Narcotics* and its progeny.[1] *Minneci* should resolve a circuit split between the Ninth Circuit, which held that employees of government contractors can be held liable under *Bivens*, and the Fourth, Tenth, and Eleventh Circuits, which held that they could not. In resolving this circuit split, the Supreme Court will need to address a number of questions that have divided lower courts for many years, such as whether employees of governmental contractors are considered federal actors; whether recognition of a *Bivens* claim is precluded if a plaintiff has alternative remedies, even if those remedies are not congressionally crafted; and how the imposition of asymmetrical liability costs on government contractors impacts availability of a *Bivens* remedy.

II. Overview of Existing Case Law

The Supreme Court's opinion in *Minneci*, regardless of which way it is decided, should resolve a question left undecided in *Correctional Services Corp v. Malesko*.[2] In *Malesko*, a divided Supreme Court[3] held that inmates in privately-operated correctional facilities could not bring a *Bivens* claim against the corporation that operated the facility.[4]

The Supreme Court found that extending *Bivens* liability to private corporations would not advance *Bivens*' goal "to deter individual federal officers from committing constitutional violations."[5] Allowing liability against an employer would undermine *Bivens*' deterrent effect because "if a corporate defendant is available for suit, claimants will focus their collection efforts on it, and not the individual directly responsible for the alleged injury."[6]

Additionally, the Court reasoned that extending *Bivens* to private corporations was in all meaningful aspects the same as allowing liability against the federal agency that employed an offending federal officer, a proposition the Supreme Court rejected in *FDIC v. Meyer*.[7] An alternative outcome would provide inmates in privately-operated facilities with a superior remedy to those enjoyed by inmates in government-operated facilities.[8] The Supreme Court explained that "[w]hether it makes sense to impose asymmetrical liability costs on private prison facilities alone is a question for Congress, not us, to decide."[9]

This concern over asymmetrical liability costs was a central factor in the Supreme Court's other main reason for refusing to extend *Bivens* to private corporations. Because inmates in privately-operated facilities could bring claims under state tort law, they "enjoy a parallel tort remedy that is unavailable to prisoners housed in Government facilities."[10] The existence of "alternative remedies [that] are at least as great, and in many respects greater, than anything that could be had under *Bivens*"[11] counseled against the "marked extension of *Bivens*"[12] sought by the plaintiff. The Supreme Court also noted that like inmates in facilities operated by the Bureau of Prisons, inmates in private facilities could bring concerns over their conditions of confinement to the attention of the BOP either through suits against the BOP for injunctive relief in federal courts or the BOP's Administrative Remedy Program.[13]

Although *Malesko* resolved the issue of whether a *Bivens* remedy was available against a private company that operates a correctional facility, both sides of the opinion recognized that they were not addressing whether the individual employees of the private contractor could be held liable under *Bivens*. The majority recognized that *Malesko* was not "seek[ing] a cause of action against an individual officer" as in prior cases extending *Bivens*.[14] Similarly, the dissent noted that "the question [of] whether a Bivens action would lie against the individual employee of a private corporation . . . is not raised in the present case."[15]

This open question regarding the liability of the employees of private contractors has vexed the lower courts for years: divided panels of the Fourth and Tenth Circuits and a unanimous panel of the Eleventh Circuit determined a *Bivens* remedy was not available, while a divided panel of the Ninth Circuit recently recognized a *Bivens* remedy against the employees of government contractors.

A. An Equally Divided Tenth Circuit Holds that Employees of Private Contractors Are Not Subject to Liability Under Bivens *in* Peoples v. CCA Detention Centers

Cornelius E. Peoples filed two *Bivens* complaints regarding his pretrial detention in a federal prison operated by CCA, a private, for-profit corporation.[16] In the first complaint (*Peoples I*), Peoples described how he feared attack by members of the "Mexican Mafia."[17] Despite filing formal and informal grievances, he was placed in the same prison unit as the

** Mr. Numbers is an attorney at Womble Carlyle Sandridge & Rice, PLLC in Raleigh, North Carolina. His practice focuses on civil rights litigation, business litigation, and appellate matters. Mr. Numbers frequently represents The GEO Group, Inc. and its employees in litigation in North Carolina that involves the question at issue in* Minneci. *However, Mr. Numbers does not represent The GEO Group, Inc. or its employees in* Minneci.

*** Mrs. Dixon lives in Winston-Salem, North Carolina, and has a J.D. from the University of Virginia School of Law.*

Mexican Mafia members and was not transferred to a new unit until after he was physically assaulted twice.[18] In the second complaint (*Peoples II*), he described how CCA had kept him in administrative segregation, where he did not have access to a law library, for thirteen months. He did not receive written notice of the reasons for administrative segregation immediately, and he did not receive a hearing for five months. He also believed that his phone calls with his attorney were unconstitutionally monitored.[19]

Citing *Malesko*, the district court dismissed *Peoples I* for lack of subject matter jurisdiction, as the availability of other remedies precluded a *Bivens* claim.[20] A different judge on the district court dismissed *Peoples II* on different grounds after assuming that a *Bivens* claim against the individual defendants was available, as the Tenth Circuit had not addressed the issue.[21] Peoples timely appealed both dismissals, and the Tenth Circuit considered them together.

After noting that the availability of a *Bivens* claim was not a jurisdictional question but a remedies question,[22] the Tenth Circuit panel held that a *Bivens* claim does not exist against individual employees of a private corporation operating a federal prison. As no courts of appeals had considered whether the existence of a state-law remedy precluded the extension of *Bivens* to employees of privately-operated prisons in the four years since *Malesko*, the court looked at two district court opinions.[23] In *Sarro v. Cornell Corrections, Inc.*, the Rhode Island district court held that while under *Malesko* a prisoner could not sue the corporation that operated the prison, the prisoner could sue the corporation's employees.[24] The *Sarro* court reasoned that this served the core purpose of *Bivens*, which was to deter individuals; that this would create parity among guards in federally-operated prisons and guards in privately-operated prisons; that no federal remedies were available to prisoners in pretrial detention like Sarro; and that allowing the availability of a *Bivens* remedy to rest on the availability of a state tort remedy would make federal prisoners' remedies vary by state, which *Bivens* sought to avoid.[25] In *Peoples I*, the court held that under *Malesko*, a *Bivens* claim was only available when the prisoner had no alternative remedy. Therefore, the availability of a state tort remedy precluded allowing a *Bivens* claim against individual employees of a private prison operator.[26]

The Tenth Circuit adopted the analysis of *Peoples I*, basing its opinion on the limited circumstances in which a *Bivens* action is available, as described in *Malesko*. The availability of an alternative state tort remedy removed Peoples from the category of plaintiffs who may pursue a *Bivens* claim.[27] The court also noted that whatever asymmetries in liability between federally- and privately-operated prisons existed, they were not created by the court; instead, the court maintained the status quo.[28] While there were policy reasons to extend *Bivens* liability to individual employees of private corporations operating federal prisons, the decision to do so is best left to Congress.[29]

Judge Ebel's dissent argued that the only "alternate 'cause of action' sufficient to preclude a *Bivens* action must be a *constitutional* cause of action."[30] Thus, state law tort remedy is insufficient. *Malesko* is best read as preserving *Bivens* claims against private individuals.[31] The best way to promote federal-state and public-private parallelism is through allowing suits against private individuals, as allowed under § 1983.[32] Allowing suits against individuals would provide uniformity of liability instead of making the protection of prisoners' constitutional rights "depend on the varying contours of state law" and on the facts.[33] And finally, the goal of individual deterrence embodied in the *Bivens* remedy is undermined by not allowing federal prisoners to sue individual private prison operators.[34]

The issue was eventually addressed by the Tenth Circuit en banc. However, the en banc court was evenly divided on the issue, which meant that while the district court's dismissal of the claim was affirmed, the case carried no precedential value.[35]

B. The Fourth Circuit Agrees that a Bivens *Remedy Is Unavailable but Is Divided over Whether GEO's Employees Are Federal Actors in* Holly v. Scott[36]

Ricky Holly, an inmate incarcerated at Rivers Correctional Institution in Winton, North Carolina, claimed that GEO's employees violated his Eighth Amendment rights by failing to properly treat his diabetes.[37] Holly brought suit under *Bivens* against the facility's warden and his treating physician.[38] The defendants were "both employed directly by GEO, and thus the only link between their employment and the federal government is GEO's contract with the BOP."[39] At the district court level, the defendants unsuccessfully sought to have Holly's claim dismissed on the basis that as employees of a private corporation they were not subject to liability under *Bivens*.[40]

On appeal, a divided panel of the Fourth Circuit Court of Appeals reversed the district court's ruling and held that GEO's employees were not subject to suit under *Bivens*. The majority opinion, authored by Judge J. Harvie Wilkinson and joined by Judge R. Bryan Harwell of the United States District Court of South Carolina (sitting by designation), began by reviewing the Supreme Court's reluctance to expand the *Bivens* doctrine almost since its inception.[41] This reluctance is based in part on the fact that "'a decision to create a private right of action is one better left to legislative judgment in the great majority of cases.'"[42] Moreover, as Congress established the statutory provisions that allowed inmates to be housed in private facilities based on "the belief that private management would in some circumstances have comparative advantages in terms of cost, efficiency, and quality of service . . . add[ing] a federal damages remedy to existing avenues of inmate relief might well frustrate a clearly expressed congressional policy."[43]

In the majority's view, extending *Bivens* to GEO's employees was precluded by two factors. "First, defendants are private individuals, not government actors. Second, Holly has an adequate remedy against defendants for his alleged injuries under state law."[44]

The defendants' status as private individuals was a key aspect of the Court's analysis because of "the importance of a party's private status in our constitutional scheme. The Bill of Rights is a negative proscription on *public* action—to simply apply it to *private* action is to obliterate 'a fundamental fact of our political order.'"[45] Restricting the applicability of the Bill of Rights to public action "'preserve[s] an area of individual freedom by limiting the reach of federal law and federal judicial power.'"[46] By maintaining this distinction between public and private action, courts "maintain the Bill of Rights as a shield

that protects private citizens from the excesses of government, rather than a sword that they may use to impose liability upon one another."[47]

Although the Court "harbor[ed] some doubt as to whether such liability would ever be appropriate," it went on to analyze whether the GEO defendants could be considered federal actors under the "state action" doctrine applied to constitutional claims under Section 1983.[48] The Court undertook this analysis despite the fact that "the Supreme Court 'ha[s] never held that the contours of *Bivens* and § 1983 are identical.'"[49]

Ultimately, the Court found that GEO's employees are not federal actors under the public function test because they did not exercise powers traditionally reserved to the state.[50] In reaching this conclusion, the majority focused on the Supreme Court's determination in *Richardson v. McKnight*[51] that the operation of correctional facilities was not a traditional public function because "the private operation of jails and prisons existed in the United States in the eighteenth and nineteenth centuries and in England, the practice dated back to the Middle Ages."[52] As Holly's alleged injury arose "out of defendants' operation of the prison [and] not the fact of Holly's incarceration," the defendants did not engage in a traditionally public function and therefore were not federal actors subject to *Bivens* liability.[53]

The Fourth Circuit also rejected Holly's contention that the Supreme Court's holding in *West v. Atkins*[54] required the court to find that "the provision of medical care to an inmate is always a public function, regardless of what entity operates the correctional facility where he is housed."[55] In *West*, the Supreme Court held that "a physician employed by North Carolina to provide medical services to state prison inmates [] acted under color of state law for purposes of § 1983."[56] The crucial distinction, according to the Fourth Circuit, was that in this case the defendants had no direct relationship with the governmental entity.[57] The Court could not conclude "that provision of medical care in a private prison is somehow a 'public function' while maintaining fidelity to *Richardson* that the prison's general operation is not."[58]

The Court also held that the existence of adequate, and perhaps superior, state tort remedies precluded an extension of *Bivens*. According to the majority, the Supreme Court has only extended *Bivens* in situations where the plaintiff lacked any alternative remedy against the allegedly offending individual.[59] Here "North Carolina law . . . supplies Holly with multiple claims against the individual defendants."[60] Thus, there was no need to recognize a *Bivens* claim against the GEO defendants.

Judge Dianna Gribbon Motz filed an opinion that, while concurring in judgment, vehemently disagreed with the majority's analysis of the federal actor issue. Under Judge Motz's analysis, the question was foreclosed by the Supreme Court's holding in *West*. She believed that the defendants qualified as federal actors because they "perform a public function delegated to them by the federal government, and they assume the necessary obligations inherent in that function."[61] However, Judge Motz concurred in the judgment because the availability of adequate state remedies precluded recognition of a new *Bivens* cause of action.[62]

C. The Eleventh Circuit Determines that the Presence of Adequate State Remedies Precludes a Bivens *Claim in* Alba v. Montford

Luis Francisco Alba filed a *Bivens* suit against individual employees of Corrections Corporation of America ("CCA"), a private corporation that operated the federal prison in Georgia in which he was incarcerated.[63] He alleged that, pursuant to a CCA policy, the employees failed to provide him with proper post-operative treatment after surgery for a benign goiter in his throat.[64] The district court dismissed Alba's claim at the initial screening stage because, as Alba had "adequate remedies in state court," it failed to state a *Bivens* claim.[65]

The Eleventh Circuit assumed without deciding that CCA was a government actor, but it unanimously agreed with the district court that the availability of remedies under state tort law rendered a *Bivens* claim unavailable.[66] An alternative remedy sufficient to defeat a *Bivens* claim does not have to be a federal remedy because *Malesko* rejected that argument and because the *Bivens* court expressed concern that *Bivens* would not be able to recover under state tort law.[67] Georgia tort law in this instance was not inconsistent with the rights protected by the Eighth Amendment and even provided Alba with superior means of recovery.[68] The court also noted that, while Alba did not sue CCA, he was challenging CCA's policy instead of the conduct of the individual employees, and the Supreme Court "made it abundantly clear" in *Malesko* that "*Bivens* will not support an action challenging the conduct or policy of a non-individual defendant."[69]

D. The Ninth Circuit Holds that GEO's Employees are Subject to Liability Under Bivens *in* Pollard v. Minneci

In 2001, Richard Pollard,[70] a federal inmate, was incarcerated at Federal Correctional Institution at Taft in California, a facility operated by the GEO Group, Inc.[71] pursuant to a contract with the Federal Bureau of Prisons.[72] On April 7, 2001, Pollard slipped and fell on a cart left in a doorway while working in the facility's butcher shop.[73] The facility medical staff took x-rays and determined that Pollard may have fractured both of his elbows. [74] He was placed in a bilateral sling and referred to an orthopedic clinic outside of the facility.[75]

As Pollard prepared to leave the facility for his orthopedic appointment, facility staff ordered him to put on a prison jumpsuit.[76] Although Pollard claimed that, as a result of the injuries to his elbows, putting his arms through the sleeves of the jumpsuit "would cause him excruciating pain," he was required to do so before leaving the facility.[77] Additionally, Pollard was required to wear a "black box" restraining device on his wrists despite complaints about the pain caused by the device.[78]

The orthopedist who saw Pollard diagnosed him with "serious injuries to his elbows and recommended that his left elbow be put into a posterior splint for approximately two weeks."[79] However, when Pollard returned to the facility he was told that "due to limitations in staffing and facilities" he would not receive the treatment recommended by the orthopedist.[80] Pollard also claimed that over the next several weeks facility staff failed to make accommodations that would allow him to feed or bathe himself, that he was required to work in spite of

his injuries, and that he was required to wear the "black box" device before he was allowed to go to a follow-up appointment with his orthopedist."[81]

Pollard, proceeding *pro se*, filed suit in the Eastern District of California against GEO and a number of GEO's employees. The complaint sought monetary damages from the defendants under *Bivens* for a violation of his Eighth Amendment rights. The district court dismissed the GEO suit based on the Supreme Court's holding in *Malesko*,[82] and subsequently dismissed the suit against the GEO employees based on the Tenth Circuit's holding in *Peoples* and the Fourth Circuit's holding in *Holly*.[83]

Pollard, now represented by counsel, appealed the dismissal to the Ninth Circuit Court of Appeals. A divided three-judge panel of the Ninth Circuit[84] reversed the district court's holding with respect to GEO's employees. The majority opinion, authored by Judge Paez and joined by Judge Hug, took direct aim at the reasoning of the various decisions that rejected an extension of *Bivens*, focusing most of its energy on the Fourth Circuit's decision in *Holly*.[85] Ultimately, the majority opinion held that Pollard should be able to bring a *Bivens* claim against GEO's employees because "(1) the GEO employees act under color of federal law for purpose of Bivens liability and (2) the availability of a state tort remedy does not foreclose Pollard's ability to seek redress under Bivens."[86]

The majority determined that GEO's employees are federal actors based on a review of the "'state action' principles developed by the Supreme Court in suits brought under 42 U.S.C. § 1983."[87] The majority did not explain why it is applying the state action principles developed under Section 1983 other than to say that both the Ninth Circuit and "[o]ther circuits have . . . recognized the similarity of the § 1983 and *Bivens* doctrines."[88]

In order to determine whether GEO's employees were engaged in "state action," the majority applied the "variation" of the public function test applied by the Supreme Court in *West*.[89] Under this test, a private employee is a state actor and subject to liability under Section 1983 if the employee is "'fully vested with state authority to fulfill essential aspects' of the state's" constitutionally-imposed responsibilities.[90] According to the majority, GEO's employees must be federal actors, and therefore amenable to suit under *Bivens*, because "there is no principled difference to distinguish the activities of the GEO employees in this case from the governmental action identified in *West*."[91] Ultimately, GEO's employees must be amenable to suit under *Bivens* because Pollard's alleged constitutional "deprivation was caused . . . by the federal government's exercise of its power to punish Pollard by incarceration and to deny him a venue independent of the federal government to obtain needed medical care."[92]

The majority went on to explicitly reject what it considered the "illogical reading of *West*" employed by the Fourth Circuit in *Holly*.[93] Under the majority's reading of *West*, there is no distinction between the actions of a private individual working directly for the governmental entity and an individual who is working for a private corporation that has a contract with a governmental entity.[94] *West* provides that "'contracting out' care 'does not relieve' the government of its 'constitutional duty' to provide adequate care or 'deprive inmates of the means to vindicate their Eighth Amendment rights.'"[95]

The majority also found unconvincing the Fourth Circuit's refusal to define GEO's employees as federal actors due to their inability to raise the defense of qualified immunity based on the Supreme Court's holding in *Richardson*.[96] Initially, the *Pollard* majority rejected the Fourth Circuit's reliance on this factor in its analysis because "the Court in *Richardson* expressly noted that it 'did not address [] whether the defendants are liable under § 1983 even though they are employed by a private firm.'"[97] But it goes on to assert that its determination that GEO's employees are federal actors is correct because "in *Malesko*, the Supreme Court explicitly left open the possibility that private prison employees could act under the color of federal law and therefore face *Bivens* liability."[98]

More central to the disagreements between the Ninth and Fourth Circuits was the question of the relevant function to be analyzed in determining whether GEO's employees performed a public function.[99] The Ninth Circuit rejected the Fourth Circuit's determination that the incarceration of prisoners and the management of prisons were separate and distinct governmental functions based on *West's* statement that a prisoners' constitutional injury from inadequate medical care is "'caused, in the sense relevant for state-action inquiry, by the State's exercise of its right to punish [the prisoner] by incarceration.'"[100] In the end, the Ninth Circuit

> decline[d] to artificially parse [the power to incarcerate] into its constituent parts—confinement, provision of food and medical care, protection of inmate safety, etc.—as that would ignore that those functions all derive from a single public function that is the sole province of the government: "enforcement of state-imposed deprivation of liberty."[101]

As the power to incarcerate has been "traditionally the exclusive prerogative of the [government]," the Ninth Circuit determined that it was appropriate to hold that GEO's employees were undertaking a public function.[102]

The Ninth Circuit then recognized that a judicially-created *Bivens* remedy is not necessarily required simply by the fact that it determined GEO's employees to be federal actors.[103] Before recognizing a new *Bivens* remedy, a court must also analyze (1) "whether any alternative existing process for protecting the [constitutionally recognized] interest amounts to a convincing reason for the Judicial Branch to refrain from providing a new and freestanding remedy in damages," and (2) where there exist "any special factors counseling hesitation before authorizing a new kind of federal litigation."[104]

The *Pollard* majority determined that the availability of state tort remedies to redress inmate injuries was insufficient to prohibit the court from recognizing a new *Bivens* remedy. Despite the Supreme Court's specific language that *Bivens* remedies are only available in the absence of any alternative remedy, the Ninth Circuit determined that "the mere existence of a potential state law claim did not suffice to preclude a *Bivens* action."[105]

The majority went on to hold that the existence of state tort remedies did not provide "convincing reasons" to refrain

from recognizing a new *Bivens* claim for two primary reasons. Initially, the majority found that Congress's failure to create a statutory remedy to address claims by federal inmates in privately-managed facilities counseled in favor of a judicially-created remedy.[106] While the Ninth Circuit cites a number of cases in support of this proposition, the only case that directly deals with the issue is *Carlson v. Green*.[107] *Carlson* is among those cases from the bygone era of "heady days in which the [Supreme] Court assumed common-law powers to create causes of action . . . ."[108] Additionally, relying on *Carlson* and the Ninth Circuit's opinion in *Castaneda v. United States*,[109] an opinion rejected by the Supreme Court,[110] the Ninth Circuit found that state tort remedies are insufficient to preclude a judicially-created *Bivens* remedy because the contours of the remedies available to each inmate will vary depending on which state the inmate's claim arises.[111]

According to the Ninth Circuit, there were also no special factors counseling hesitation from recognizing a new *Bivens* remedy.[112] Adopting a *Bivens* remedy for inmates in private facilities would produce a workable cause of action because under Eighth Amendment jurisprudence "the applicable standards are clear. There is no need for the district court to craft new standards or remedies to address Pollard's claims."[113] The Ninth Circuit also held that recognizing a *Bivens* cause of action would enhance the doctrine's core purpose of deterring individual officers from committing unconstitutional acts because (1) it would allow inmates to avoid liability caps, prefiling certification requirements, and other limitations placed on state action; and (2) "*Bivens* may allow for recovery of greater damages in some cases than a state tort law remedy."[114] However, existence of these asymmetrical liability costs did not rise to such a level as to "counsel hesitation in recognizing a *Bivens* remedy here."[115]

Based upon its finding that GEO's employees were federal actors, that inmates in privately-operated facilities lacked a sufficiently adequate remedy to preclude recognizing a *Bivens* remedy, and that there were no special factors counseling hesitation against recognizing such a claim, the majority reversed the district court's decision and allowed Pollard's *Bivens* claim against GEO's employees to proceed.[116]

Dissenting from the majority opinion, Judge Restani commented that "[t]he majority overlooks the reality that the Supreme Court has recognized *Bivens* causes of action only where federal officials, by virtue of their position, enjoy impunity, if not immunity, from damages liability because of gaps or exemptions in statutes or in the common law."[117] Such gaps did not exist in Pollard's case because "his 'alternative remedies [under state tort law] are at least as great, and in many respects greater than anything that could be had under *Bivens*."[118] Given the existence of an adequate remedy to address Pollard's alleged injury, "bedrock principles of separation of powers foreclose[] judicial imposition of a new substantive liability.[119] Judge Restani also noted that, contrary to the majority's assertion, the Supreme Court has considered state tort remedies sufficient to preclude the recognition of a new *Bivens* remedy.[120] Moreover, she asserted that "[i]t is to much of a stretch to infer, as the majority does," that the Supreme Court would have reached the same result in *Wilkie* had the case involved a handful of state law tort claims instead of an amalgamation of state, federal, administrative, and judicial remedies.[121] Moreover, there was no compelling need to ensure uniformity in this area of the law because employees of private entities do not receive the same immunities as federal officials and the basic elements of state law tort claims "are fundamentally the same in every state."[122]

Judge Restani also disagreed with the majority's analysis regarding the presence of special factors counseling hesitation in recognizing a new type of *Bivens* claim. In her view, feasibility concerns did not counsel in favor of a *Bivens* remedy for all inmates in private facilities because "allowing a *Bivens* action to go forward only where a plaintiff would otherwise have no alternative remedy [under state law] is not unduly complicated," and she could not conceive of any circumstances in which state tort law would not provide a remedy for an inmate's claim.[123] Additionally, Judge Restani did not believe that recognizing a *Bivens* cause of action would further the deterrence goals of Bivens because state law provided an adequate deterrent effect through awards of "compensatory and punitive damages for the same conduct . . . ."[124] Finally, recognizing a *Bivens* cause of action would only exacerbate the existing asymmetrical liability costs between inmates in private and public facilities because of the increased types of claims that may be brought against the employees of private facilities and their lack of qualified immunity from *Bivens* claims.[125]

Joining Judge Restani in her rejection of the majority's opinion were the eight judges of the Ninth Circuit who dissented from the denial of the GEO defendants' petition to have the matter heard en banc.[126] Judge Bea and those who joined him believed that it was contrary to Supreme Court precedent to recognize a *Bivens* cause of action because "Pollard has a viable suit in state court against each of the jailor defendants under theories of intentional or negligent tort or medical malpractice."[127] Moreover, the dissenting judges found the majority's concerns regarding lack of uniformity were misplaced because of the existence of "an adequate, and arguably superior, tort claim under state law."[128] Ultimately, the dissenting judges found that

> [t]he panel's explanation for this disagreement [with other circuits] reduces to a policy judgment that plaintiffs in this situation should have another forum in which to pursue these claims even though an adequate state remedy exists. Whatever may be the merits of that policy judgment, it is for Congress, not us, to make.[129]

III. Analysis

A review of the various circuit court opinions, and particularly *Holly* and *Minneci*, demonstrates where the fault lines are on this issue. First, there is a dispute over whether the employees of private corporations that operate correctional facilities constitute federal actors. Second, the courts disagree whether the availability of state tort remedies precludes the recognition of a *Bivens* remedy. Finally, there is controversy over whether and to what extent the difference between the private and public entities that operate correctional facilities impacts a court's ability to recognize a *Bivens* remedy.

With regard to the federal actor question, there is a clear dispute over whether this question is governed by *West*'s holding that governmental entities cannot contract away their constitutional responsibilities or *Richardson*'s statement that the operation of correctional facilities is not a traditional public function.[130] Interestingly, at the Supreme Court neither Minneci, nor the United States in its amicus brief in support of Minneci, spend any meaningful time discussing the federal actor issue. Minneci asserts that the resolution of this matter does not "require a determination of whether employees of private prison operators exercise governmental powers as a general matter."[131] Similarly, the United States indicates that "the Court need not reach this issue to decide this case," but goes on to indicate that if the Court does take up the issue, "the government submits that private prison contractors do act 'under color of law' for certain purposes, including for purposes of federal criminal law."[132] Pollard does not make any arguments in support of the position that GEO's employees are federal actors, but instead asserts that Minneci's failure to address the issue constitutes a concession of that point.[133] While Minneci and the United States are correct in their assertion that resolution of the federal actor issue is not necessary to resolve the case, addressing the issue of when private action reaches the level of government action could provide much-needed guidance on this unsettled question. This is particularly true in light of the reality that while *Minneci* only deals with prison operators, its reasoning will be employed in litigation involving government contractors beyond the corrections industry.

The courts of appeals disagree over whether the availability of state tort remedies preclude recognition of a *Bivens* claim. Primarily the dispute centers over whether congressionally-crafted remedies are the exclusive means of prohibiting a *Bivens* claim or whether any adequate remedy will do. According to the Ninth Circuit, only remedies crafted by Congress should be considered remedies adequate to defeat a *Bivens* cause of action: "The mere availability of a state law remedy does not counsel against allowing a *Bivens* cause of action. . . . [O]nly remedies crafted *by Congress* can have such a preclusive effect."[134] The majorities in *Holly*, *Peoples*, and *Alba* held that the availability of *any* alternative remedy precludes a *Bivens* remedy: "[A] *Bivens* claim should not be implied unless the plaintiff has no other means of redress or unless he is seeking an otherwise nonexistent cause of action against the individual defendant."[135] The court's judgment of what constitutes an *adequate* alternative remedy will be central to its decision.

Based upon the Supreme Court's holdings in *Malesko*, where state tort law remedies seem to have been sufficient to bar a *Bivens* cause of action, it appears that Minneci has the better argument. However, the Ninth Circuit is correct in its statement that the Supreme Court has been less than clear on this point.[136]

Regardless of how the Supreme Court resolves this matter, asymmetrical liability costs will exist between private and public providers of correctional services. If there is a *Bivens* remedy against the employees of privately-operated correctional facilities, inmates in these facilities will have both *Bivens* and state law claims at their disposal. In addition to having an additional set of claims, inmates in privately-operated facilities will have an easier path to recovery on their *Bivens* claims because the defendants will not be entitled to the defense of qualified immunity. A prisoner in a privately-operated prison would be able to recover damages from individual prison officials for violations that were not clearly established, while a prisoner in a federally-operated prison would not be able to recover for the same violation. Similarly, a prison official in a privately-operated prison would be subjected to personal liability in more situations than a prison official in a federally-operated prison due to the multiple causes of action available to plaintiffs. Of course, if the Supreme Court rejects the Ninth Circuit's reasoning, inmates in privately-operated facilities will only be able to recover through state law claims. Either way inmates in privately-operated facilities are in a vastly different position than inmates in federally-operated facilities.

However, a case pending in the Eastern District of North Carolina has presented a potential route for inmates in privately-operated facilities to seek redress for constitutional violations. In *Mathis v. The GEO Group, Inc.*,[137] the court has indicated that the Bureau of Prisons may be held liable for an Eighth Amendment violation if the BOP's on-site contract monitor is aware of unconstitutional conduct by the contractor or its employees and is deliberately indifferent to the unconstitutional acts.[138] Although *Mathis* does not involve a *Bivens* claim against the BOP's on-site contract monitor, the potential exists that such a claim could be viable. If such a claim were recognized, it would address at least some of the concerns raised over the government contracting away its constitutional responsibilities.[139]

Ultimately, the only way the asymmetrical liability issue can be addressed, barring some wholesale change in the law, is if Congress addresses the issue. Congress has the ability and the authority to produce a remedial scheme to address tortious conduct, both of a constitutional nature and otherwise, by both government contractors and their employees. A congressionally-crafted cause of action would, most likely, have the additional benefit of preempting state law tort claims, which would provide the uniformity of liability sought by those who support Pollard's position.

Despite the existence of the controversy over the potential liability of the employees of government contractors for several years, Congress has shown neither the interest, nor the will, to act on this issue. It is unlikely that it will do so at any point in the near future and nearly certain that the issue will not be addressed prior to the Supreme Court resolving *Minneci*.

IV. Conclusion

A decision to uphold the Ninth Circuit could signal a shift in the Supreme Court's approach to recognizing judicial causes of action. The Court may be less willing to wait for Congress to act to protect constitutional rights and more willing to fill in the gaps where Congress has been silent: a return to the "heady days in which [courts] assumed common-law powers to create causes of action."[140] As in *Davis* and *Carlson*, it could also be a narrow expansion limited to the facts in *Minneci* and only apply to suits by prisoners against prison officials employed by a private corporation that operates a federal prison.

Practically, a decision upholding the Ninth Circuit could lead to increased costs in government contracting and increased litigation. If employees of government contractors may be subject to individual liability for possible constitutional violations, they will demand higher pay or indemnification, which will in turn drive up the cost of the contracts to the government. The courts will also face more suits filed by prisoners. Given the increasing number of prisons operated by private corporations, the number of *Bivens* suits could dramatically increase.[141]

However, given the Supreme Court's reluctance to expand *Bivens*,[142] it is likely that the Court granted certiorari to undo the expansion of *Bivens* approved by the Ninth Circuit. The unavailability of the *Bivens* remedy where another adequate remedy is available is relatively uncontroversial, and this will likely provide a firm basis for the Supreme Court to agree with the Fourth, Tenth, and Eleventh Circuits and refuse to extend *Bivens* to individual employees of private prison operators.

Endnotes

1 403 U.S. 388 (1971).

2 534 U.S. 62 (2001).

3 Chief Justice Rehnquist authored the majority opinion and was joined by Justices O'Connor, Scalia, Kennedy, and Thomas. Justice Scalia wrote a concurring opinion, joined by Justice Thomas, that explained his broad rejection of the reasoning underlying *Bivens* and his desire to "limit *Bivens* and its two follow on cases to the precise circumstances they involved." *Id.* at 524 (Scalia, J., concurring) (citations omitted). Justice Stevens filed a dissenting opinion, which was joined by Justices Souter, Ginsburg, and Breyer.

4 *Id.* at 63.

5 *Id.* at 70.

6 *Id.* at 71.

7 *Id.* at 70-71 (citing FDIC v. Meyer, 510 U.S. 471 (1994)).

8 *Id.* at 71-72.

9 *Id.* at 72.

10 *Id.* at 72-73.

11 *Id.* at 72.

12 *Id.* at 74.

13 *Id.*

14 *Id.* At one point, Malesko attempted to add the individuals allegedly responsible for his injuries to the suit as defendants, but they were dismissed on statute of limitations grounds. *Id.* at 65.

15 *Id.* at 79 n.6 (Stevens, J., dissenting).

16 Peoples v. CCA Detention Ctrs., 422 F.3d 1090, 1093 (10th Cir. 2005), *vacated in relevant part by equally divided court*, 449 F.3d 1097 (10th Cir. 2006) (en banc).

17 *Id.* at 1093.

18 *Id.* at 1093-94.

19 *Id.* at 1094.

20 *Id.*

21 *Id.* at 1094-95.

22 *Id.* at 1096.

23 *Id.* at 1100.

24 Sarro v. Cornell Corrections, Inc., 248 F. Supp. 2d 52 (D.R.I. 2003). The District Courts for the Districts of New Jersey and Kansas subsequently adopted the *Sarro* opinion. Jama v. INS, 343 F. Supp. 2d 338, 362-63 (D.N.J. 2004); Purkey v. Corrections Corp. of Am., 339 F. Supp. 2d 1145, 1148-51 (D. Kan. 2004).

25 *Peoples*, 422 F.3d at 1100-01.

26 *Id.* at 1101.

27 *Id.*

28 *Id.* at 1102-03.

29 *Id.* at 1103.

30 *Id.* at 1109 (Ebel, J., dissenting).

31 *Id.* at 1110.

32 *Id.* at 1110-12.

33 *Id.* at 1112-13.

34 *Id.* at 1113.

35 Peoples v. CCA Detention Ctrs., 449 F.3d 1097 (10th Cir. 2006) (en banc).

36 Womble Carlyle represented the GEO Defendants in *Holly v. Scott*, but Mr. Numbers did not participate in the case.

37 Holly v. Scott, 434 F.3d 287, 288 (2006).

38 *Id.*

39 *Id.*

40 *Id.*

41 *Id.* at 290.

42 *Id.* at 289 (quoting Sosa v. Alvarez-Machain, 542 U.S. 692, 727 (2004)).

43 *Id.* at 290.

44 *Id.*

45 *Id.* at 291 (emphasis in original) (quoting Lugar v. Edmondson Oil Co., 457 U.S. 922, 937 (1982)).

46 *Id.* (emphasis in original) (quoting *Lugar*, 457 U.S. at 936).

47 *Id.* at 292 (citing Moose Lodge No. 107 v. Irvis, 407 U.S. 163, 172 (1972)).

48 *Id.* at 291-92.

49 *Id.* at 292 (quoting Correctional Servs. Corp. v. Malesko, 534 U.S. 62, 82 (2001)).

50 *Id.* at 293.

51 521 U.S. 399 (1997).

52 *Holly*, 434 F.3d at 293 (quoting *Richardson*, 521 U.S. at 405-07).

53 *Id.* at 293.

54 487 U.S. 42 (1988).

55 *Holly*, 434 F.3d at 293-94.

56 *Id.* at 294 (quoting *West*, 487 U.S. at 54).

57 *Id.*

58 *Id.*

59 *Id.* at 295.

60 *Id.* at 296.

61 *Id.* at 299 (Motz, J., concurring in judgment only).

62 *Id.* at 302-03.

63 Alba v. Montford, 517 F.3d 1249, 1251 (11th Cir. 2008).

64 *Id.* at 1251.

65 *Id.* at 1251-52.

66 *Id.* at 1254.

67 *Id.* (citing Correctional Servs. Corp. v. Malesko, 534 U.S. 62, 69

(2001)).

68 *Id.* at 1255-56.

69 *Id.* at 1255.

70 As this case never proceeded past the frivolity review stage at the district court, the facts are all taken from Pollard's complaint and, presumably, would be denied in whole or in part by the defendants if the case moves forward.

71 At the time of the events set out in the complaint, the GEO Group, Inc. was known as Wackenhut Correctional Corporation. Pollard v. GEO Group, Inc., No. 07-16112, slip op. at 8147 n.2 (9th Cir. June 7, 2010).

72 *Id.* at 1-2.

73 *Id.* at 2.

74 *Id.* at 8147.

75 *Id.*

76 *Id.* at 8147-48.

77 *Id.* at 8148.

78 *Id.*

79 *Id.*

80 *Id.*

81 *Id.*

82 *Id.*

83 Pollard v. Wackenhut Correctional Corp., 1:01-cv-06078, slip op. at 4-6 (E.D. Cal. Sept. 14, 2006).

84 The panel was comprised of The Honorable Proctor Hug, Jr. and Richard A. Paez of the Ninth Circuit and The Honorable Jane A. Restani, chief Judge of the United States Court of International Trade, sitting by designation.

85 Pollard v. GEO Group, Inc., No. 07-16112, slip op. at 8153 (9th Cir. June 7, 2010) ("We recognize that the former holding directly conflicts with the Fourth Circuit's holding in Holly v. Scott and the latter conflicts with both Holly and the Eleventh Circuit's holding in *Alba v. Montford*.") (citations omitted).

86 *Id.* at 8152.

87 *Id.* at 8154.

88 *Id.* at 8154-55.

89 *Id.* at 8155 (citing West v. Atkins, 487 U.S. 42, 49-51 (1988)).

90 *Id.* at 8155-56 (quoting *West*, 487 U.S. at 57).

91 *Id.* at 8156.

92 *Id.* at 8157.

93 *Id.*

94 *Id.* at 8156.

95 *Id.* (quoting *West*, 487 U.S. at 55-56).

96 *Id.* at 8158.

97 *Id.* (quoting Richardson v. McKnight, 521 U.S. 399, 413 (1997)).

98 *Id.* at 8160 (citing Correctional Servs. Corp. v. Malesko, 534 U.S. 62, 65 (2001)).

99 *Id.* at 8159.

100 *Id.* (quoting *West*, 487 U.S. at 55).

101 *Id.* at 8160 (quoting *Richardson*, 521 U.S. at 416 (Scalia, J., dissenting)).

102 *Id.* (quoting Rendell-Baker v. Kohn, 457 U.S. 830, 842 (1982)).

103 *Id.* at 8161 ("Even where defendants have engaged in federal action, we do not always allow *Bivens* suits to go forward.").

104 *Id.* at 8162-63 (quoting Wilkie v. Robbins, 551 U.S. 537, 550 (2007)).

105 *Id.* at 8165-66 (quoting Correctional Servs. Corp. v. Malesko, 534 U.S. 62, 70 (2001); *Wilkie*, 551 U.S. at 551-54).

106 *Id.* at 8166-67.

107 446 U.S. 14, 23(1980).

108 *Malesko*, 534 U.S. at 75 (Scalia, J., concurring).

109 546 F.3d 682 (9th Cir. 2008).

110 Hui v. Castaneda, 130 S. Ct. 1845 (2010).

111 *Pollard*, slip op. at 8167-69.

112 Among the factors recognized by the Supreme Court as counseling hesitation are "(1) whether it is feasible to create a workable cause of action; (2) whether extending the cause of action would undermine *Bivens*'s deterrence goals; (3) whether an extension of *Bivens* would impose asymmetric liability costs on privately operated facilities as compared to government-operated facilities; and (4) whether unique attributes of an area, like the military, give reason to infer that congressional inaction is deliberate." *Id.* at 8169 (citations omitted).

113 *Id.* at 8167-70 (citations omitted). The majority also found that allowing state law remedies to control in this area would be inappropriate because "many acts [constituting an Eighth Amendment violation] may not be covered by state tort law." *Id.* at 8170. As examples, the Ninth Circuit questioned whether an inmate would have a claim under state tort law for denial of access to a toilet, denial of outside exercise, or the denial of basic necessities such as socks, toilet paper, and soap. *Id.*

114 *Id.* at 8174-77.

115 *Id.* at 8178.

116 *Id.*

117 *Id.* at 8179 (Restani, J., dissenting).

118 *Id.*

119 *Id.* at 8183 (quoting Correctional Servs. Corp. v. Malesko, 534 U.S. 62, 69 (2001)).

120 *Id.* at 8184-85.

121 *Id.* at 8185.

122 *Id.* at 8187.

123 *Id.* at 8188.

124 *Id.* at 8190.

125 *Id.* at 8191.

126 Pollard v. GEO Group, Inc., No. 07-16112, slip op. at 19767 (9th Cir. Dec. 10, 2010) (order amending opinion and denying petition for rehearing en banc). (Bea, J., dissenting from denial of rehearing en banc). Judge Bea's dissent was joined by Chief Judge Kozinski and Judges O'Scannlain, Gould, Tallman, Callahan, Ikuta, and N.R. Smith.

127 *Id.* at 19680.

128 *Id.* at 19682.

129 *Id.* at 19684.

130 *Compare* Holly v. Scott, 434 F.3d 287, 292 (2006) ("The alleged actions of these defendants were not of a sufficiently federal character to create constitutional liability."), *with id.* at 297 (Motz, J., concurring in the judgment) ("[I]ndividual private correctional providers *are* government actors subject to liability as such.").

131 Brief for Petitioners at 37 n.8, Minneci v. Pollard, No. 10-1104 (U.S. 2011).

132 Brief for the United States as Amicus Curiae Supporting Petitioners at 13 n.6, Minneci v. Pollard, No. 10-1104 (U.S. 2011).

133 Brief for Respondent at 8, Minneci v. Pollard, No. 10-1104 (U.S. 2011)

134 Pollard v. GEO Group, Inc., No. 07-16112, slip op. at 8164 (9th Cir. June 7, 2010); *see also* Peoples v. CCA Detention Ctrs., 422 F.3d 1090, 1109 (10th Cir. 2005) (Ebel, J., dissenting).

135 *Peoples*, 422 F.3d at 1103; *see also* Alba v. Montford, 517 F.3d 1249, 1254-56 (11th Cir. 2008) (finding the alternative state remedies adequate

even where the prisoner would face considerable procedural difficulties in filing a state action); *Holly*, 434 F.3d at 296-97 (holding by both the majority and Judge Motz's concurrence that available state remedies defeat a *Bivens* action).

136 *See Pollard*, slip op. at 8164-66.

137 Mathis v. The GEO Group, Inc., No. 08-ct-00021 (E.D.N.C. filed May 28, 2008). Mr. Numbers represents The GEO Group Inc. in this litigation.

138 Mathis v. The GEO Group, Inc., No. 08-ct-00021, slip op at 7-13 (E.D.N.C. Sept. 29, 2010)

139 *See Pollard*, slip op. at 8157; *Holly*, 434 F.3d at 299 n.1 (Motz, J., concurring in judgment).

140 Correctional Servs. Corp. v. Malesko, 534 U.S. 62, 75 (2001) (Scalia, J., concurring).

141 In mid-2007, 30,379 out of 199,118 total federal prisoners (15.2%) were held in privately-operated prisons, and the percentage has been steadily increasing annually. William J. Sabol & Heather Couture, Bureau of Justice Statistics Bulletin: Prison Inmates at Midyear 2007, at 1, 5 (2008), *available at* http://bjs.ojp.usdoj.gov/content/ pub/pdf/pim07.pdf.

142 The Supreme Court has refused to extend *Bivens* to new circumstances in the seven opportunities it has had since 1980. Brief for Petitioners at 13, Minneci v. Pollard, No. 10-1104 (U.S. 2011).

Financial Services & E-Commerce

An Overview and Analysis of the Consumer Financial Protection Bureau

*By Sarah Riddell**

Overview

The Dodd-Frank Wall Street Reform and Consumer Protection Act (the "Act"), signed into law by President Obama on July 21, 2010, created a major overhaul of the financial industry.[1] For years, advocates have praised the benefits of financial reform and promoted legislation that would provide such reform. Specifically, these advocates have focused their support on consumer protection legislation.[2] The Act addresses many of these concerns by creating an entirely new regulatory regime with the purpose of "ensuring that all consumers have access to markets for consumer financial products and services and that markets for consumer financial products and services are fair, transparent, and competitive."[3] This paper describes the new regime and its powers and analyzes the effectiveness of the new bureau, which is still in its early stages.

The Consumer Financial Protection Bureau

The Act not only creates the Bureau of Consumer Financial Protection (the "CFPB") and gives it enforcement and regulatory authority, but it also transfers enforcement power from the Federal Trade Commission to the new agency. The CFPB is created as an independent executive branch agency, regulating consumer financial products and services under federal consumer financial laws.[4] A Director, appointed by the President and approved with the advice and consent of the U.S. Senate, serves for five years.[5] The Director is permitted to establish regional offices.[6] The CFPB is authorized to implement federal consumer financial laws by issuing rules, orders, interpretations, guidance, statements of policy, examinations, and enforcement actions.[7] Any rules or orders created by the CFPB are not subject to the review or approval of the Board of Governors of the Federal Reserve System, which is composed of seven appointees of the President.[8] However, the CFPB's proposed rules and regulations can be denied by the Financial Stability Oversight Council, a separate and distinct entity created by the Act consisting of ten voting members and five nonvoting members.[9]

The Act requires the CFPB to establish several mini-bureaus that focus on specific areas of consumer protection. For example, the Office of Fair Lending and Equal Opportunity provides "oversight and enforcement of Federal laws intended to ensure the fair, equitable, and nondiscriminatory access to credit for both individuals and communities that are enforced by the CFPB, including the Equal Credit Opportunity Act and the Home Mortgage Disclosure Act."[10] A new Office of Financial Education is responsible for developing and implementing a strategy to increase consumers' financial literacy.[11] The Office strategies, goals, and objectives include providing opportunities for access to financial counseling and mainstream financial institutions' services, such as savings and borrowing.[12] Moreover, the Office of Financial Education is responsible for providing consumers with methods to evaluate credit products and understand their credit scores and histories.[13]

The Act targets two sub-groups of particularly vulnerable citizens by creating the Office of Service Member Affairs[14] and the Office of Financial Protection for Older Americans.[15] The former is established to help educate service members and their families, enabling them to make informed decisions about consumer financial products and services. The latter has the goal, among others, of alerting seniors to financial advisers who possess unfair, deceptive, and abusive certifications.

Funding & Civil Penalties

The CFPB is funded with the Federal Reserve Board's earnings, but only as much as is "reasonably necessary to carry out the authorities of the CFPB."[16] Funding in 2011 cannot be more than 10% of the Federal Reserve System's total operating expenses, with gradual adjustments over the following two years.[17] Based on the Federal Reserve System's 2010 budget, which allocates $4,368,400,000 toward total operating expenses, the maximum amount of funding the CFPB would receive in 2011 is approximately $436.84 million.[18] Ultimately, the Act provides a cap on the CFPB's funding of 12% of the Federal Reserve System's total operating expenses.[19]

The Act directs the CFPB to collect civil penalties won against a person under the federal consumer financial laws and deposit the funds into a Consumer Financial Civil Penalty Fund.[20] The money in this account is to be distributed as payments to victims or for the CFPB's use for consumer education and financial literacy programs.[21]

Scope of Coverage

The Act applies to a variety of financial institutions, including non-depository institutions that provide loan origination, brokerage, or servicing for loans secured by real estate and obtained by consumers primarily for personal, family, or household purposes.[22] Larger participants of the consumer financial products or services market are covered.[23] Also covered are entities that have engaged in or are engaging in conduct that poses a risk to consumers, as well as those that offer private education loans or payday loans.[24] The Act expressly rejects from the CFPB's coverage various types of entities, including certain merchants, retailers, small businesses, real estate brokers, manufactured and modular home retailers, accountants, tax preparers, and lawyers.[25] Entities engaged in providing employee benefit and compensation plans are also excluded.[26] State-regulated entities are excluded to a limited extent.[27] Other exclusions are provided for entities regulated by the CFTC and Farm Credit Administration and those involved in charitable contribution activities.[28]

** Sarah Riddell is a recent graduate of Loyola University Chicago School of Law and currently works for Eurex in Chicago. Any views expressed in this article are hers alone and are not necessarily shared by her employer.*
The author would like to thank Professor Spencer Weber Waller and Laura Kotelman for their invaluable assistance with this article.

Covered entities are required to submit reports and are subject to periodic examinations in order to permit the CFPB to assess compliance, obtain information, and detect and assess risks to consumers and the consumer financial market.[29] Failure or flat-out refusal by covered entities to follow the Act's requirements is unlawful.[30] Covered entities are prohibited from offering or providing consumers any financial products or services not in conformity with federal consumer financial laws, or otherwise commit acts or omissions in violation of these laws.[31] They are also prohibited from engaging in any unfair, deceptive, or abusive acts or practices.[32]

Rulemaking and Enforcement Authority

The Act grants the CFPB with sweeping power to "administer, enforce, and otherwise implement the provisions of Federal consumer financial law."[33] The CFPB is authorized to issue rules, orders, and guidance on federal consumer financial law.[34] The CFPB has the responsibility to monitor for risks and developments in the consumer financial products or services market.[35] Although the Act grants the CFPB with the exclusive authority to make rules to regulate the consumer financial markets, the Financial Stability Oversight Council is permitted to set aside a final regulation if it believes that the regulation would threaten the "safety and soundness of the United States banking system or the stability of the financial system of the United States . . . ."[36] Thus, the Act appears to create a conflict by giving priority to the maintenance of the banking and financial systems over the CFPB's goal of protecting consumers. The political climate in which the new CFPB operates might be the key to the success of the CFPB.[37]

The CFPB also has limited authority to define unfair, deceptive, and abusive acts or practices. The Act expressly prohibits the CFPB from defining unfairness.[38] Abusive acts or practices are defined as those that

> (1) materially interfere with a consumer's ability to understand a term or condition of a consumer financial product or service or (2) take unreasonable advantage of a consumer's (a) lack of financial savvy, (b) inability to protect himself in the selection or use of consumer financial products or services, or (c) reasonable reliance on a covered entity to act in the consumer's interests.[39]

The Act gives the CFPB authority to investigate possible violations of federal consumer financial law, hold hearings, and commence civil litigation. The CFPB can issue cease-and-desist orders against covered entities that violate CFPB laws.[40] The CFPB gives notice to the entity about the violation and holds a hearing between thirty and sixty days after such notice, where it makes a decision about whether a violation occurred.[41] If the covered entity does not appear at the hearing, a presumption that the covered entity consents to the order is made.[42] A covered entity may appeal the CFPB's decision in federal court.[43]

The CFPB may also institute a civil action against an entity in violation of federal consumer financial law in order to impose a civil penalty or an injunction.[44] While no exemplary or punitive damages are available, many other types of relief are provided within the Act. For example, the Act includes the following types of relief: rescission of contracts, refund of money or return of real property, restitution, disgorgement for unjust enrichment, damages payments, costs of public notification of the violation, limits on the covered entity's activities or functions, and civil money penalties.[45] The civil money penalties are harsh, with the penalties categorized into three tiers. The first tier provides for a maximum penalty of $5,000 per day during which such violation or failure to pay continues.[46] The second tier provides for a maximum penalty of $25,000 for each day, and the third tier provides for a maximum $1 million penalty per day.[47] Mitigating factors may be considered when assessing the penalty, however. These factors include the size of financial resources of the covered entity, good faith, gravity of the violation or failure to pay, severity of consumers' risks or losses, and history of previous violations.[48]

The Act gives the CFPB the authority to take ancillary actions as they pertain to the CFPB's duties. For example, the CFPB is permitted to provide the Commissioner of Internal Revenue information, including the periodic reports or examinations, provided by covered entities, when tax law noncompliance is suspected.[49] The CFPB is also permitted to give evidence of federal criminal law violations to the U.S. Attorney General.[50] The CFPB can also restrict or prohibit mandatory pre-dispute arbitration agreements between covered entities and consumers.[51]

Post-Transfer Date Analysis

The "transfer date," the date on which the various consumer protection laws are transferred from other agencies to the Consumer Financial Protection Bureau and when the CFPB can exercise new authorities, arrived on July 21, 2011. In the months following the transfer date, many challenges still loom large before the CFPB. First, there is still no Senate-confirmed Director, and Republicans refuse to confirm any such Director until the CFPB's structure and its funding are changed. Second, the CFPB must negotiate its turf with the Federal Trade Commission ("FTC"), which is itself threatened by the removal of its jurisdiction over consumer financial policy and enforcement. Third, businesses covered under the CFPB may find implementing vague rules, such as the "unfair, deceptive, or abusive acts or practices" rule, to be difficult. With all of these challenges, the question remains: Will the CFPB become a toothless agency weakened by the current state of politics, or will it rise to the occasion and, to Professor Elizabeth Warren's vision, protect ordinary consumers from risky financial products and services that threaten the American Dream?

Recent Developments at the CFPB

Since the Dodd-Frank Act was passed, the CFPB has hired over 400 staff, yet it still remains without a Senate-confirmed Director.[52] Professor Warren was passed over for Director of the CFPB, and instead President Obama nominated Richard Cordray, former Ohio Attorney General and previous head of the CFPB's enforcement division, on July 18, 2011. As of the time this article was submitted, the Senate had not confirmed his appointment.[53] Even without a Director, the CFPB has

started fulfilling its obligations. Today, it is working to create a single, simple mortgage disclosure form that allows consumers to comparison shop when obtaining a mortgage (a combined RESPA/TILA form).[54] Under its "Know Before You Owe" project, the CFPB is testing two potential forms that consumers would receive upon applying for a mortgage loan.[55]

The CFPB is also working to define its "larger participant" rule, which must be defined by July 21, 2012. The CFPB has supervisory authority over nondepository businesses, including those in the payday lending, private education lending, and residential mortgage markets. The new agency also has supervisory authority over other markets that provide consumer financial products or services, but only over the larger participants of those markets. The definition of "larger participant" will thus create a broader supervisory role for the CFPB, and greater compliance for those companies that fall within the larger participant definition. The CFPB is currently seeking public comment on various aspects of this rule, including the primary consideration of specific markets that should be covered by the rule. The CFPB has proposed inclusion of six markets in its initial definition, which incorporates debt collection; consumer credit and related activities; prepaid cards; debt relief services; consumer reporting; and money transmitting, check cashing, and related activities.[56]

Once the markets are defined, the CFPB will seek comment on the appropriate way to measure the threshold for the "larger participants" within those markets.[57] The CFPB proposed several methods of calculating a larger participant, but wants feedback on whether to use just one or a combination of several criteria in the calculation.[58] The threshold measurement may be tailored to each specific market.[59] The CFPB is considering an absolute approach, which would dictate a larger participant to be one with an annual loan volume of a specific dollar amount.[60] Another consideration is using a relative approach based on market share or some other calculation that compares the market participant to others in the market.[61]

Challenges the CFPB Faces

As the CFPB moves ahead with these projects, Republicans object to its institutional design and power. House Republicans have introduced several bills that would substantially alter the way the CFPB operates.[62] Senator Shelby says that any Director is "dead on arrival" and will not be confirmed by the Senate until President Obama comes to the negotiating table to discuss the reform found in the House bills.[63]

The three House bills would dramatically alter the structure of the CFPB and the veto procedure for new regulations proposed by the CFPB. For example, one bill proposes to move the CFPB from the Federal Reserve to the Department of the Treasury.[64] Another brings the CFPB into the regular congressional appropriations process.[65]

Representative Duffy introduced H.R. 1315, or the "Consumer Financial Protection Safety and Soundness Improvement Act of 2011." It allows for a simple majority of the Financial Stability Oversight Council instead of a two-thirds vote to veto the CFPB's proposed rules and regulations.[66] A recent amendment to H.R. 1315 required that the two-thirds majority vote be restored, but this amendment failed to pass.[67] A different amendment that was successfully passed eliminates any potential conflicts of interest by prohibiting members of the Financial Stability Oversight Council from voting on a proposed regulation if that regulation would affect an institution at which the member was employed in the preceding two years.[68]

The Duffy bill would also change the language in the Dodd-Frank Act that permits the Financial Stability Oversight Council to set aside a final CFPB regulation if it believes that the regulation would threaten the "safety and soundness of the United States banking system or the stability of the financial system of the United States . . . ."[69] Rep. Duffy's legislation replaces "may" with "shall," thus requiring the Financial Stability Oversight Council to intervene when a CFPB regulation is inconsistent with the safe and sound operations of United States financial institutions.[70] The legislation also replaces "regulation or provision would put the safety and soundness of the United States banking system or the stability of the financial system of the United States at risk" with "regulation which is the *subject of the petition* is *inconsistent* with the safe and sound operations of United States financial institutions."[71] The change of language begs the question: What types of regulations are inconsistent with the operation of our financial institutions? Those that increase costs and risks for banks and decrease costs and risks for consumers? If recent history is any indication of how U.S. financial institutions operate "safely and soundly" in a political environment where special interests thrive and banks are "bailed out," then this legislation would greatly reduce the CFPB's intervention power as set out under the Act.

The "Responsible Consumer Financial Protection Regulations Act of 2011," introduced by Representative Bachus, establishes a five-member commission to head the bureau, and the Vice Chairman for Supervision of the Federal Reserve System must be one of the five members.[72] Each member serves staggered five-year terms.[73] An interesting aspect of this proposed legislation is that one commissioner has the special responsibility for the oversight of the CFPB's consumer protection activities, specifically focusing on protecting minorities, older citizens, youth and veteran consumers from unfair, deceptive, and abusive lending practices.[74] The legislation requires the commissioners to coordinate with state enforcers.[75]

A Turf Battle with the FTC?

The FTC will likely lose funding for the employees who have moved to the CFPB and for its financial programs and enforcement activities involving consumer finance schemes. Former FTC Commissioner William Kovacic is concerned that the CFPB will actually diminish current consumer financial protection policies by overseeing functions typically performed by the FTC but without the FTC's institutional design.[76] Specifically, the FTC's policy perspective, carefully crafted after insight from the FTC's Bureau of Economics and Bureau of Competition, is a unique component of the FTC, making it more than a mere enforcement agency.[77] Kovacic believes that the FTC already conducts important research and provides educational programs in an independent manner that may not carry over to the CFPB or, if it does, it will be inferior to the FTC's pre-existing structure.[78] Kovacic questions why the FTC must abandon its consumer protection functions

and transfer all of these important roles to the CFPB when states will continue to enforce consumer protection laws.[79] In fact, the federal consumer financial protection law does not nullify or exempt people from complying with state law unless the state law is inconsistent with the Act.[80] State laws that provide greater protection to consumers than the Act are not "inconsistent" with the Act and must be adhered to.[81] When a majority of states enact a resolution supporting the establishment or modification of a CFPB regulation, the CFPB will propose a rule in response to the state action.[82]

Another major concern is that the Dodd-Frank Act defines the CFPB's consumer protection functions so broadly as to overlap and threaten the FTC's seemingly non-financial research and enforcement responsibilities, such as telemarketing fraud.[83] The Memorandum of Understanding due six months from July 21, 2011 will clarify jurisdiction, but many grey areas may crop up unexpectedly if the battle lines are not clearly drawn. For example: will the FTC retain enforcement jurisdiction over violations of advertising rules when the violators are banks and other "larger participants," or will the CFPB be responsible for enforcing those rules under the unfair, deceptive, or abusive acts or practices language found in the Dodd-Frank Act?

Vague Rules Will Impact Financial Product Providers

The expanded FTC § 5 language found in the Dodd-Frank Act covers abusive acts or practices along with the standard unfair or deceptive acts or practices. What exactly is an abusive act or practice? The Dodd-Frank Act defines it, but not very clearly. An abusive act or practice is one that causes a consumer to fail to understand the financial product or service's terms or takes "unreasonable advantage" of a consumer's lack of understanding or inability to protect his own interests.[84] "Enhanced amorphousness" of this language will cause a much higher risk for the consumer financial services industry, especially considering that a dedicated "cop on the beat" will enforce this vague rule.[85] In fact, the CFPB can enforce the rule and investigate, hold hearings, litigate and seek remedies, including substantial civil penalties of up to $1 million per day.[86] For a compliance officer at a bank or one of the as-of-yet undefined "larger participants," such vagueness can threaten conformity with the rules; clearer rules are necessary to avoid these large penalties. Another chilling prospect for covered entities is the fact that a single credit disclosure violation could potentially lead to liability under the FTC Act, states' Little FTC Acts, TILA, and the new unfair, deceptive, or abusive acts or practices language the CFPB can enforce.[87] Look for enforcement actions to clarify these rules, and for state enforcers and plaintiffs' attorneys to follow the CFPB's lead in these actions.[88]

Litigation will also clarify the extent to which the Dodd-Frank Act alters federal preemption of state consumer financial laws. Although somewhat peripheral to the CFPB, because the Office of the Comptroller of the Currency ("OCC") retains authority to issue preemption regulations, orders, and determinations on a case-by-case basis, the Dodd-Frank Act mandates that the OCC consult with the CFPB before making a preemption determination.[89] The Dodd-Frank Act codifies a Supreme Court case and empowers state enforcers to bring lawsuits against national banks that are not in compliance with non-preempted state laws.[90] The changes in federal preemption standards, similar to the change in the "unfair or deceptive acts and practices" language, remain somewhat vague; future cases will help illuminate the contours of the law. The way in which the OCC consults with the CFPB on future determinations will be interesting and may give rise to another turf battle with the OCC.

Going Forward

As the CFPB sets up shop and hires more people, it will be that much more difficult to dismantle or change the structure of the new agency. The continuing uncertainty over the CFPB's relationship with the FTC poses a threat to the CFPB's jurisdiction. The FTC and CFPB have six months from July 21, 2011 to negotiate an agreement on areas over which each will possess jurisdiction. Things may be clearer in November, when the ABA Antitrust Section hosts its Fall Forum and presents a panel on how the FTC and CFPB will engage with industries and coordinate their enforcement and policymaking efforts.[91] The panel will address enforcement priorities in areas that focus on consumers, including privacy, marketing, and the internet. President Obama has a lot on his plate (a health care bill in jeopardy, a new jobs bill criticized by Congressional Democrats, etc.). Will he have time and political capital to ensure that the CFPB is established as intended in the Dodd-Frank Act?[92] Until its first Director is confirmed, the CFPB will lack the ability to command the respect and wield the power that the Dodd-Frank Act intended.

Endnotes

1 Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, Pub. L. 100-203, 124 Stat. 1376 (2010) [hereinafter "Act"].

2 *See, e.g.,* Oren Bar-Gill & Elizabeth Warren, *Making Credit Safer*, 157 U. Pa. L. Rev. 1 (2008).

3 Act § 1021(a).

4 *Id.* § 1011(a).

5 *Id.* §§ 1011(b)(1), (2).

6 *Id.* §§ 1011(c)(1); 1011(e).

7 *Id.* § 1012(a)(10)

8 Board of Governors of the Federal Reserve System, About the Fed, http://www. federalreserve.gov/aboutthefed/bios/board/default.htm (last visited Nov. 17, 2011). The board members serve a fourteen-year term; current board members include Ben Bernanke, Janet Yellen, Kevin Warsh, Elizabeth Duke, Daniel Tarullo, and Sarah Bloom Raskin. Act § 1012(c)(3). In fact, the Board of Governors may not:

> (A) intervene in any matter or proceeding before the Director, including examinations or enforcement actions, unless otherwise specifically provided by law;
>
> (B) appoint, direct, or remove any officer or employee of the Bureau; or
>
> (C) merge or consolidate the Bureau, or any of the functions or responsibilities of the Bureau, with any division or office of the Board of Governors or the Federal reserve banks.

Act §1012(c)(2).

9 Act §§ 111(a); 111(b). Voting members include the Secretary of the

Treasury, Chairman of the Board of Governors of the Federal Reserve System, Comptroller of the Currency, Director of the Bureau, Chairman of the Commission, Chairperson of the Corporation, Chairperson of the Commodity Futures Trading Commission, Director of the Federal Housing Finance Agency, Chairman of the National Credit Union Administration Board, and an independent member with insurance experience appointed by the President and confirmed by the Senate. *Id.* Nonvoting members serve in an advisory capacity and include the Director of the Office of Financial Research, Director of the Federal Insurance Office, a state insurance commissioner, a state banking supervisor, a state securities commissioner. *Id.* The latter three members are to be designated by a selection process determined by their peers. *Id.*

10 *Id.* § 1013(c)(1)-(2)(A).

11 *Id.* § 1013(d)(1)-(2).

12 *Id.* § 1013(d)(2).

13 *Id.*

14 *Id.* § 1013(e)(1). The Office of Service Member Affairs has authority to develop and implement initiatives intended to "educate and empower service members and their families to make better informed decisions regarding consumer financial products and services." *Id.* § 1013(e)(1)(A). Moreover, this office must monitor and respond to complaints made by service members and their families. *Id.* § 1013(e)(1)(B). The Bureau has authority to assign this Office's functions and the maintenance of these functions to regional offices near military bases. *Id.* § 1013(e)(2)(A).

15 *Id.* § 1013(g)(1). The duties of this office include developing goals for programs that offer financial literacy and counseling to seniors; monitoring the certification of financial advisors who advise seniors and alert the Bureau and State Regulators about unfair, deceptive, or abusive certifications; recommending best practices for seniors to learn about financial advisors; conducting research to learn these best practices; coordinating with other agencies; and working with community and non-profit organizations. *Id.* § 1013(g)(3)(A)-(F).

16 *Id.* § 1017(a)(1).

17 *Id.*

18 *See* Bd. of Governors of the Fed. Reserve Sys., Annual Report: Budget Review 2010, at 9 (2010), *available at* http://www.federalreserve.gov/boarddocs/rptcongress/budgetrev/br10.pdf.

19 Act § 1017(a)(2).

20 *Id.* § 1017(d)(1).

21 *Id.* § 1017(d)(2).

22 *Id.* § 1024(a)(1)(A)-(E). The Bureau also has supervisory power over very large insured depository institutions or credit unions (those with total assets of more than $10 billion) and their affiliates, as well as insured depository institutions or credit unions with total assets of $10 billion or less. *Id.* §§ 1025(a), 1026(a).

23 *Id.* § 1024(a)(1)(A)-(E).

24 *Id.*

25 *Id.* § 1027. However, the carve-out is inapplicable to those entities that (i) extend credit to a consumer to enable the consumer to purchase a nonconsumer financial product or service, (ii) collect debt, or (iii) sell or convey debt.

Id. § 1027(a)(2)(A)(i)-(iii). There is also an exclusion for certain auto dealers. *Id.* § 1029.

26 *Id.*

27 *Id.* § 1027. For example, entities regulated by a state securities commission or a state insurance regulator are excluded. *Id.*

28 *Id.*

29 *Id.* § 1024(b)(1)(A)-(C).

30 *Id.* § 1036(a)(2).

31 *Id.* § 1036(a)(1).

32 *Id.*

33 *Id.* § 1022(a).

34 *Id.* §§ 1022(b)(1), 1022(b)(4)(A).

35 *Id.* § 1022(c).

36 *Id.* § 1023(a). An agency must first petition the Council to stay or set aside a regulation; the Council Chairperson "may stay the effectiveness of a regulation for the purpose of allowing appropriate consideration of the petition by the Council." *Id.* § 1023(c)(1)(A). The Council may set aside or issue a stay of any regulation upon a vote of the Council and approval by two-thirds of the then-serving Council members. *Id.* § 1023(c)(3)(A).

37 Mark E. Budnitz, *The Development of Consumer Protection Law, the Institutionalization of Consumerism, and Future Prospects and Perils*, 26 Ga. St. U. L. Rev. 1147, 1188 (2010) ("The robustness of enforcement efforts by government agencies such as the FTC, bank agencies and state attorneys general depends upon the political climate in which those agencies operate, the resources they receive, and other priorities needing a share of those resources.").

38 Act § 1031(a).

39 *Id.* § 1031(b).

40 *Id.* § 1053(b).

41 *Id.*

42 *Id.* §1053(b)(1)(C).

43 *Id.* § 1053(b)(3).

44 *Id.* § 1054(a).

45 *Id.* § 1054(a)(2).

46 *Id.* § 1054(c)(2)(A).

47 *Id.* § 1054(c)(2)(B)-(C).

48 *Id.* § 1054(c)(3).

49 *Id.* §§ 1024(b)(6), 1025(b)(5), 1026(b)(3).

50 *Id.* § 1056.

51 *Id.* § 1028(a)-(b). This provision permits the prohibition or imposition of conditions on mandatory pre-dispute arbitration agreements after the Bureau conducts a study of the use of such agreements and concludes that such actions are in the public interest and protect consumers. *Id.*

52 Consumer Financial Protection Bureau, Building the CFPB: A Progress Report 24 (2011), *available at* http://www.consumerfinance.gov/wp-content/uploads/2011/07/Report_BuildingTheCfpb1.pdf.

53 The White House Blog, http://www.whitehouse.gov/blog/2011/07/18/president-obama-nominates-richard-cordray-lead-consumer-financial-protection-bureau (July 18, 2011, 15:55 EDT).

54 Consumer Financial Protection Bureau, *supra* note 52, at 10.

55 *Id.* at 11.

56 *Id.* at 13.

57 Defining Larger Participants in Certain Consumer Financial Products and Services Markets, 76 Fed. Reg. 38,059, 38,060 (June 29, 2011) (to be codified at 12 C.F.R. ch. X).

58 *Id.* The CFPB seeks feedback on whether one or a combination of the following factors should be used to determine whether a market participant is a "larger participant": number of transactions in the market, annual value of transactions, annual receipts or revenues, geographic coverage (number of states the participant is located), asset size, and outstanding loan balances. *Id.*

59 *Id.*

60 *Id.*

61 *Id.*

62 H.R. 1121, 112th Cong. (2011); H.R. 1315, 112th Cong. (2011).

63 Reid J. Epstein, *Richard Shelby: Cordray is DOA*, Politico, July 21, 2011, *available at* http://www.politico.com/news/ stories/0711/59545.html.

64 Consumer Financial Protection Oversight Act of 2011, H.R. 557, 112th

Cong. (2011).

65 H.R. 1640, 112th Cong. (2011).

66 Consumer Financial Protection Safety and Soundness Improvement Act of 2011, H.R. 1315, 112th Cong. §2 (2011).

67 H.AMDT. 687 (amending H.R. 1315), 112th Cong. (2011) (Congresswoman Sheila Jackson Lee offered the amendment on July 21, 2011, and it failed by recorded vote (170-239), roll no. 615.).

68 H.AMDT. 688 (amending H.R. 1315), 112th Cong. (2011).

69 Act § 1023(a).

70 H.R. 1315, 112th Cong. § 103(A) (2011). This legislation passed in the House, but will likely fail in the Democrat-controlled Senate. *See* Seung Min Kim, *House Bill Revamps Consumer Agency*, Politico, July 21, 2011, *available at* http:// www.politico.com/news/stories/0711/59624.html.

71 H.R. 1315 §103(B) (emphasis added).

72 H.R. 1121, 112th Cong. § 2 (2011).

73 H.R. 1121 § 2.

74 *Id.*

75 *Id.*

76 William E. Kovacic, *The Consumer Financial Protection Agency and the Hazards of Regulatory Restructuring*, 1 Lombard Street 12, 20 (2009).

77 *Id.* Kovacic points out that the FTC, as an enforcement agency, brought seventy consumer protection cases in the five years preceding the article's writing. *Id.* at 22.

78 *Id.* at 21.

79 *Id.* at 23-24; *see also* note 75 and accompanying text.

80 Act § 1041(a)(1).

81 *Id.* § 1041(a)(2).

82 *Id.* § 1041(c). In proposing a rule in response to the state action, the CFPB will consider the following factors: potential greater protections to consumers offered by the new regulation and the benefits of the regulation weighed against the increased costs or inconveniences to consumers. *Id.* The CFPB will also consider whether unfair discrimination against a class of consumers would result. *Id.*

83 Kovacic, *supra* note 76, at 24.

84 Act § 1023(a).

85 Martin Bishop, *Regulatory: Unfair, Deceptive or Abusive Acts or Practices. Amorphous New Statutory Provisions Create Serious Compliance Risks.*, Inside Counsel, July 27, 2011, *available at* http://www.insidecounsel.com/2011/07/27/regulatory-unfair-deceptive-or-abusive-acts-or-pra.

86 *Id.*

87 Martin Bishop, *Regulatory: Unfair, Deceptive, or Abusive Acts or Practices—Part II. Symbiotic Relationships Among State and Federal Laws, Regulators and Plaintiffs' Bar Creates Difficult Compliance Environment.,* Inside Counsel, Aug. 10, 2011, *available at* http://www.insidecounsel.com/2011/08/10/regulatory-unfair-deceptive-or-abusive-acts-or-pra.

88 *Id.*

89 Act § 1044.

90 Cuomo v. The Clearing Housing Ass'n, 129 S. Ct. 2710 (2009).

91 Julie Brill, Commissioner, Fed. Trade Comm'n, and Peggy Twohig, Director of the Office of Consumer Protection at the Dept. of Treasury and Policy Lead for the CFPB, Panel at the ABA Antitrust Section Fall Forum: Working with the FTC and CFPB (forthcoming Nov. 17, 2011).

92 Editorial, *Consumers vs. the Banks*, N.Y. Times, July 24, 2011, at A20 (Cordray will need Obama's support to win confirmation; the President's "decision to jettison Ms. Warren is not a reassuring sign.").

Free Speech & Election Law

Illuminating *Citizens United*: What the Decision Really Did

*By William R. Maurer**

In January 2010, the U.S. Supreme Court issued one of its most controversial decisions in decades, *Citizens United v. FEC*.[1] The response among politicians supporting restrictions on campaign finances was immediate and fierce. President Obama said he could not "think of anything more devastating to the public interest" and criticized it during the State of the Union address with members of the Court present.[2] Senator Al Franken called it "an incredible act of judicial activism,"[3] while Rep. Chris Van Hollen called it "a very, very sad day for American democracy," and a "radical, radical decision."[4]

Politicians were not the only ones to denounce *Citizens United*. One law professor compared it to *Plessy v. Ferguson*[5] and *Dred Scott*.[6] A *Huffington Post* writer compared the five Justices in the majority to concentration camp prisoners who cooperated with the Nazis and called the beneficiaries of the decision "vampires" who treat humans "as sources of profit, with zero consideration for their humanity."[7]

The decision remains a sore spot for many. A cable-TV-talk-show-host for the cable channel MSNBC, Dylan Ratigan, is attempting to lead an effort to amend the U.S. Constitution to reverse *Citizens United*,[8] while a recent "Occupy DC" event concentrated on undoing the decision.[9]

Many of the assumptions underlying this opposition are simply incorrect, however. If the arguments employed against *Citizens United* are any indication, the opponents' positions are based on an erroneous understanding of the American constitutional system and a fundamental misreading of the First Amendment itself. Indeed, the most common critiques of *Citizens United* are based on beliefs about what the decision did—recognizing corporate personhood and ignoring that the Founders never meant to "give" free speech rights to corporations—that are either entirely false or, at the least, reflect a serious misunderstanding of American government. Read correctly, with an accurate understanding of history and Supreme Court precedent, *Citizens United* is a decision consistent with both the words and intent of the First Amendment.

What Did *Citizens United* Actually Say?

Citizens United concerned a provision in the U.S. Code, Section 441b of Title 2, that made it a crime for corporations and unions to use general treasury money to make "independent expenditures" (that is, spending that is not coordinated with candidates) that expressly advocated the election or defeat of a federal candidate.[10] Prior to *Citizens United*, corporations and unions could only participate in the political process by creating separate political action committees (PACs). PACs operate under complex and expensive administrative requirements, however, and these associations could not use general treasury funds for political purposes, so this was an "alternative" of which very few corporations availed themselves.[11]

Citizens United is a nonprofit corporation that wished to use its general treasury funds to distribute a film about Hillary Clinton—then a candidate for the Democratic Party's nomination for President in 2008—via video-on-demand. Citizens United sued the Federal Election Commission to enjoin Section 441b's application to their distribution of the film. Citizens United lost at the trial court and then sought review at the U.S. Supreme Court, which took up the case in 2009.

In an unusual move, the U.S. Supreme Court held oral argument twice in the case. In the first argument, the U.S. Solicitor General's office admitted that "a corporation could be barred from using its general treasury fund to publish [a] book . . . ."[12] In other words, the position of the government was that, if a group of citizens pooled their money in a corporate form, the government could fine or imprison them if they published a book, or made a film, about politics. During the second oral argument, then-Solicitor General Elena Kagan attempted to back away from this statement, saying that the FEC had never applied the provision to a book, to which Chief Justice Roberts responded, "But . . . we don't put our First Amendment rights in the hands of FEC bureaucrats . . . ."[13]

In January 2010, a five-Justice majority struck down Section 441b. The Court stated unequivocally that the First Amendment restricts the ability of the government to abridge the freedom of speech of corporations. The Court found that Section 441b was an outright ban on speech and that the PAC alternative was not a real alternative for corporations because PACs are separate associations and expensive and difficult to establish and administer.

The Court also noted that the government's reasoning would also allow it to ban media publications, but that it had so far exempted media corporations from the law's broad reach. The Court rejected the government's proffered justifications for the law. It overturned two relatively-recent decisions, *Austin v. Michigan Chamber of Commerce*[14] and portions of *McConnell v. FEC*,[15] which held that the government may ban the independent expenditures of corporate and union entities.

Justice Stevens, joined by three other Justices, filed a lengthy dissent, arguing that Congress could constitutionally make it a felony for corporations and unions to pay for political advertisements using money from their general treasury.

What *Citizens United* Did Not Say

Many critics of the decision argue that *Citizens United* hinge on the assumption that the decision granted corporations the same constitutional rights as individuals and that this grant of rights was incorrect because the First Amendment only applies to individuals. For instance, in one of its criticisms of

....

** Institute for Justice. A previous version of this article was published in* The Weekly Standard*, Vol. 16, No. 11. The article is republished here with permission.*

Citizens United, *The New York Times* (ironically, a for-profit corporation that routinely uses its general treasury funds to expressly advocate for the election or defeat of federal candidates) said the following: "Most wrongheaded of all is its insistence that corporations are just like people and entitled to the same First Amendment rights."[16] Similarly, Rep. Van Hollen stated that *Citizens United* "is a decision that equates, for the purposes of expending monies in elections, says [sic] that corporations equal individuals. I think it is an un-American decision . . . ."[17] Justice Stevens also accepted this argument, arguing that the "speech" referred to in the First Amendment only applies to "oral communications by individuals," and that because corporations are "artificial entities," they "do not have the technical capacity to speak."[18]

These criticisms misunderstand the American constitutional order and the purpose of the First Amendment itself. The critics are correct, of course, that the First Amendment does not say anything about corporations having free speech rights. This is because it does not say anything about which individuals and groups have free speech rights. The First Amendment provides: "Congress shall make no law . . . abridging the freedom of speech." It does not say, "Only persons have the right to free speech." It does not say, "Corporations do not have free speech rights," nor does it say, "Congress shall make no law abridging the freedom of speech of individuals." It does not say, "Congress shall make no law abridging freedom of speech, except for two or more people." It does not say, "Congress shall make no law abridging freedom of speech, except when the speaker is capable of amassing immense aggregations of wealth that are accumulated with the help of the corporate form and that have little or no correlation to the public's support for the corporation's political ideas."[19]

The critics of *Citizens United* thus forget an important fact about the First Amendment: it is not a grant of rights. Instead, the First Amendment is a restriction on government power.[20] The First Amendment restricts the ability of government to restrict the rights listed in the amendment—it certainly does not create a limitation on those rights so that they apply only to individuals acting by themselves. It does not lay out who does and does not have the right of free speech because it assumes every American, acting collectively or individually, does.

Instead, the First Amendment defines what legislation Congress can pass that affects this right, which is none. The Founders believed the rights guaranteed by the Bill of Rights are inherent in the American people and that these rights prevail whether they act independently or in concert with one another. The Bill of Rights is an explicit recognition that Congress cannot interfere with these inherent rights. It was not a positive grant of privileges, but a restriction on government.

It is difficult to believe that the critics of *Citizens United* really believe that only individuals may exercise constitutional rights. Taken to its logical conclusion, the belief that "only individuals have constitutional rights" would have serious consequences for American liberty and would reduce the U.S. to little more than a legislative dictatorship. If only individuals are protected by the Bill of Rights, can the government seize Apple's intellectual property without paying for it, regardless of the Fifth Amendment? Can the government quarter troops at the AFL-CIO's headquarters, despite the Third Amendment? Can it search the ACLU's offices without a warrant because the Fourth Amendment does not apply? Why would any corporation continue to operate in the United States if the fundamental protections that have made America a free and prosperous nation can be ignored by the government? Could the government destroy organized labor by means commonly viewed as forbidden by the Bill of Rights?

In his majority opinion, Justice Kennedy also recognized the implications of a holding that Congress could ban speech by groups of individuals acting in concert: it would allow the government to ban speech by newspapers, magazines, television news organizations, and other members of the press.[21] If the government may ban speech because it is produced by corporations, unduly influential, and drowns out other speakers, why should media corporations—which are sometimes owned by some of the largest multi-national corporations in the world—be immune? Dylan Ratigan, who works for a cable network owned by General Electric, seems to assume that his speech and that of his employer would be immune from a constitutional amendment banning corporate political speech, but it is not clear why this would be so.

Corporations and unions are not individuals, but they are comprised of individuals that have banded together for common purposes. Marriages, partnerships, neighborhood organizations, and rock groups are all also not individuals, but are rather associations of individuals that have decided that acting cooperatively is more effective than acting alone. To hold that the First Amendment rights (or any other constitutional rights) dissipate the minute one person begins to act in concert with another would neuter the Bill of Rights as an effective check on unrestrained government power. As Chief Justice Roberts put it in his concurrence in *Citizens United*, "The First Amendment protects more than just the individual on a soapbox and the lonely pamphleteer."[22]

Conclusion

Many critics of *Citizens United* do not realize that this decision was the third time the Supreme Court considered the constitutionality of this law. The first two times the Supreme Court considered the ban, however, it went out of its way to avoid reaching the issue of its constitutionality. Justice Douglas, joined by Chief Justice Warren and Justice Black, filed a dissent that argued that not only should the Court hear the case, it should strike down the ban as an obvious violation of the First Amendment:

> Some may think that one group or another should not express its views because it is too powerful, because it advocates unpopular ideas, or because it has a record of lawless action. But these are not justifications for withholding First Amendment rights from any group—labor or corporate. First Amendment rights are part of the heritage of all persons and groups in this country. They are not to be dispensed with or withheld because we or Congress thinks the person or group is unworthy.

Justice Douglas concluded by calling the ban "a broadside

assault on the freedom of political expression guaranteed by the First Amendment."[23]

The critics of *Citizens United* too often ignore what the case actually said and disregard the meaning and intent of the First Amendment. Justice Kennedy and Justice Douglas recognized that the right of free speech is not a privilege dispensed by the Court or the government, but an inherent right that the First Amendment protects from government action. When viewed correctly, *Citizens United* was perfectly consistent with the wording, spirit, and intent of the First Amendment.

Endnotes

1 130 S. Ct. 876 (2010).

2 *Quoted in* Darlene Superville, *Obama Weekly Address VIDEO: President Blasts Supreme Court over* Citizens United *Decision*, Associated Press, Jan. 23, 2010, *available at* http://www.huffingtonpost.com/2010/01/23/obama-weekly-address-vide_n_434082.html.

3 Press Release, Charles E. Schumer, U.S. Senator, Senate Democrats Unveil Legislation to Limit Fallout from Supreme Court Ruling that Allows Unlimited Special-Interest Spending on Elections—Announce Plan for Senate Passage by July 4 (Apr. 29, 2010), *available at* http://schumer.senate.gov/new_website/record.cfm?id=324343.

4 Press Release, Chris Van Hollen, Member, U.S. House of Representatives, Van Hollen Remarks on Supreme Court Ruling in *Citizens United* Case (Jan. 21. 2010), *available at* http://vanhollen.house.gov/News/DocumentSingle.aspx?DocumentID=167326.

5 163 U.S. 537 (1896).

6 Scott v. Sanford, 60 U.S. 393 (1857).

7 Rob Kall, *Real Vampires, Their Human Cattle and Supreme Court Capos: Time to Declare War on Corporate Personhood*, Huffington Post, Jan. 25, 2010, http://www.huffingtonpost.com/rob-kall/real-vampires-their-human_b_434612.html.

8 Get Money Out, http://www.getmoneyout.com/ (last visited Nov. 14, 2011).

9 Arin Greenwood, *OccupyDC Protest: Group Pushes Repeal of* Citizens United*, Corporate Personhood*, Huffington Post, Oct. 1, 2011, http://www.huffingtonpost.com/2011/10/01/occupydc-citizens-united-corporate-personhood_n_989690.html.

10 2 U.S.C. 441b.

11 2 U.S.C. § 441b(b)(2).

12 Transcript of Oral Argument at 20, Citizens United v. FEC, 130 S. Ct. 876 (2010), *available at* http://www.supremecourt.gov/oral_arguments/argument_transcripts/08-205.pdf.

13 Transcript of Oral Argument at 65-66, Citizens United v. FEC, 130 S. Ct. 876 (No. 08-205), *available at* http://www.supremecourt.gov/oral_arguments/argument_transcripts/08-205%5D.pdf.

14 494 U.S. 652 (1990).

15 540 U.S. 93 (2003).

16 Editorial, *The Court's Blow to Democracy*, N.Y. Times, Jan. 21, 2010, *available at* http://www.nytimes.com/2010/01/22/opinion/22fri1.html?_r=1.

17 Van Hollen, *supra* note 4.

18 *Citizens United*, 130 S. Ct. 876, 950 n.55 (Stevens, J., dissenting).

19 Austin v. Mich. Chamber of Commerce, 494 U.S. 652, 660 (1990).

20 *See* Barron v. Baltimore, 32 U.S. 243, 250 (1833) (The amendments "demanded security against the apprehended encroachments of the general government . . . .").

21 *Citizens United*, 130 S. Ct. at 905-07.

22 *Id.* at 917 (Roberts, C.J., concurring).

23 United States v. UAW-CIO, 352 U.S. 567, 597 (1957) (Douglas, J., dissenting). In an earlier case, Justice Rutledge, joined by Justices Black, Douglas, and Murphy, also dissented from the Court's refusal to consider the ban. Justice Rutledge argued, "A statute which, in the claimed interest of free and honest elections, curtails the very freedoms that make possible exercise of the franchise by an informed and thinking electorate, and does this by indiscriminate blanketing of every expenditure made in connection with an election, serving as a prior restraint upon expression not in fact forbidden as well as upon what is, cannot be squared with the First Amendment." United States v. CIO, 335 U.S. 106, 155 (1948) (Rutledge, J., dissenting).

Intellectual Property

Hot News: The "Hot-News" Doctrine Is Hot Again! Or Is It?

*By David L. Applegate and Ryan Schermerhorn**

Copyright law, like contract law, is deceptively complex. Just as the familiar elements of offer, acceptance, and consideration can give rise to endless disputes not easily resolved, the seemingly simple notion that an author has a time-limited monopoly on rights to a particular work of authorship gives rise to many questions. Given constantly-changing technology for fixing works of original expression in tangible media, ever-evolving means of copying and piracy, and repeated revisions to U.S. copyright law, both the courts and creators have had a difficult time understanding core concepts and keeping up with how the law is applied. Nowhere is this difficulty more apparent than in the Second Circuit's attempts to keep the law of copyright straight in the area of "hot news."

A fundamental concept of United States copyright law today is that although original expressions of ideas fixed in tangible media are copyrightable, both the ideas and the facts themselves are not.[1] Thus one may copyright a play, a song, a telephone directory, or a map, but one may not copyright words themselves, musical notes, the names in the phone book, or the jurisdictional boundaries or rivers shown on the map, each of which anyone is free to use. (The express purpose of copyright law, after all, as embodied in the Constitution's Copyright Clause, is to promote the progress of "Science," meaning knowledge.[2])

In the same way, a newspaper, website, or blog can copyright a story (its particular original expression recounting the facts of, and opinion regarding, an event), but it can never copyright the events depicted in that story nor prevent someone else from reporting those facts—or can it? Under the "hot-news" doctrine, as it has become known, a narrow exemption for protection of certain facts may still exist, at least in the Second Circuit.

International News Service v. Associated Press

The United States Supreme Court first formulated the "hot-news" doctrine in 1918 in *International News Service v. Associated Press*.[3] At that time, of course, the current Copyright Act had not yet been enacted, and the 1909 Act was still in effect. Neither radio nor television effectively existed, and the most immediate means of communication was by wire. The Associated Press ("AP") and the International News Service ("INS") then competed in the "wire services" market and independently employed journalists to cover news events and to generate news articles based on those events, which they then supplied to affiliated newspapers throughout the country.

Seeing an opportunity to decrease its reporting costs and thereby increase profits, the INS began republishing and presenting as its own—without attribution—information and facts obtained from news articles that AP had originally created. AP then sued INS, seeking to enjoin its copying activities. On appeal, the United States Supreme Court found in favor of AP and granted AP's request for an injunction. Although the Supreme Court recognized that AP's news reports—particularly the facts reported—represented the "history of the day" and therefore were not copyrightable, the Court nonetheless found that because of the cost, skill, labor, and money involved in reporting and generating news, particularly news of the time-sensitive nature ("hot-news"), AP maintained a "quasi-property" right in its reports.[4]

To safeguard this "quasi-property" right, the Court created the "hot-news" doctrine, applicable to cases in which one party used its labor, skill, and money to follow and to report time-sensitive news. This doctrine, the Court observed, would prevent competitors "from reaping the fruits of the complainant's efforts and expenditures" and thereby provide an incentive for parties to collect "hot-news" (in much the same way, without the Court noting it, that the Copyright Clause is intended to promote the growth of knowledge).[5]

Applying the newly-created doctrine to the facts at hand, the Court found that INS's unpaid use of economically valuable and time-sensitive news constituted unlawful misappropriation of AP's quasi-property. To hold otherwise, said the Court, would "essentially divert profits away from those (AP) who earn or properly deserve them and toward those (INS) who did not."[6] This holding, not explicitly grounded in copyright law, was essentially equitable in nature and emphasized the property rights aspect of copyright protection, effectively protecting what the Court would later, in *Feist Publications Inc. v. Rural Telephone Service Co.*, characterize as mere "sweat of the brow."[7]

The 1976 Federal Copyright Act

Between 1918, when the Supreme Court decided *INS*, and 1976, when Congress passed the first significant revision to the Federal Copyright Act since 1909, the Federal Copyright Act did not expressly preempt state law misappropriation claims that were often based, at least loosely, on *INS*. By enacting the 1976 Copyright Act,[8] however, Congress included a two-part test to determine whether the Copyright Act preempts a state-law claim.[9]

According to the new Section 301, U.S. copyright law preempts a state-law claim (i) if the claim "seeks to vindicate 'legal or equitable rights that are equivalent' to one of the bundle of exclusive rights already protected by copyright law under 17 U.S.C. §106"[10] (commonly known as the "general scope requirement"); and (2) "if the work in question is of the type of works protected by the Copyright Act under 17 U.S.C. §§102 and 103"[11] (commonly known as the "subject

** David Applegate is a partner and the chair of the Intellectual Property Law Practice Group at Chicago's Williams Montgomery & John Ltd., a firm of trial lawyers dedicated to business litigation nationwide. Ryan Schermerhorn is a patent lawyer with Chicago's Lempia Summerfield Katz LLC. Both authors are members of the Federalist Society's Intellectual Property Law Practice Group and the Chicago Lawyers Chapter of the Federalist Society.*

matter requirement").[12] At first glance, Section 301 would seem to preempt all state law misappropriation claims. If one accepts the legitimacy of legislative history, however,[13] then the background of the 1976 Act suggests a general desire that "hot-news *INS*-like claims [survive] preemption."[14] The scope and breadth of such an exception nonetheless remain unclear, even within the Second Circuit.

Computer Associates Intern., Inc. v. Altai Inc.

In the years following the 1976 Act, the Second Circuit in particular has attempted to put teeth into the legislative history by developing an "extra element" exception to preemption. In *Computer Associates Intern., Inc. v. Altai Inc.*, it found:

> if an "extra element" is "required instead of or in addition to the acts of reproduction, performance, distribution or display [four of the copyright rights granted by the 1976 Copyright Act] in order to constitute a state-created cause of action, then the right does not lie 'within the general scope of copyright' and there is no preemption."[15]

At the same time, the Second Circuit in *Altai* expressed reservations about applying the test overly broadly: the "extra element test should not be applied so as to allow state claims to survive preemption easily."[16] Other than indicating that some but not all misappropriation claims could survive preemption, however, *Altai* provided no specific guidance concerning the breadth or application of the exception.

National Basketball Association v. Motorola, Inc.

More than fifty years after the "hot-news" doctrine was first established and five years after *Altai*, the Second Circuit considered, in *National Basketball Association v. Motorola, Inc.*,[17] (1) the extent to which a state-law "hot-news" misappropriation claim based on *INS* involves "extra elements" and thus survives preemption, and (2) the breadth of any surviving "hot-news" misappropriation cause of action.

Factual Background and Procedural History

At issue in *NBA* was whether the "real-time" transmission of National Basketball Association ("NBA") scores and information tabulated from in-progress television and radio broadcasts of NBA games to a sports pager constituted an unlawful misappropriation of *INS*-type "hot-news." The case largely centered on the Sportstrax handheld pager, manufactured and sold by Motorola and operated by Sports Team Analysis and Tracking Systems ("STATS"). The Sportstrax pager was designed to provide users with up-to-date "real time" information—*i.e.,* score, possession, quarter, time—for in-progress NBA games. STATS employed a team of reporters to collect this information, either by watching NBA games on television or listening to them on the radio, and to transmit it to STAT's central computer, which would then compile, analyze, and format the relevant data for transmission to and display on the Sportstrax pager.

The NBA then sued Motorola and STATS, alleging that both parties had misappropriated time-sensitive NBA game information. Accordingly, the NBA sought to enjoin Motorola and STATS from selling and operating the Sportstrax pager. The district court found for the NBA on these grounds and granted the NBA's request for a permanent injunction against Motorola and STATS. On appeal, the Second Circuit found (1) that a narrow "hot-news" exception does survive federal preemption under the Copyright Act, but (2) that by transmitting NBA game information Motorola and STATS had not unlawfully misappropriated "hot-news" property belonging to the NBA.[18]

Federal Preemption

Using the two-part preemption test of Copyright Act Section 301, the Second Circuit found that federal law did not preempt the NBA's misappropriation claim. Turning first to the "subject matter requirement," the court noted that although the essence of the NBA's claim involved NBA basketball games and facts associated with those games—both of which are uncopyrightable on their own—because the games and facts about those games were taken from a copyrighted NBA broadcast, the subject matter of the NBA's claim initially "f[ell] within the ambit of copyright protection."[19] Accordingly, the court concluded that the NBA's claim satisfied the second part of the preemption test, that of subject matter traditionally covered by copyright.

The court next addressed the "general scope" requirement of the Section 301 test. Recognizing that the NBA's claim for tortious behavior involved reproducing, distributing, and displaying facts taken from copyrighted broadcasts—each of which was consistent with the exclusive rights normally protected by federal copyright law—the court concluded that the NBA's claim also satisfied the general scope requirement for preemption.

Finding the two-prong preemption test satisfied, the court could have looked to the rule that copyright law does not protect facts, including such "data and information" as basketball scores, even quarter-by-quarter or minute-by-minute, to decide the preemption question. Instead, the court then turned its attention to the "extra element" exception to preemption it had proffered in *Altai*. In doing so, the court observed that a hot-news misappropriation claim is "not the equivalent of exclusive rights under a copyright," because misappropriation claims, for example, allegedly involve factors that are not considered central, much less peripheral, to a claim for copyright infringement, such as free-riding by a defendant.[20] Accordingly, the court held that a narrow "hot-news" misappropriation claim involves the extra elements necessary to survive preemption.[21]

But this explanation is problematic for at least two reasons. The first is that the Second Circuit's finding that a hot-news misappropriation claim is "not the equivalent" of exclusive copyright rights contradicts its finding that the NBA's claim sought to vindicate "legal or equitable rights that are equivalent" to one or more of the exclusive rights protected by copyright law under Section106 (the "general scope requirement"). Second, protection against "free-riding" by a defendant on a plaintiff's fact-gathering and assimilation is part of what the U.S. Supreme Court rejected in repudiating the "sweat of the brow" requirement in *Feist*. To allow such a state law misappropriation claim on the grounds that copyright law does not preempt it after first having found that the claim is preempted by copyright law disregards *Feist*.

"Hot-News" Unlawful Misappropriation Claim

Using *INS* as a guidepost (a 1918 case that the 1976 Copyright Act and *Feist* would seemingly have overturned), the Second Circuit turned next to the breadth of surviving "hot-news" claims and concocted a five-part test for an *INS*-like "hot-news" claim to succeed:

> (i) the plaintiff must generate or collect information at some cost or expense;
>
> (ii) the value of the information must be highly time-sensitive;
>
> (iii) the defendant's use of the information must constitute free-riding on the plaintiff's efforts to generate or to collect it;
>
> (iv) the defendant's use of the information must directly compete with a product or service that the plaintiff offers;[22] and
>
> (v) the ability of other parties to free-ride on plaintiff's efforts would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened.[23]

Analysis

In creating this test, the Second Circuit seems not only to have usurped the role of the legislature in an effort to fashion its own common law of copyright but also to have ignored the Supreme Court's *Feist* decision. Nonetheless, the court in turn applied this test to the NBA's misappropriation claim, placing particular emphasis on elements (ii), (iv), and (v). At the outset, the court conceded that both elements (ii) and (iv) were likely met, because the information transmitted to Sportstrax, although not done in full real-time, was "nevertheless time-sensitive," and Sportstrax was a direct competitor of the NBA's "Gamestats" product, a service that provides play-by-play logs and box scores for each game.[24] The court thus turned its focus to whether the operations behind the Sportstrax product constituted "free-riding" on the NBA's three main products: (1) generating relevant information by playing the games; (2) transmitting those games, in real-time, via copyrighted broadcasts; and (3) collecting and retransmitting facts and information for those games.

Here, the court found that Sportstrax did not have any competitive effect on the NBA's first two products because there "is no evidence that anyone regards Sportstrax . . . as a substitute for attending NBA games or watching them on television."[25] As to the third product—collecting and retransmitting facts and information—the court found that Motorola and STATS had expended their own resources (both money and time) to collect factual information generated from NBA games and to upload that information to their own network (the Sportstrax database) for eventual transmission to Sportstrax handheld pagers. (As an aside, the court did note that if STATS and Motorola were, at some point in the future, to collect facts and information from the NBA's Gamestats product and transmit it to their Sportstrax network and pagers, this would constitute free riding.) Without evidence of such activity, however, the court concluded that Motorola and STATS had not engaged in unlawful misappropriation of any "hot-news" that the NBA has generated.

All of this discussion, however, seems superfluous. If sufficiently original, performances, broadcasts, and the arrangement of collected facts may be copyrighted,[26] so as the Second Circuit's analysis conceded by seeking an exception in the first place, all three of the NBA's "products" at issue in *NBA* are at least potentially subject to copyright protection. Therefore, under a straightforward application of the 1976 Copyright Act—without importing any *INS*-inspired exceptions dating back to World War I—the results of *NBA* should have been the same. As the Second Circuit itself stated, Motorola and STATS plainly did not compete with the NBA in either performing, recording, or broadcasting the games, and at least since *Feist*—but under any plain understanding of the purpose and intent of the Copyright Clause—the NBA could not plausibly claim copyright in such reportable facts as the names of the players (which the NBA did not originate); which players scored, fouled out, or blocked shots; or what the score was at any given point in the game. Thus, all of *NBA*'s ruminations about a five-point "test" to determine whether a "sweat-of-the-brow" misappropriation claim otherwise preempted by copyright law should nonetheless be "excepted" from preemption were logically unnecessary to the result—hence, dicta.

Barclays Capital Inc. v. Theflyonthewall.com, Inc.

Undaunted, almost fifteen years after *NBA*, the Second Circuit has once again considered the viability and breadth of a "hot-news" misappropriation claim in *Barclays Capital Inc. v. Theflyonthewall.com*.[27] At issue in that case was whether Theflyonthewall.com's publication of recommendations by Barclays Capital, Merrill Lynch, Pierce, Fenner & Smith, and Morgan Stanley (collectively, "the Brokers"), constituted an unlawful misappropriation of "hot-news" belonging to the Brokers.

Factual Background

The Brokers are major financial institutions that provide brokerage recommendations regarding buying, holding, and/or selling securities of various companies. The Brokers prepare these recommendations following extensive (both expensive and time-consuming) research about those companies, the industries in which those companies compete, and the securities markets, then distribute them to institutional clients and prospective clients such as hedge funds, private equity firms, and wealthy investors. These recommendations purportedly provide recipients with an informational advantage over non-recipients, with the idea that the recipients will profit financially from this informational advantage and that the Brokers will, in turn, profit from commissions earned as a result of purchases or sales that the recipients execute based on these recommendations.

Theflyonthewall.com ("Fly") is a news aggregator that, on a number of occasions and through various means, timely acquired the Brokers' recommendations and distributed them through its website and third-party distributors to paid subscribers, including individual investors, institutional investors, brokers, and day traders. Like the Brokers, Fly aimed to distribute the acquired recommendations before the U.S. securities markets opened each day.

Procedural History and District Court Opinion

The Brokers sued Fly in the United States District Court for the Southern District of New York, alleging that Fly's unauthorized acquisition and publication of the Brokers' recommendations constituted unlawful "hot-news" misappropriation of the Brokers' property. In particular, the Brokers asserted that unauthorized publication of these recommendations threatened the Brokers' financial viability because clients and prospective clients would allegedly learn of the recommendations from sources other than the Brokers, thereby reducing the Brokers' abilities to derive commission income from the significant resources needed to create the recommendations in the first place.

The district court found for the Brokers on the grounds that their misappropriation claim was not preempted and that it satisfied the elements of the "hot-news" test set forth in *NBA.*[28] Applying the five-factor *NBA* test, the district court found that the first two elements (that the Brokers generated or collected information at some cost or expense and that the value of the information was highly time-sensitive, which Fly did not dispute) were easily met. Regarding the third factor—whether Fly's use of the information constituted free-riding on the Broker's efforts—the district court found that because Fly "does no equity research of its own, nor . . . undertakes any original reporting or analysis," it contributes "nothing to the actual [r]ecommendations" provided by the Brokers.[29] Fly's activities, in the district court's opinion, thus constituted "free-riding" on the Brokers' costly efforts to generate or to collect their recommendations.

In finding that the fourth factor was present—i.e., that Fly's use of the information directly competed with a product or service offered by the Brokers—the district court relied on the fact that Fly and the Brokers were both in the business of "disseminating [r]ecommendations to investors for their use in making investment decisions" and that both companies used "similar distribution channels."[30] With regard to the fifth factor—whether others' ability to free-ride on the Brokers' efforts would so reduce the incentive to produce the product or service that its existence or quality would be substantially threatened—the district court found that "common sense and the circumstantial evidence about the [Brokers'] business model" supported a finding that the Brokers would have a reduced incentive to continue generating their recommendations if Fly could legally retransmit them.[31]

Having concluded that the Brokers demonstrated a valid claim for "hot-news" misappropriation, the district court promptly permanently enjoined Fly from reporting the Brokers' recommendations for periods ranging from thirty minutes to several hours following their release by the Brokers, when the "news" would no longer be "hot."

Second Circuit—Majority Opinion

On appeal, the Second Circuit reversed the district court's judgment in favor of the Brokers, holding that the Copyright Act preempts the Brokers' "hot-news" misappropriation claim.[32] Applying Section 301's two-part preemption test, the court found that the Brokers' recommendations satisfy both the "subject matter" and the "general scope" requirements. Although the facts contained in the recommendations are, themselves, not copyrightable, the recommendations constitute a work "of a type covered by section 102," namely original works of authorship.[33] Likewise, the claim fulfilled the "general scope" requirement of Section 301 because Fly's acts of reproducing and distributing the Brokers' recommendations were of the type that "would infringe one of the exclusive rights' provided by federal copyright law," namely reproduction and distribution.[34]

Turning next to the "extra elements" exception to preemption invented in *Altai* and explained in *NBA*, the court found a noticeable absence of any significant "extra elements" that would warrant finding the Brokers' claims not preempted. In particular, the court relied heavily on its finding that Fly was not "free-riding" on the Brokers' activities. Again, as in *NBA*, the court looked to *INS* for guidance. According to the *INS* court, "hot-news" and—more particularly—"free riding" are defined as "taking material that has been acquired by complainant as the result of organization and the expenditure of labor, skill, and money . . . and appropriating it and selling it as defendant's own."[35] In contrast, the court found, once Fly obtained news of a recommendation, it did not sell that recommendation "as its own" but instead sold the information with "specific attribution" to the issuing broker.[36] Moreover, the court found, Fly was unlikely to profit or to gain from selling that information "as its own," because "it is not the identity of Fly" but rather the identity of the financial institution that lends credibility to the recommendation.[37]

To support its finding that Fly had not engaged in "free riding," the court also drew parallels between the case at hand and the facts in *NBA.* Like STATS and Motorola in *NBA*, the court found, Fly "has its own network and assembles and transmit[s] data itself."[38] Moreover, much like the SportsTrax service in *NBA*, which "b[ore] its own costs of collecting factual information on NBA games," Fly's news service utilized a significant amount of its resources (fourteen of twenty-eight total employees) to collect the Brokers' recommendations.[39] As a result, the Second Circuit found that Fly's service was not the type of *INS*-like product "that could support a non-preempted cause of action for misappropriation."[40]

But just as in *NBA*, the Second Circuit's discussion in *Flyonthewall* of exceptions to preemption seems wholly unnecessary. Having already found that the Brokers' recommendations satisfy both the "subject matter" and the "general scope" requirements of Section 301, the court could easily—and properly—have reached the same result directly under federal copyright law without having to wander off into thickets of exceptions and forests of five-part tests. Instead, the Second Circuit could readily have found that U.S. copyright law provided the exclusive remedy for the Brokers (the very meaning of "preemption") and that, under *Feist*, the Brokers' sweat of the brow in researching and assembling their recommendations did not protect either the underlying information (i.e., facts about the subject companies) or the recommendations themselves, which the court could have found lacked sufficient originality under *Feist*. Either way the result would be the same, without a lot of unnecessary—and arguably unconstitutional—exposition by the court.

Second Circuit—Concurring Opinion

Although ultimately concurring with the majority in favor of the Brokers, Judge Raggi wrote separately to express her disagreement with the majority's reasoning. Like her colleagues in the majority, Judge Raggi concluded that the Brokers' claims satisfied both the "subject matter" and "general scope" requirements of Section 301; in contrast, she criticized the majority for essentially treating the five-part test from *NBA* as nothing more than "dictum."

According to Judge Raggi, the majority improperly concluded that "*NBA* 'held' only that the facts presented could not establish a non-preempted 'hot news' claim," and thus dismissed the *NBA* test as "an unnecessary discussion of hypothetical circumstances giving rise to a 'hot news' claim"—hence, dictum that need not be followed.[41] Instead, Judge Raggi noted, the Second Circuit in *NBA* was "required to determine the 'breadth' of the 'hot news' claim that survives preemption."[42] In response, she said, the court "identified five factors required to state a non-preempted 'hot news' claim [the five-part *NBA* test], applied them to the facts presented, and concluded that the plaintiffs' claim failed."[43] Despite having her own reservations about the *NBA* test, therefore, Justice Raggi observed that because the *NBA* test was necessary to the *NBA* opinion, "it is not dictum," and thus should be applied to the facts at hand.[44]

In applying that test to the facts at hand, Judge Raggi found that the Brokers' claim failed to satisfy the test and was therefore preempted. Although Judge Raggi disagreed with the majority (and agreed with the district court) that Fly's conduct was "strong evidence of free-riding . . . [because] Fly is usurping the substantial efforts and expenses of the [Brokers] to make a profit without expending any time or cost to conduct research of its own,"[45] she nonetheless found the Brokers' claim preempted based on a lack of direct competition between the Brokers' product and Fly's newsfeed containing the Brokers' recommendations.

Although the Brokers and Fly broadly share the overall goal of disseminating the Brokers' recommendations to clients and subscribers, Judge Raggi found, the Brokers and Fly do not compete directly with each other because (1) the Brokers do not collect or disseminate other Brokers' recommendations, whereas Fly collects and disseminates recommendations from over sixty-five different firms; and (2) the Brokers limit access to their recommendations and the underlying research to those clients that generate sufficient trading revenue, whereas Fly disseminates that "'financial news' to anyone interested" and does not seek trading commissions of its own.[46] As a result, Judge Raggi reasoned, Fly's product was "sufficiently distinct from the [Brokers'] business model" that the Brokers and Fly did not directly compete with one another.[47] Finding the *NBA* test unmet, Judge Raggi concluded that the Copyright Act preempted the Brokers' misappropriation claim[48] and joined the majority in reversing the district court's judgment.

Conclusion

In our view, both the majority and the concurring judges have engaged in unnecessary analysis while reaching the right result. Simply put, as the U.S. Supreme Court clarified in *Feist*, U.S. copyright law does not protect facts or ideas, no matter how carefully or cleverly arranged, but only the particular expression of those ideas if sufficiently original. Even giving credence to the dubious value of Section 301's "legislative history," it is difficult to envision a claim that (1) "seeks to vindicate 'legal or equitable rights that are equivalent' to one of the bundle of exclusive rights already protected" under Section106[49] regarding (2) a "work . . . of the type of works protected" under Sections 102 and 103[50] and is not by those very terms preempted.

Although a cautious Congress may have acted wisely to ensure the possibility of future exceptions for hard cases it could not then envision, interpreting the statute as written, without grafting on court-made exceptions concerning which the Second Circuit cannot even agree—either over time or within its contemporaneous members—would have allowed the same judicial flexibility without the unnecessary machinations exhibited in the Second Circuit's opinions. (All it would take would be for a court to determine, on the facts of the hypothetical hard case before it, that the rights sought to be protected are not precisely equivalent to the rights that Section 106 already protects, or that the work is not "of the type . . . protected" under Sections 102 and 103.) Under our constitutional system of both delegated powers and separation of powers, such a process would be preferable to the continuing and unsatisfactory spectacle of panels of unelected judges creating federal common law on the validity and application of which they themselves apparently cannot even agree.

Endnotes

1 *See* Feist Publ'ns Inc. v. Rural Tel. Serv. Co., 499 U.S. 340 (1991).

2 U.S. Const. art. I, § 8, cl. 8.

3 Int'l News Serv. v. Associated Press, 248 U.S. 215 (1918).

4 *Id.* at 234, 242.

5 *Id.* at 241.

6 *Id.* at 240.

7 *See* Feist Publ'ns Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 352, 359, 360 (1991) (The "sweat of the brow" doctrine refers to the "underlying notion that a copyright was a reward for the hard work that went into compiling facts. . . . [T]he 1976 revisions to the Copyright Act leave no doubt that originality, not "sweat of the brow," is the touchstone of copyright protection in directories and other fact-based works.").

8 1976 Copyright Act, 17 U.S.C. §§ 101-810, 18 U.S.C. § 2318, 44 U.S.C. §§ 505, 2113 (2010).

9 17 U.S.C. § 301.

10 17 U.S.C. § 106 (2010). The owner of the copyright is generally vested with the exclusive rights to "do and authorize any of the following":

(1) to reproduce the copyrighted work in copies or phonorecords; (2) to prepare derivative works based upon the copyrighted work; (3) to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending; (4) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly; (5) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and (6) in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission.

Id.

11 17 U.S.C. §§ 102-103 (2010). Sections 102 and 103 generally protect works of authorship, fixed in any tangible medium of expression, that fall into any one of the following categories: "(1) literary works; (2) musical works, including any accompanying words; (3) dramatic works, including any accompanying music; (4) pantomimes and choreographic works; (5) pictorial, graphic, and sculptural works; (6) motion pictures and other audiovisual works; (7) sound recordings; and (8) architectural works." *Id.*

12 17 U.S.C. § 301 (2010).

13 *See, e.g.,* Zedner v. United States, 547 U.S. 489, 511 (2006) (Scalia, J., concurring in judgment) ("The use of legislative history is illegitimate and ill advised in the interpretation of any statute."); *see also* Conroy v. Aniskoff, 507 U. S. 511, 518-528 (1993) (Scalia, J., concurring in judgment).

14 Nat'l Basketball Ass'n v. Motorola, Inc., 105 F.3d 841, 845 (2d. Cir. 1997); *see also* H.R. Rep. No. 94-1476, at 132 (1976), reprinted in 1976 U.S.C.C.A.N. 5659, 5748 ("'Misappropriation is not necessarily synonymous with copyright infringement . . . . [S]tate law should have the flexibility to afford a remedy (under traditional principles of equity) against a consistent pattern of unauthorized appropriation by a competitor of the facts (i.e., not the literary expression) constitution 'hot news'); Fin. Info., Inc. v. Moody's Investors Serv., Inc., 808 F.2d 204, 208 (2d Cir. 1986) ("[M]isappropriation of hot news, under *INS*, is a branch of the unfair competition doctrine not preempted by the Copyright Act according to the House Report.").

15 Computer Assocs. Int'l, Inc. v. Altai Inc., 982 F.2d 693, 716 (2d. Cir. 1992).

16 105 F.3d at 851 (citing *Altai*, 982 F.2d at 717). "An action will not be saved from preemption by elements such as awareness or intent, which alter the action's scope but not its nature." *Altai*, 982 F.2d at 717.

17 105 F.3d at 841.

18 *Id.*

19 105 F.3d at 848 (citing Harper & Row, Inc. v. Nation Enter., 723 F.3d 195, 200 (1983), rev'd on other grounds, 471 U.S. 539 (1985)); *see also* 105 F.3d at 848-49 ("Although game broadcasts are copyrightable while the underlying games are not, the Copyright Act should be read to distinguish between the two when analyzing the preemption of a misappropriation claim based on copying or taking from the copyrightable work.").

20 105 F. 3d at 850.

21 *Id.* at 841.

22 Generally, this requires that the defendant's use of the information results in direct competition in the plaintiffs' primary, rather than collateral, market. *Id.* at 853.

23 *Id*.

24 *Id.*

25 *Id.*

26 *See* 17 U.S.C. § 106. The owner of the copyright is generally vested with the exclusive rights to "do and authorize any of the following":

> (1) to reproduce the copyrighted work in copies or phonorecords; (2) to prepare derivative works based upon the copyrighted work; (3) to distribute copies or phonorecords of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending; (4) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and motion pictures and other audiovisual works, to perform the copyrighted work publicly; (5) in the case of literary, musical, dramatic, and choreographic works, pantomimes, and pictorial, graphic, or sculptural works, including the individual images of a motion picture or other audiovisual work, to display the copyrighted work publicly; and (6) in the case of sound recordings, to perform the copyrighted work publicly by means of a digital audio transmission.

Id.

27 Barclays Capital Inc. v. Theflyonthewall.com, No. 10-1372-cv, 2011 WL 2437554 (2d Cir. June 20, 2011).

28 Barclays Capital, Inc., v. Theflyonthewall.com, 700 F. Supp. 2d 310 (S.D.N.Y. 2010).

29 *Id.* at 336.

30 *Id.* at 339.

31 *Id.* at 343.

32 2011 WL 2437554, slip op. at *1.

33 *Id.* at *13 (citing 17 U.S.C. § 301 (2010)).

34 *Id*.

35 *Id*. at *22 (citing Int'l News Serv. v. Associated Press, 248 U.S. 215, 239 (1918)).

36 2011 WL 2437554, slip op. at *22.

37 *Id*.

38 *Id*. at *20.

39 *Id.* at *21 (citing Nat'l Basketball Ass'n v. Motorola, Inc., 105 F.3d 841, 854 (2d Cir. 1997)).

40 2011 WL 2437554, slip op. at *24.

41 *Id.* at *29 (Raggi, J., concurring).

42 *Id*.

43 *Id*.

44 *Id*.

45 *Id*. at *31.

46 *Id*.

47 *Id*. at *32.

48 *Id*.

49 17 U.S.C. § 106 (2010).

50 17 U.S.C. §§ 102-103 (2010). Sections 102 and 103 generally protect works of authorship, fixed in any tangible medium of expression, that fall into any one of the following categories: "(1) literary works; (2) musical works, including any accompanying words; (3) dramatic works, including any accompanying music; (4) pantomimes and choreographic works; (5) pictorial, graphic, and sculptural works; (6) motion pictures and other audiovisual works; (7) sound recordings; and (8) architectural works." *Id.*

International & National Security law

Multinational Businesses and the Matrix of Human Rights Governance Networks

*By James P. Kelly, III**

For decades, human rights activists have successfully petitioned state and national governments in developed countries to fund such economic rights as the right to housing, the right to education, the right to a clean and safe environment, the right to work, the right to social security, and the right to health. Just at the time that the bill is coming due for such expenditures and developed countries are facing the resulting global economic crisis, these advocates are pursuing the realization of these economic rights in developing countries. However, a lack of government funding is forcing them to look to transnational corporations and other multinational business enterprises for the funding of their social welfare ambitions. In doing so, they are relying on a matrix of human rights governance networks (the "Matrix") first described in a 2008 article in this journal.[1] This follow-up article explains 1) how human rights activists and multinational institutions are using the Matrix to govern the operations of multinational business enterprises; 2) how the Matrix has become an "intellectual complex adaptive system" that, after facing initial resistance from the business community, has evolved to increase its scope and effectiveness; and 3) some steps that multinational businesses might take to resist the Matrix.

The Matrix Revisited: The Business and Human Rights Context

International non-governmental organizations ("NGOs") and national civil society organizations ("CSOs") are using a matrix of human rights governance networks to bypass national courts, democracy, and the rule of law to develop "soft law" human rights norms, with which multinational business enterprises will have to comply from the early stages of project research, design, and planning through project completion and beyond. As will be described in this paper, this matrix is not a conspiratorial undertaking pursued by a few like-minded, non-transparent NGOs; rather, it consists of an observable and increasingly institutionalized group of interconnected networks through which NGOs and CSOs realize their human rights governance agenda outside the ordinary democratic process.

The ten human rights governance networks comprising the Matrix include:

1. *Advocacy networks*: The networks of international human rights activists that articulate and advocate for human rights, including so-called "emerging" economic and social human rights.

2. *Research networks*: The networks of social scientists and academics that conduct research on how the lack of human rights protection negatively impacts individuals and society.

3. *Policy networks*: The networks of government officials and other policy makers that discuss and formulate human rights policies.

4. *Standards-setting networks*: The networks of multilateral international organizations that meet to adopt treaties or declarations listing human rights norms or standards.

5. *Interpretive networks*: The networks of human rights treaty committees and UN-sanctioned expert committees that interpret the norms and standards contained in human rights treaties and declarations.

6. *Explanatory networks*: The networks of international organizations and their NGO and CSO partners that explain the human rights interpretations to members of civil society at the local, national, and regional levels.

7. *Implementation networks*: The networks of national legislatures and government officials that, upon the recommendation of the human rights experts, adopt and implement laws and regulations promoting and protecting human rights.

8. *Assessment networks*: The networks of NGOs and government officials that encourage the use of human rights impact assessments by legislatures and businesses to measure the potential human rights impact of proposed legislation or products.

9. *Enforcement networks*: The networks of local, national, and regional courts; government agencies; and human rights treaty committees that decide cases or rule on alleged human rights violations.

10. *Funding networks*: The networks of governments, multilateral institutions, and private foundations that fund the promotion and protection of human rights by supporting one or more of the other human rights governance networks.

The ten human rights governance networks comprising the Matrix work in successive stages. The advocacy networks generate the idea for an emerging economic right; the research networks conduct the research necessary to support the right; the policy networks design the policy that embodies the right; the standards-setting networks publicly adopt or declare the right as a norm or standard; the interpretive networks determine the nature and scope of the right; the explanatory networks explain the right to the affected parties and their supporters in civil society; the implementation networks adopt the legislation that promotes or protects the right; the assessment networks encourage government and business respect for the right; the

** James P. Kelly, III serves as the Director of International Affairs for the Federalist Society. He also is the President of Solidarity Center for Law and Justice, P.C., in which capacity he wrote this article.*

enforcement networks penalize those who violate the right; and the funding networks help sustain one or more of the human rights governance networks comprising the Matrix.

Beginning about a decade ago, NGOs, CSOs, and multilateral institutions began using the Matrix in an attempt to hold multinational business enterprises accountable for 1) assessing their human rights responsibilities in the developing countries in which they operated and 2) funding the fulfillment of the economic rights of the residents therein who, in the opinion of the NGOs and CSOs, are being adversely impacted as a result of such operations. The Matrix produced three primary mechanisms in the area of business and human rights: the Guidelines for Multinational Enterprises produced by the Organisation for Economic Co-operation and Development; the Ten Principles of the United Nations Global Compact; and the Norms on the Responsibilities of Transnational Corporations and Other Businesses with Regard to Human Rights adopted by the United Nations Sub-Commission on the Promotion and Protection of Human Rights.

Pursuant to, or coincident with, these mechanisms, from 2000 to 2008, transnational corporations and other multinational business enterprises were subjected to the Matrix in the following manner:

1. *Advocacy networks*: In 2000, leading activists in the field of economic, social, and cultural rights ("ESCR") from key human rights organizations in the Americas, Africa, and Asia came together to develop an international network for the promotion of economic, social, and cultural rights. The process culminated in the founding of a General Assembly and the Inaugural ESCR-Net Conference, titled "Creating New Paths towards Social Justice," held in Chiang Mai, Thailand in 2003. Over 250 human rights activists from fifty different countries came together to launch the network and to elect the first ESCR-Net Board. The ESCR-Net Corporate Accountability Working Group (the "Working Group") advocates for corporate accountability at the international level.

2. *Research networks*: In 2001, the United Nations Sub-Commission on the Promotion and Protection of Human Rights (the "Sub-Commission") asked its Working Group on the Working Methods and Activities of Transnational Corporations (the "Working Group") to contribute to the drafting of relevant norms concerning human rights and transnational corporations and other economic units whose activities have an impact on human rights. In 2002, the Sub-Commission requested that the Working Group's report and the annexed draft norms be widely circulated in the expectation that comments would be taken into account when the Working Group next considered its draft norms in August 2003.

3. *Policy networks*: In 2000, the Global Reporting Initiative ("GRI"), a network-based organization, released its first sustainability reporting guidelines, which are designed to mainstream disclosure by businesses on environmental, social, and governance performance. GRI's reporting framework is developed through a consensus-seeking, multi-stakeholder process, with participants being drawn from global business, civil society, labor, academic and professional institutions.

4. *Standards-setting networks*: Four multinational institutions played an early role in creating (or attempting to create) human rights standards that could be used to hold multinational business enterprises accountable for protecting and fulfilling economic rights.

First, in 1976, after being ratified by the necessary number of States Parties, the International Covenant on Economic, Social and Cultural Rights ("ICESCR") entered into force. The ICESCR commits its parties to work toward the granting of economic, social, and cultural rights to individuals, including labor rights, the right to health, the right to education, and the right to an adequate standard of living. As of July 2011, 160 States had ratified the ICESCR; however, the United States has not done so.

Also in 1976, the Organisation for Economic Co-operation and Development ("OECD") adopted the OECD Guidelines for Multinational Enterprises (the "OECD Guidelines"). The OECD Guidelines constitute a set of voluntary recommendations to multinational enterprises in all the major areas of business ethics, human rights, environment, information disclosure, combating bribery, consumer interests, science and technology, competition, and taxation. Adhering governments have committed to promoting the OECD Guidelines among multinational enterprises operating in or from their territories. All of the thirty-four OECD member countries and eight non-OECD countries have agreed to adhere to the OECD Guidelines and encourage multinational enterprises to comply with their provisions.

Third, in 2000, the UN Global Compact was launched to bring businesses together with UN agencies, labor unions, civil society, and governments to advance ten universal principles in the areas of human rights, labor, environment, and anti-corruption (the "Ten Principles"). Although companies are asked to mainstream the Ten Principles within their spheres of influence, the UN Global Compact explicitly denies that it is a regulatory initiative. Instead, it claims to offer a values-based platform for voluntary peer review and institutional learning. Participants are encouraged to share case studies of good practices and to participate in policy dialogues.

Finally, in August 2003, the Sub-Commission approved the Norms on the Responsibilities of Transnational Corporations and Other Businesses with Regard to Human Rights (the "Norms"). In part, the Norms provided that:

> 12. Transnational corporations and other business enterprises shall respect economic, social and cultural rights as well as civil and political rights and contribute to their realization, in particular the rights to development, adequate food and drinking water, the highest attainable standard of physical and mental health, adequate housing, privacy, education, freedom of thought, conscience, and religion and freedom of opinion and expression, and shall refrain from actions

which obstruct or impede the realization of those rights.[2]

The UN Commission on Human Rights (the "Commission") considered the Norms in April 2004; however, it did not approve them and adopted the position that the Norms had no legal standing. At the time, it was obvious that the Sub-Commission had taken a position on the obligation of transnational businesses to respect and fulfill economic rights that exceeded the more limited position held by a majority of UN member states. As a result, the United Nations took no further actions on the Norms. To satisfy disappointed human rights activists and to explore a more independent and reasonable position on the subject of norms on the human rights responsibilities of multinational businesses, in July 2005, then UN Secretary-General Kofi Annan appointed John Ruggie as the Special Representative of the UN Secretary-General on the issue of human rights and transnational corporations and other business enterprises (the "Special Representative").

5. *Interpretive networks*: Although the Commission did not adopt the Norms, as part of the process of formulating the Norms, the Sub-Commission prepared commentary on them (the "Commentary"). The Commentary, which interpreted each provision of the Norms, provided an in-depth look at the provisions comprising the most ambitious agenda for holding multinational enterprises responsible for realizing economic rights.

Also, since 1991, the Committee on Economic, Social and Cultural Rights (the "ICESCR Committee") has been developing and publishing General Comments that have interpreted the meaning and scope of various economic rights contained in the ICESCR, including the rights to adequate housing, adequate food, education, the highest attainable standard of health, water, work, and social security.

6. *Explanatory networks*: Since the OECD Guidelines were adopted in 1976 and significantly revised in 2000, CSOs have gone to great lengths to explain them to multinational business enterprises and government officials. In 2003, a group of CSOs meeting in Amersfoort, the Netherlands, established OECD Watch, a network that seeks to strengthen cooperation between CSOs worldwide, build CSO capacity, and promote a corporate accountability framework in the interest of sustainability and poverty eradication. To do so, OECD Watch primarily aims to help facilitate NGO activities around the OECD Guidelines through a membership that consists of a diverse range of national CSOs working on human rights, labor rights, consumer rights, transparency, the environment, and sustainable development.

7. *Implementation networks*: Primary responsibility for implementing the OECD Guidelines rests with National Contact Points ("NCPs"). The NCP is a national government office responsible for encouraging observance of the OECD Guidelines in a national context and for ensuring that the Guidelines are well-known and understood by the national business community and by other interested parties. The NCP gathers information on national experiences with the OECD Guidelines; handles inquiries; discusses matters related to the OECD Guidelines; and assists in solving problems that may arise in their implementation.

As for the implementation of the UN Global Compact's Ten Principles, participating companies are required to issue an annual Communication on Progress ("COP"), a public disclosure to stakeholders (e.g., investors, consumers, civil society, governments, etc.) on progress made in implementing the Ten Principles and in supporting broad UN development goals.

8. *Assessment networks*: In 1993, the UN General Assembly adopted Principles relating to the Status of National Institutions (the "Paris Principles"), which led to the creation of national human rights institutions ("NHRIs") in many countries. According to the Paris Principles, NHRIs must have a broad mandate under national law to promote and protect human rights, including through monitoring and advising home governments, investigating human rights abuses, engaging with international human rights bodies, public education, and research. Also, in 1993, NHRIs established the International Coordinating Committee of National Institutions for the Promotion and Protection of Human Rights (the "ICC") to secure greater integration of NHRI activities. Presently, NHRIs exist in more than 100 countries.

Also, in 2005, the Commission requested the Special Representative to develop materials and methodologies for undertaking human rights impact assessments ("HRIAs") for business activity. Although the Special Representative determined that developing such materials and methodologies was beyond his mandate's time and resource constraints, in 2007, the Special Representative published a report describing the principles and characteristics of HRIAs for business, including similarities to environmental and social impact assessments, and providing updates on current HRIA initiatives (the "HRIA Report"). In the HRIA Report, the Special Representative explained that, prior to engaging in a proposed business activity, a business enterprise should conduct a HRIA to examine whether human rights protections have been adequately considered. In his view, HRIAs should catalogue the relevant human rights standards, including those set out in international conventions to which the home and host countries are signatories, other standards such as indigenous customary laws and traditions, and, in cases of armed conflict, international humanitarian law.[3]

9. *Enforcement networks*: Until the 2000 revision of the OECD Guidelines, no complaint mechanism existed through which parties could pursue relief from alleged violations of the voluntary recommendations contained therein. Since the 2000 revision, when issues arise concerning implementation of the OECD Guidelines in relation to specific instances of business conduct, the NCP is expected to help resolve them. Under the OECD Guidelines Procedural Guidance, as revised in 2000 (the "Procedural Guidance"), when the NCP receives a complaint, it has to "make an initial assessment of whether the issues raised merit further examination and respond to the party or parties raising them" and where "the issues

raised merit further examination, offer good offices to help the parties involved to resolve the issues."[4] The Procedural Guidance did not explain when issues that are raised in the complaint merit further consideration. Under the Procedural Guidance, if a NCP decides to proceed with the complaint, and provided that the parties involved consent, it plays a mediating role in bringing parties together to resolve the issue.

10. *Funding networks*: The United Nations and OECD contribute significant funds to develop and sustain the different networks comprising the Matrix, with the United States of America, in turn, providing the largest amount of the regular funding for those two organizations.

Inside the Matrix: The Matrix as an Intelligent Complex Adaptive System

By 2008, the Matrix provided NGOs and CSOs with a loosely organized framework and mechanism for protecting and realizing (i.e., funding) economic rights in developing and developed countries. Yet, the Matrix had evolved from the primarily independent and uncoordinated efforts of individuals who had been pursuing an economic rights agenda within the context of their individual networks. These NGOs and CSOs were unaware of the degree to which their individual efforts had given birth to a comprehensive, integrated, complex system for holding businesses accountable for protecting and fulfilling economic rights. Thus, it is completely understandable that transnational corporations and other multinational business enterprises were likewise unaware of what had transpired. Until that time, businesses were under the impression that they could satisfy the demand for protecting and fulfilling economic rights by engaging in basic corporate social responsibility or sustainability measures, joining the UN Global Compact, participating in the World Economic Forum in Davos, Switzerland, or contributing to one or two UN-sponsored humanitarian relief programs.

Yet, from the perspective of those frustrated NGOs and CSOs that were seeking greater corporate accountability for protecting and fulfilling economic rights, the following shortcomings in the Matrix existed:

1. The UN Global Compact and the UN Secretary-General were not serious about holding members accountable for complying with the Ten Principles.

2. The OECD Guidelines did not adequately address business and human rights and national governments were not constructing or operating National Contact Points in a manner that could credibly and fairly resolve disputes over whether a multinational business enterprise was adhering to the OECD Guidelines.

3. There was no mechanism whereby individuals could communicate to the ICESCR Committee cases where a State Party was not holding transnational corporations or other multinational businesses accountable for failing to protect or fulfill the economic rights of indigenous peoples or other groups.

4. NHRIs were not being adequately educated about the need for them to monitor the degree to which multinational businesses were failing to protect or fulfill economic rights.

5. Little progress had been made to require businesses to conduct HRIAs prior to launching a new product or project.

6. By failing to adopt the Norms or any comparable internationally-approved standards, the UN had failed to provide the necessary leadership on linking businesses and human rights.

In a historic development having evolutionary significance, within the short span of the past four years, the Matrix has adapted to address all of these perceived shortcomings. In many ways, the Matrix represents a type of "intellectual complex adaptive system" ("ICAS"), a system explained by Alex and David Bennet in their 2004 book, *Organizational Survival in the New World*. A "complex adaptive system" is one composed of a large number of self-organizing components that seek to maximize their own goals but operate according to rules and in the context of relationships with other components and the external world. Examples include ant colonies, cities, the brain, the immune system, ecosystems, computer models, and organizations.[5] The ICAS is a type of complex adaptive system:

> The ICAS, as a complex organization, is composed of a large number of individuals, groups, and human subsystems that have nonlinear interaction and the capability to make many local decisions and strive for specific end states or goals. These components build many relationships both within the organization and external to the organization's boundaries that may become highly complex and dynamic. Together, these relationships and their constituents form the organization and its enterprise. The word *adaptive* implies that the organization and its subcomponents are capable of studying and analyzing the environment and taking actions that internally adjust the organization and externally influence the environment in a manner that allows the organization to fulfill local and higher-level goals.[6]

The success of an ICAS depends on the competency and freedom of individual participants in the system in terms of learning, decision-making, and taking actions. The ability of individuals to learn, decide, and take actions in an ICAS are leveraged through multiple and effective networks that provide sources of knowledge, experience, and insights from others.

> Dynamic networks will represent the critical infrastructure of the next-generation knowledge-based organization. Made available by increased bandwidth and processing power of both silicon and biotechnology, they will offer the opportunity for virtual information and knowledge support systems that connect data, information, and people through virtual communities, knowledge repositories, and knowledge portals. The foundation and grounding of future firms will be strengthened through a common set of strong, stable values held by all employees. Such values

not only provide a framework that enhances empowerment but also motivate and strengthen the self-confidence of the workforce, thereby magnifying the effectiveness of the self-organized teams within the ICAS. To survive and compete in the future world, these organizations will need to possess a number of emergent characteristics that taken together result in resilience, agility, adaptivity, and learning, all well-known traits of survival.[7]

In essence, the Matrix is an ICAS comprised of dynamic networks sharing common values associated with the mission of holding multinational business enterprises accountable for protecting and fulfilling economic rights.

The Matrix Reloaded: Adapting to the Resistance of Multinational Business Enterprises

During the past four years, the Matrix has adapted to the resistance of multinational business enterprises and "reloaded" by enhancing the networks comprising the Matrix as follows:

1. *Strengthened the UN Global Compact*. In order to provide clear benchmarks for corporate adherence to the Ten Principles, in May 2010, officials from the UN Global Compact and GRI agreed to cooperate in amending the GRI Guidelines to include performance indicators that address the Ten Principles. In March 2011, the GRI released version 3.1 of the GRI Guidelines, including two new indicators on human rights that call upon businesses to disclose 1) the percentage and total number of business operations that have been subject to human rights reviews and/or impact assessments and 2) the number of grievances related to human rights that have been filed, addressed, and resolved through formal grievance mechanisms. Meanwhile, in 2010, the UN Joint Inspection Unit published a report that severely criticized the UN Global Compact for the lack of a clear and articulated mandate, the lack of any adequate entry criteria for participants, the lack of an effective monitoring system to measure actual implementation of the principles by participants, the lack of adherence to existing rules and regulations relating to "normal" UN offices, the lack of any representation of either UN Member States or other UN agencies on its Board, and the lack of regular unbiased and independent performance evaluation of its operations. As a result of pressure from the office of the UN Joint Inspection Unit and the NGO community, the UN Global Compact is asking its member businesses to adopt a three-part "leadership blueprint" called the Global Compact LEAD Platform. Under the LEAD Platform, businesses would be expected to implement the Ten Principles into strategies and operations (using the GRI Human Rights Performance Indicators), take action in support of broader UN development goals and issues, and work with the UN Global Compact in creating global and local working groups on issue-based and sector initiatives relating to the Ten Principles. In September 2011, the UN Global Compact's Human Rights Working Group, composed of representatives of business, civil society, trade unions, the UN, and academia, met under newly-revised Terms of Reference that focus on promoting the business and human rights agenda in the context of the UN Global Compact.

2. *Amended the OECD Guidelines and Procedural Guidance Relating to NCPs.* In May 2011, the OECD Guidelines were amended to include a new chapter on human rights, which requires multinational enterprises to "carry out human rights due diligence as appropriate to their size, the nature and context of operations and the severity of the risks of adverse human rights impacts."[8] At the same time, the Procedural Guidance was amended to clarify and enhance the role of NCPs in contributing to the resolution of issues that arise relating to the implementation of the OECD Guidelines in specific instances of alleged violations by multinational enterprises.

3. *Adopted the Optional Protocol to the ICESCR to Permit Individual Communications*. In late 2008, the UN General Assembly unanimously adopted an Optional Protocol to the ICESCR that permits the ICESCR Committee to receive and consider communications (i.e., complaints) from individual citizens alleging the failure of a State Party to implement the provisions of the ICESCR. While the decisions of the ICESCR Committee in relation to the communications are not formally binding, ratifying States and domestic courts, under pressure from NGOs and CSOs, may treat the decisions as authoritative. To date, the Protocol has been signed by thirty-six states but only ratified by three, well short of the ten ratifications needed in order to enter into force. If and when the Optional Protocol enters into force, it is anticipated that individual communications will be used to pressure States Parties to hold businesses accountable for not protecting or fulfilling economic rights.

4. *Strengthened the Role of NHRIs in Monitoring Business and Human Rights*. In 2009, the ICC established a Working Group on Business and Human Rights (the "Working Group"). The Working Group's purpose is to promote capacity building, strategic collaboration, advocacy, and outreach by NHRIs in the area of business and human rights. In 2010, at the Tenth International Conference of the ICC, the participating representatives from NHRIs adopted the Edinburgh Declaration (the "Declaration"). The Declaration sets forth the practical functions NHRIs can fulfill in promoting enhanced protection against corporate-related human rights abuse; greater accountability and respect for human rights by business actors; access to justice for victims; and establishing multi-stakeholder approaches.[9]

5. *Increased the Availability of Tools for Conducting Human Rights Impact Assessments*. Beginning in 2007, the International Finance Corporation of the World Bank, the UN Global Compact, and the International Business Leaders Forum engaged in a three-year road-testing process for a guide that provides practical advice to companies on how to identify and assess the human rights risks and impacts of their business activities, integrate the results into their management system, and ultimately improve their performance. In 2010, during the UN Global Compact Leaders Summit, the revised online version of the Guide to Human Rights Impact Assessment and Management ("HRIAM") was launched.

In 2008, the Danish Institute for Human Rights launched the Human Rights Compliance Assessment ("HRCA") tool. The HRCA is a comprehensive tool designed to detect human rights risks in company operations. It covers all internationally-recognized human rights and their impact on all stakeholders, including employees, local communities, customers and host governments. The HRCA tool incorporates a database of 195 questions and 947 indicators, each measuring the implementation of human rights in company policies and procedures. In 2010, HRCA 2.0 was released.

6. *Produced a UN Framework and Guiding Principles on Business and Human Rights*. In June 2008, after three years of extensive research and consultations with governments, businesses, and civil society, the Special Representative concluded that one reason cumulative progress in the business and human rights area had been difficult to achieve was the lack of an authoritative focal point around which actors' expectations could converge—a framework that clarified the relevant actors' responsibilities and provided the foundation on which thinking and action could build over time. In June 2008, the Special Representative presented such a framework to the UN Human Rights Council. The "Protect, Respect and Remedy" Framework rests on three pillars: the state duty to protect against human rights abuses by third parties, including business, through appropriate policies, regulation, and adjudication; the corporate responsibility to respect human rights, which means to act with due diligence to avoid infringing on the rights of others and to address adverse impacts that occur; and greater access by victims to effective remedies, both judicial and non-judicial.[10] In a June 2011 resolution, the Council endorsed the Special Representative's "Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework" (the "UN Guiding Principles"). The UN Guiding Principles explain the implications of existing human rights standards and practices for States and businesses; integrate them within a single, logically-coherent, and comprehensive template; and identify where the current normative and regulatory regime falls short and how it should be improved.[11] Each Principle is accompanied by a commentary, further clarifying its meanings and implications. The Council's resolution also established a Working Group on business and human rights consisting of five independent experts, the mandate of which includes promoting implementation of the UN Guiding Principles; providing advice regarding the development of domestic legislation and policies relating to business and human rights; conducting country visits; making recommendations for enhancing access to effective remedies for those whose human rights are affected by corporate activities; and guiding the work of the Council's new annual Forum on Business and Human Rights.

The Matrix Revolutions: Businesses Must Choose

Multinational businesses are facing a reloaded Matrix that has high expectations for their protection and fulfillment of economic rights. The Matrix will no longer be satisfied with "mere" business ethics, corporate philanthropy, corporate social responsibility, or environmental sustainability programs. The Matrix will no longer limit itself to pursuing claims for damages from egregious human rights violations. Instead, the Matrix will expect businesses to assess the impact their normal operations and policies have on the economic rights of others.

Multinational business enterprises need to choose whether to comply with demands of the Matrix. Specifically, businesses must be prepared to:

1. Consider carefully whether to conduct comprehensive and invasive HRIAs that, in essence, make human rights activists partners in corporate strategic planning and operations.

2. Decide whether to embrace the efforts of human rights activists to convert the UN Global Compact from a voluntary program that promotes best practices in the areas of the Ten Principles to a program that requires its members to comply with the detailed GRI Guidelines.

3. Decline offers by government-run NCPs to mediate unfounded NGO-instigated complaints of alleged corporate failures to protect or fulfill economic rights under the OECD Guidelines.

4. Wage effective media and other public education campaigns against NHRIs that engage in hearings, investigations, or reports designed to shame businesses for not protecting or fulfilling economic rights for which they have no legal responsibility.

5. Monitor the degree to which, in promoting the implementation of the UN Guiding Principles at the country level, the Council's Working Group on business and human rights interferes with national sovereignty.

A long-term objective of human rights activists is to generate court decisions, government agency or quasi-governmental rulings, international human rights treaty committee determinations, international organization instruments, and academic or other commentary that create "soft law" norms that can be used to hold multinational business enterprises accountable for protecting and fulfilling economic rights. Businesses will have to make some tough choices regarding whether to spend time and resources trying to help the Matrix achieve that objective or focus instead on the operation of their businesses in compliance with existing laws, thereby increasing the value of the investments made by their shareholders. In turn, individual shareholders could decide the degree to which, through their private philanthropy, political action, or support for corporate social responsibility, they can help others meet their basic economic needs.

Endnotes

1 James P. Kelly,, *The Matrix of Human Rights Governance Networks*, Engage, Feb. 2008, at 92, *available at* http://www.fed-soc.org/doclib/20080313_KellyGovernanceEngage9.1.pdf.

2 *Norms on the Responsibilities of Transnational Corporations and Other Business Enterprises with Regard to Human Rights*, U.N. Doc. E/CN.4/Sub.2/2003/12/Rev.2 (2003), *available at* http://www1.umn.edu/humanrts/links/norms-Aug2003.html.

3 John Ruggie, human rights impact assessments—Resolving Key methodological questions 6 (2007), *available at* http://www.reports-and-materials.org/Ruggie-report-human-rights-impact-assessments-5-Feb-2007.pdf.

4 Organisation for Econ. Co-operation & Dev., OECD Guidelines for Multinational Enterprises 34 (2008), *available at* http://www.oecd.org/dataoecd/56/36/1922428.pdf.

5 Alex Bennet & David Bennet, Organizational Survival in the New World 26 (2004).

6 *Id.*

7 *Id*. at 29.

8 Organisation for Econ. Co-operation & Dev., OECD Guidelines for Multinational Enterprises 29 (2011), *available at* http://www.oecd.org/dataoecd/43/29/48004323.pdf.

9 Int'l Co-ordinating Comm. of Nat'l Institutions for the Promotion & Prot. of Human Rights, The Edinburgh Declaration (2010), *available at* http://www.humanrightsbusiness.org/files/1098615587/file/edinburgh_declaration.pdf.

10 John Ruggie, Protect, Respect and Remedy, a Framework for Business and human Rights (2008), *available at* http://www.reports-and-materials.org/Ruggie-report-7-Apr-2008.pdf.

11 John Ruggie, Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework (2011), *available at* http://www.business-humanrights.org/media/documents/ruggie/ruggie-guiding-principles-21-mar-2011.pdf.

Labor & Employment Law

The States and the NLRB: A Study in Comparative Sovereignty

*By Thomas M. Christina**

Under a system of government that diffuses power and makes institutional "[a]mbition . . . counteract ambition,"[1] sudden power grabs by a federal agency are rare. Nevertheless, they do occur, particularly when they can be conducted "under the radar." A lawsuit can be a very successful means for launching a power struggle without arousing much public attention. As Justice Scalia famously observed, most lawsuits involving the allocation of governmental power arrive in court "clad, so to speak, in sheep's clothing: the potential of the asserted principle to effect important change in the equilibrium of power is not immediately evident, and must be discerned by a careful and perceptive analysis."[2]

National Labor Relations Board v. Arizona, pending in federal district court in Arizona, is one such case. The issue on the merits is whether the National Labor Relations Act ("NLRA" or "the Act") preempts an amendment to the Arizona constitution dealing with the right to vote by secret ballot in the election of bargaining representatives. This issue, however important it may be, is unlikely to attract widespread public attention. In fact, the controversy is unlikely to stir much interest even among lawyers. Preemption jurisprudence is a relatively narrow category of federal constitutional law, and the merits issue in *National Labor Relations Board v. Arizona* arises under only one branch of that jurisprudence. Moreover, the element of suspense is lacking, since the few observers who have taken notice of the preemption issue agree by and large that the National Labor Relations Board ("the Board") has the better case on the merits.

This paper does not deal with the merits of the preemption issue in *National Labor Relations Board v. Arizona*. Instead, it focuses on several procedural arguments the Board advanced (explicitly or implicitly) to convince the court to decide the merits of its preemption claim. The procedural issues in the case warrant serious professional and public attention, partly because of their novelty, partly because the district court ruled in the Board's favor on some of them,[3] but perhaps most significantly because they reflect a disturbing conception of federal-state relations and of the proper allocation of authority within the federal government. If the Board's positions on these procedural questions ultimately are sustained, *National Labor Relations Board v. Arizona* could have very troubling consequences for how governmental authority will be exercised in the future, not just over the states but over individual citizens, and in areas of the law far removed from labor relations.

Three procedural issues in the case deserve particular attention. First, the Board maintains that its request for a declaratory judgment presents a justiciable "case or controversy" for resolution by the district court. The arguments it advances in support of that contention burst the envelope of traditional standing doctrines, which generally prohibit judicial resolution of abstract legal questions at the request of a party with no legally cognizable interest of its own in the answer. Second, the Board effectively alleges that the very existence of the Arizona constitution's "secret ballot" provision is a wrong of constitutional dimension, which means that the relief the Board seeks necessarily is a direct interference with a sovereign function of a state that resulted from its internal political process. Interference of this kind can be justified in certain carefully-defined circumstances. However, the litigating position adopted by the Board would expand those circumstances in a way that would diminish public accountability of government officials to an unprecedented degree. Third, in pursuing its case against Arizona, the Board has advanced arguments that call for expanding the bounds within which so-called "independent regulatory agencies" can act. Unlike the Attorney General or other heads of executive branch departments, members of the Board are not subject to full supervision by the President because they cannot be removed from office at will. In bringing this lawsuit, the Board nonetheless implicitly asserts that it can decide unilaterally when and how the Supremacy Clause should be "enforced," and against which state laws. For the Board's assertion of its power to be upheld, a court would be required to decide that a so-called "independent regulatory agency" can make discretionary policy decisions pitting the national government against a state government without express congressional authorization and despite the President's limited authority to control how that power will be employed.

The Board clearly grasps the potential significance of its litigating position for augmenting the power of federal regulatory agencies and shielding the exercise of that power from control by an elected official. In its papers filed in opposition to Arizona's motion to dismiss the complaint, the Board describes itself not only as an "independent agency of the United States," but also as "a *sovereign* federal agency."[4] The Board presumably understands what sovereignty means, and that there can be only so much of it to go around in any given country. It must therefore also realize that judicial recognition of the Board's asserted status necessarily would alter the division of powers, the separation of powers, or both.

Despite the extraordinary nature of the Board's procedural arguments, it is fair to ask why any of those arguments matter. The traditional answer is that the Framers thought individual liberty would be protected both by federalism and by the Constitution's separation of federal powers among three distinct branches of government, one of which would be under

....

** J.D. Harvard Law School, 1980; Shareholder, Ogletree, Deakins, Nash, Smoak & Stewart, P.C. The author gratefully acknowledges the research and editorial assistance of his colleagues, Ashley Prickett Cuttino (J.D., University of South Carolina Law School, 2002), and James R. Silvers (J.D., Charleston School of Law, 2009). The opinions expressed in this paper are the author's alone and neither reflect any opinion nor are intended as the expression of any opinion on the part of Ogletree, Deakins, Nash, Smoak & Stewart, P.C., or any of its attorneys or staff.*

the control of a single individual.[5] The Supreme Court has elaborated on this thought by explaining the linkage among three concepts: protection of all the political processes that are necessary for constitutional federalism, the principle of accountability underlying the structural norms ordained by the Constitution, and the promotion of individual liberty.[6] The Board's litigating posture in *National Labor Relations Board v. Arizona* asks the federal courts to ignore each of these concepts.

Background and Context

To fully appreciate how the Board's procedural arguments in the case would enhance the powers of independent agencies and pose a threat to the individual liberties of citizens, it is first necessary to appreciate the extraordinary novelty of the Board's litigating position. That appreciation, in turn, requires some background.

The National Labor Relations Act of 1935 established certain rights and responsibilities in connection with employment in the private sector. One such right springs from Sections 7 and 9, which taken together generally require that a private-sector employer bargain over the terms and conditions of employment with a representative of his or her employees if the representative was "designated or selected for the purposes of collective bargaining by a majority of the employees in a unit appropriate for such purposes," and that such a representative will be the exclusive agent for bargaining purposes.[7] The Section 7 obligations of employers are generally enforceable under Section 8(a)(5), which makes it an "unfair labor practice" as defined in the Act for an employer "to refuse to bargain collectively with the representatives of his employees, subject to the provisions of section 159(a)."[8] Thus, a determination of what is sometimes called "majority status" logically precedes a determination of a duty to bargain.

The ordinary means for ascertaining whether a majority concurs in a proposition is to hold a referendum of some kind, usually involving each individual's expression of his or her sentiments, i.e., voting. Traditionally, the Board has expressed a decided preference for secret ballot elections conducted under the supervision of its agents.[9] At the same time, there also has been something of a running (if syncopated) dialogue at the federal level regarding the extent to which a labor organization's majority status might be or even must be recognized in the absence of an election under the auspices of the Board.[10] The Board's position is that Section 7 of the Act affords a second path to the consequences of majority status, which it describes as "voluntary recognition [by an employer] based on other convincing evidence of majority status."[11]

Cards purporting to authorize representation can be signed under a variety of circumstances, including home visitation by union organizers.[12] Thus, one driver of the debate over the "second path" described in the Board's Opposition to Arizona's motion to dismiss is concern over whether card signing results as reliably as secret ballot elections in the authentic expression of the individual's preference as he or she immediately perceives it.[13]

Over the decades, the mutual pushback between the Board and Congress changed the line of scrimmage but did not result in a victory for one school of thought over the other.[14] For example, Congress amended Section 9 in 1947 to provide that "[i]f the Board finds upon the record of [an unfair labor practice] hearing that . . . a question of representation exists, it shall direct an election by secret ballot and shall certify the results thereof."[15] Based in part on the language of this amendment, the Supreme Court held in *Linden Lumber Division, Summer & Co. v. NLRB* that "unless an employer has engaged in an unfair labor practice that impairs the electoral process, a union with authorization cards purporting to represent a majority of the employees, which is refused recognition, has the burden of taking the next step in invoking the Board's election procedure."[16]

Yet, notwithstanding the 1947 amendment of Section 9(c) and the decision in *Linden Lumber*, the Board has determined that majority status can be found in some scenarios based on the presentation of authorization cards bearing signatures equal to 50% plus one of the number of individuals in a bargaining unit. Arguably, this represents a dilution of the requirement for "*convincing* evidence" of majority status that the Board invoked in its papers in *National Labor Relations Board v. Arizona*. Elsewhere in its Proposed Opposition to the Motion to Dismiss in *National Labor Relations Board v. Arizona,* for example, the Board seems to blur the distinction between "evidence" and "convincing evidence" by asserting that the Section 7 right to representatives of their own choosing is enforceable "[i]f private sector employees can persuade their employer to recognize their choice of a representative on the basis of evidence of majority status."[17] A standard based on "evidence" means less than one might think, because an employer need not actually adjudicate majority status based on "evidence," whether or not accompanied by persuasive argument regarding how to interpret or weigh the evidence. To the contrary, an employer and a union can agree in advance that majority status will be recognized based on the union's accumulation of signed cards. The Board has generally held that neutrality and card-check agreements between a union and an employer are enforceable.[18] As a result, combined neutrality and card-check agreements have eclipsed elections as a means of organizing new bargaining units.[19]

* * *

On November 2, 2010, against this general backdrop, voters in four states (Arizona, South Carolina, South Dakota, and Utah) adopted a "secret ballot" amendment to their state constitutions, each worded somewhat differently from the others. The amendment to Arizona's constitution provides that "[t]he right to vote by secret ballot for employee representation is fundamental and shall be guaranteed where local, state or federal law permits or requires elections, designations or authorizations for employee representation."[20]

On January 13, 2011, the Board's Acting General Counsel wrote to the Attorney General of each of the four states "to apprise [them] of the National Labor Relations Board's conclusion that a recently-approved amendment to [the state's constitution] conflicts with the rights afforded individuals covered by the National Labor Relations Act . . . [and] . . . to explain the Agency's position and to advise you that I have been authorized to bring a civil action in federal court to seek to invalidate the [a]mendment."[21] In a Fact Sheet issued on

January 14, 2011, the Board took a slightly more modest position, stating that on January 6, 2011, it had authorized the Acting General Counsel to bring lawsuits against the states to enjoin the application or enforcement of the states' secret ballot amendments "insofar as they conflict with the federal rights of private sector employees to designate a union to represent them."[22]

The Attorneys General responded to the Acting General Counsel in a single letter dated January 27, 2011. They flatly rejected the invitation to concede that portions of their state constitutions were invalid, and pointed out ways in which the substance of the secret ballot amendments was consistent with the Board's interpretation of the NLRA and with the freedom of association guaranteed by the First Amendment.[23] In response, the Board's Acting General Counsel offered to hold discussions at the staff level on the issues.[24] However, these discussions came to an immediate and abrupt halt when the Acting General Counsel's staff refused to discuss the merits of the Board's position in the absence of a confidentiality and non-disclosure agreement.[25] The Attorneys General made clear that they would not hold discussions regarding the validity of state constitutional provisions behind the backs of their fellow-citizens (to say nothing of their governors and state legislatures).[26]

On April 22, 2011, the Board's Acting General Counsel wrote to the Attorneys General again, this time to say that he had "directed [his] staff to initiate lawsuits in federal court seeking to invalidate Arizona Constitution Article 2 § 37 and South Dakota Constitution Article 6 § 28 as preempted by operation of the NLRA . . . and the Supremacy Clause. . . ."[27] However, the Board followed through on the Acting General Counsel's threatened legal action by singling out Arizona as the first target. On May 6, 2011, the Board filed its complaint "seek[ing] a declaratory judgment . . . that Arizona Constitution Article 2 § 37 . . . conflicts with the rights of private sector employees under the National Labor Relations Act . . . and is therefore preempted."[28]

The Standing Issue

Whether a state law is preempted by federal law is an abstract question unless and until some regulated party is actually affected by the alleged conflict between federal and state law. For this reason, it is one thing for a federal agency to announce its views on whether a state law is preempted by federal law, but quite another to seek the endorsement of those views by a federal court.

In fact, the Constitution imposes general limitations on the circumstances in which federal courts can rule on a matter of law, and some of those limitations apply to any would-be plaintiff, including a federal agency. To issue a ruling deciding a contested issue of law (or of fact, for that matter), a federal court first must have jurisdiction over a "case or controversy," i.e., an actual dispute between parties whose legally cognizable rights and/or obligations with respect to each other actually might be altered by how the court decides the legal issue presented to it.[29]

Although the district court ruled against Arizona on the standing concerns it raised in its motion to dismiss, a powerful argument can be made that *National Labor Relations Board v. Arizona* does not meet the "redressability" component of the traditional standing doctrine. The Board and Arizona will stand in precisely the same place with respect to their own labor law rights and obligations vis-à-vis each other regardless of any possible outcome in the Board's lawsuit. No matter what the outcome of the case may be on the merits, the district court's decision cannot determine Arizona's rights and obligations in its capacity as an employer under the Act, because the term "employer" is defined in the NLRA to *exclude* states.[30] Nonetheless, the district court held that "[a] declaratory judgment that Article 2 § 37 is preempted would redress plaintiff's injuries by rendering the amendment unenforceable and affirming plaintiff's exclusive power to enforce § 7 and § 8 of the NLRA."[31] However, as a general rule, "[i]t is a violation of due process for a judgment to be binding on a litigant who was not a party or a privy and therefore has never had an opportunity to be heard."[32] Thus, a declaratory judgment in *National Labor Relations Board v Arizona* to the effect that its secret ballot amendment is preempted will not by its own force bar a single employer or employee from invoking it in any subsequent case.[33]

Despite the seeming futility of the judgment the Board seeks in this case, it would be a mistake to dismiss the Board's decision to sue as harmless. That decision is an assault on important principles that keep the power of the federal judiciary, legislature, and executive separated. To be sure, one important purpose of the "case or controversy" requirement is to limit the authority of the judiciary alone by preventing the federal courts from deciding abstract generalities outside the context of adjudicating specific rights between specific parties.[34] However, it also serves three broader purposes, each of which promotes the ideals of self-government by avoiding instances in which distinct constitutional powers might be blended, thereby reducing the accountability of elected officials. First, it deters the political branches from running to the judiciary for political cover or moral reinforcement, for example by seeking an advisory opinion on the constitutionality of pending legislation, thereby closing a route by which elected officials might escape accountability to the voters. Second, it limits the circumstances under which the judiciary might accept an invitation to shut down popular or political debate before a majority has coalesced around a specific proposal or course of action, thereby forestalling at least one means by which the processes of representative democracy might be short-circuited. Third, it prevents the courts from deciding what Chief Justice Marshall called a "mere question of right" at the behest of a petitioner who requests the judiciary "to control the legislature" of a state, a dubious enterprise that Chief Justice Marshall observed "savours too much of the exercise of political power to be within the proper province of the judicial department."[35]

National Labor Relations Board v. Arizona does not arise from a "case or controversy" under established principles of law, because the Board's own rights and obligations are not at stake in the case.[36] In fact, the Board essentially admits as much by arguing that it has standing to vindicate the federal statutory rights of workers in Arizona under a doctrine called "*parens patriae*," under which a sovereign can bring suit to assert the rights of its citizens.[37]

Endorsing the Board's *parens patriae* argument would extend that doctrine far beyond its current or historical boundaries, because the Board does not have "citizens," i.e., persons whose interests it is authorized to represent as against other governments. At best, the Board has various "constituencies," but only in the sense of that word when used to describe interest groups. In any event, the Board itself does not take seriously the notion that it is trying to vindicate the rights of individuals employed in Arizona: it has opposed a motion to intervene in the case by dozens of such individuals on the grounds that the State of Arizona adequately represents the interests of its citizens.[38]

There is a second way in which the Board's *parens patriae* argument could set a dangerous precedent. The NLRA establishes a wide range of rights and obligations on the part of employers, employees, and employee organizations. These rights and obligations are cast in general terms, which means that when it comes to the exercise of these rights by their holders, interests may differ. For example, the Act establishes an individual right to refrain from forming, joining, or assisting a labor organization.[39] It also establishes a right to engage in "concerted activities," including those that are alternatives to collective bargaining, for "mutual aid or protection."[40] It seems to follow that employees have a right to collaborate in resisting the selection of a supposed "bargaining representative" if they consider the method of selection illegitimate.

It is impossible to act in any representative capacity for a group consisting of members with conflicting interests unless the representative is authorized to ignore or subordinate (that is, to sacrifice) the interests of certain members of the group in favor of others. Thus, the Board's *parens patriae* argument necessarily entails the assertion that it may determine *which* of various abstract statutory rights to protect at the expense of others (a process that would ultimately call for determining that one portion of its organic statute is not to be advanced in light of another).[41] A federal regulatory agency endowed with this authority genuinely would be "sovereign" as opposed to merely "independent"—and not in a good way.

Separation of Powers and the Unitary Executive

The implications of the Board's *parens patriae* argument serve as an introduction to a separation of powers issue implicit in the Board's litigating posture. This issue becomes apparent only when one realizes that an agency not directly accountable to the President has brought an action against a state seeking to invalidate a particular state law as repugnant to the Federal Constitution without any established statutory authorization to bring such a suit. Viewed in that light, the assertion of agency power underlying the complaint in *National Labor Relations Board v Arizona* is genuinely revolutionary because it can be justified, if at all, only by some non-statutory authority to perform the quintessential executive function—that is, to "see" that the Supremacy Clause "be faithfully executed." The Board has no such authority, nor could it.

The NLRB's statutory authority to litigate is confined to two areas. Under various subsections of the act, it can sue to enforce its orders and/or to prevent an unfair labor practice, and it can seek enforcement of its subpoenas in a federal court.[42] Since a state is not a labor organization, a state cannot commit an unfair labor practice as defined by Section 8(b), the Section relating to unfair labor practices by labor organizations. Moreover, since the term employer is defined in Section 2(2) to exclude the states, a state cannot commit an unfair labor practice as defined in Section 8(a), the Section relating to unfair labor practices by an employer. Thus, it does not appear that the NLRB has statutory authority to bring litigation of the kind contemplated here.[43]

But imagine that the Board had such a power. Its exercise necessarily would involve discretion regarding the enforcement of federal law. "The authority to bring such suits [i.e., suits to enforce federal law as *parens patriae*] includes the discretionary authority not to bring them, if the responsible officers of the government are of the opinion that a suit is not warranted or would be of disservice to the national interest."[44] The exercise of discretion in the enforcement of the law is traditionally thought of as a quintessentially executive function. Indeed, *National Labor Relations Board v. Arizona* involves at least two individual examples of the Board's exercise of executive discretion. First, it chose to pursue its case against Arizona but not against any of the other states it threatened to sue. Second, it chose to promote what it perceives to be the interests of Arizona employees who might wish to have a bargaining representative recognized through alternatives to secret ballot elections and to oppose the interests of Arizona employees who have a right under the Act not to engage in any concerted activity. Yet federal executive power involving such a degree of discretion may be wielded only by the President or an executive officer of the United States, and an executive officer of the United States must be removable at will by the President or by someone who the President can remove at will.[45]

To be sure, so-called "independent regulatory agencies" are permitted to play a role in the administration of federal law, despite the President's inability to remove the heads of those agencies at will. However, this exception to the normal rule for the exercise of executive power applies only to the conclusion (or at any rate the polite fiction) in *Humphrey's Executor* that what the independent agencies do is entirely "quasi-legislative" or "quasi-judicial."[46]

Judging from the majority opinion written by Chief Justice Roberts in *Free Enterprise Fund v. Public Co. Accounting Oversight Board*, the Supreme Court may be losing its enthusiasm for the reasoning behind *Humphrey's Executor*. To begin with, Chief Justice Roberts leaves no room for doubt that the President's removal power stems from the Faithful Execution Clause, not solely from the Appointments Clause. Moreover, the majority opinion in *Free Enterprise Fund* studiously avoids even using the terms "quasi-legislative" or "quasi-judicial" except in quotations.[47] In any event, to recognize that an independent agency's assertion of discretionary authority to seek declaratory judgments regarding the constitutionality of state laws is at best constitutionally dubious, it suffices to observe that no court has ever tried to justify agency action on the grounds that it was merely "quasi-executive." To do so would be to diminish the authority (and therewith the accountability) of the only federal officeholder whose authority is conferred by the people of the United States as a whole. A concept even remotely approaching

a diminution of that authority by transferring some part of it to an agency has no place in the constitutional design.

Federalism

And so we come to the third constitutional novelty in the Board's litigating position. The Board implicitly asserts that the very existence of Arizona's secret ballot amendment gives rise to a "case or controversy" and triggers the Board's authority to approach a federal court for declaratory relief. This assertion is based in part on the possibility that the amendment might be invoked affirmatively or defensively to support or defeat a claim. However, the assertion also is based on the Board's contention that "the enactment of the Amendment has a . . . chilling effect on the free exercise of organizational rights."[48]

In a letter dated January 13, 2011, to the Arizona Attorney General, the Acting General Counsel's office elaborated on its concerns about the potential chilling effect of the Arizona amendment. The letter makes clear that the Board's concern applies not only when the amendment became law, but even before it became law:

> I understand that the Amendment adopted by the voters in November is not technically in effect and must still be proclaimed by the Governor of Arizona. A.R.S. Const. Art 4 Pt. 1 § 1(5). Accordingly, I am hopeful that, after a review of the NLRB's legal position, Arizona might be willing to take voluntary measures to ensure that the Amendment will not be proclaimed, and that the public will be so notified. . . .
>
> In light of the significant impact of this Amendment, I request that any response to this letter on behalf of Arizona be made within two weeks. Absent any response, I intend to initiate the lawsuit.
>
> In similar situations, where offending enactments have not yet ripened into actual enforcement actions, the courts have nonetheless permitted suits to bar their enforcement where a danger exists that public knowledge of the provision may result in "self-censorship; a harm that can be realized even without an actual prosecution." . . . That principle is applicable here since it is foreseeable that widespread public knowledge of the Amendment will deter some employers from granting voluntary recognition or abiding by their commitments to recognize a union on the basis of a card check.[49]

On a chilling effect theory as expansive as the one advocated by the Board, there would seem to be no reason in principle why the Board could not seek injunctive relief to preclude a referendum on a state constitutional amendment or to preclude legislative action to propose such an amendment. (Ripeness considerations might convince a court not to proceed in such a case, but ripeness is generally regarded as a prudential doctrine, not a constitutional one, so the court would have the authority to decide whether or not to stay its hand.) If the courts endorsed this model of agency authority, the political processes of state and local governments would continue solely at the sufferance of a host of federal agencies—some accountable to an elected official, others not—each with what amount to censorial powers, including the power of discretionary interdiction. The Framers declined to give that power even to the federal courts. There can be no doubt that they would have denied it to an "independent agency."

Republican Sovereignty

The Board is right about one thing: in the last analysis, *National Labor Relations Board v. Arizona* is a case about sovereignty. For those who believe that the residual sovereignty of the states includes certain immunities from federal interference in their political and/or governmental processes,[50] the case represents either a new threat or a new opportunity. Which it turns out to be depends at least in part on whether the courts will recognize that the case raises both a grave federalism issue and an equally important separation of powers question. Recognizing that *National Labor Relations Board v. Arizona* involves both of the basic structural underpinnings of republican democracy could lead a court to decide the case on the basis of the general accountability principles that are the constitutional postulates of both federalism and the separation of powers. Those principles of accountability are vital to the sovereignty of republics.[51]

One reason to believe courts are basing their decisions on these principles can be found in a recent ruling from a district court, which coincidentally also deals with sovereignty as a limitation on the Board's authority. *Chickasaw Nation v. National Labor Relations Board*[52] arose from two unfair labor practice charges brought by the Board against a federally-recognized Indian tribe, with respect to a casino operated under the auspices of the Chickasaw Nation's executive branch.

A federally-recognized Indian tribe is "a domestic, dependent nation" under federal law, and therefore a type of sovereign.[53] On that ground, the Chickasaw Nation brought an action in federal court alleging that the NLRA has no application to it because of its sovereign status as a tribe, and moved for injunctive relief to restrain the Board from conducting further proceedings on the unfair labor practice charges.[54] The plaintiff argued that allowing the Board to exercise jurisdiction over the Chickasaw Nation would unlawfully interfere with its inherent powers as a sovereign, including its right to make its own laws and be governed by them.[55]

The Board's principal defense did not engage the sovereignty issue directly. Instead, it relied on the "primary jurisdiction" doctrine, under which the federal courts decline to exercise their power to determine in the first instance whether a given controversy falls within the scope of an independent agency's adjudicatory purview.[56] In effect, the Board argued that the district court should decline to render any judgment and defer to the Board's determination of the scope of its own authority with regard to the unfair labor practice charges against the casino operation.[57] In support of this argument, the Board pointed to the provision of the NLRA which the Board urged gives it "[power] to prevent any person from engaging in any unfair labor practice . . . [unaffected] by any other means of adjustment or prevention . . . established by . . . law." [58]

The district court acknowledged that the NLRA's grant of the power to prevent unfair labor practices had been interpreted to limit the authority of federal district courts to

enjoin proceedings before the Board.[59] However, the court held that federal regulatory schemes such as the NLRA do not apply to tribal governments exercising their sovereign authority absent express congressional authorization.[60] It also found that the NLRA makes no specific reference to tribes and the legislative history does not indicate an intention to abrogate the sovereignty of recognized tribes by subjecting their executives to regulation under the Act.[61]

Chickasaw Nation v. National Labor Relations Board will not eliminate all agency over-reaching, but, along with *National Labor Relations Board v. Arizona*, it is an opportunity to establish a limitation on such conduct. To be sure, the limitation would come into play only in the rare event that agency action threatened an immunity that is a component of the residual sovereignty of a state or of the provisional sovereignty of a federally-recognized tribe. Nonetheless, for close to a century, it has been regarded as a major achievement to establish *any* limitation on the power of an independent agency. Moreover, whether the states' status as sovereigns is a limitation on the power of independent agencies is merely a specific form of a more general (and more pressing) question: whether liberty-reinforcing and accountability-reinforcing constitutional norms are limitations on the extent of otherwise ungoverned agency power.

Endnotes

1 The Federalist No. 51 (James Madison).

2 Morrison v. Olson, 487 U.S. 654, 699 (1988) (Scalia, J., dissenting).

3 *See* Order, Nat'l Labor Relations Bd. v. Arizona, No. 2:11-cv-913, Dkt. No. 18, at 2-5, and 8-10.

4 *See* National Labor Relations Board's Opposition to Motion to Dismiss, Nat'l Labor Relations Bd. v. Arizona, No. 2:11-cv-913, Dkt. No. 13, 2-40, at 3, 16 [hereinafter Proposed Opposition] (emphasis added). The Board's Proposed Opposition exceeded the page limitations applicable in the Arizona federal court. The Board lodged the Proposed Opposition with the court clerk (who maintains it as part of the docket in the case) and submitted a motion for leave to file an Opposition exceeding the court's page limitations. Although the Board's motion was denied, the Proposed Opposition spells out the Board's arguments in more detail than the shorter brief it ultimately filed. For that reason, this paper relies to a great degree on the Board's Proposed Opposition.

5 Printz v. United States, 521 U.S. 898, 920, 922-23 (1997); U.S. Term Limits v. Thornton, 514 U.S. 779, 838 (1995); New York v. United States, 505 U.S. 144, 168-69 (1992).

6 *See, e.g.,* Bond v. United States, 131 S. Ct. 2355, 2364 (2011) ("The Framers concluded that allocation of powers between the National Government and the States enhances freedom, first by protecting the integrity of the governments themselves, and second by protecting the people, from whom all governmental powers are derived."); Free Enter. Fund v. Pub. Co. Accounting Oversight Bd., 130 S. Ct. 3138, 3154 (2010) (holding that where an executive officer of the United States is removable only for cause and only by other executive officers protected from removal at will by the President, the latter "can neither ensure that the laws are faithfully executed, nor be held responsible for [the officer's] breach of faith").

7 29 U.S.C. § 159(a).

8 *Id.* § 158(a)(5).

9 *Cf.* Nat'l Labor Relations Bd. v. Gissel Packing Co., 395 U.S. 575, 596 (1969) ("The most commonly traveled route for a union to obtain recognition as the exclusive bargaining representative of an unorganized group of employees is through the Board's election and certification procedures under § 9(c) of the Act (29 U.S.C. § 159(c)); it is also, from the Board's point of view, the preferred route."). By contrast, the preference for elections has come under sharp academic criticism. *See, e.g.,* Joel Dillard & Jennifer Dillard, *Fetishizing the Electoral Process: The National Labor Relations Board's Problematic Embrace of Electoral Formalism*, 6 Seattle J. Soc. Just. 819, 819 (2008).

10 *See* Sheila Murphy, *A Comparison of the Selection of Bargaining Representatives in the United States and Canada:* Linden Lumber, Gissel, *and the Right to Challenge Majority Status*, 10 Comp. Lab. L.J. 65, 69-70 (1988) (chronicling decisions before 1947).

11 Proposed Opposition, *supra* note 3, at 2.

12 *See, e.g.,* Congressional Documents, *Senator Specter Speaks on Employee Free Choice Act/Card Check* (Mar. 24, 2009) ("Testimony shows union officials visit workers' homes with strong-arm tactics and refuse to leave until cards are signed.").

13 *See, e.g.,* Nat'l Labor Relations Bd. v. Village IX, Inc., 723 F.2d 1360, 1371 (7th Cir. 1983) ("Workers sometimes sign union authorization cards not because they intend to vote for the union in the election but to avoid offending the person who asks them to sign, often a fellow worker, or simply to get the person off their back. . . ."). For a survey of the arguments pro and con, see A. E. Eaton & J. Kriesky, *NLRB Elections versus Card Check Campaigns: Results of a Worker Survey*, 62 Indus. & Lab. Rel. Rev. 157 (2009).

14 The Board itself has changed its views regarding the reliability of non-electoral expressions of an employee's preference. In *Dana Corp.*, 351 N.L.R.B. 434, 439 (2007), the majority of the Board found that "There is good reason to question whether card signings . . . accurately reflect employees' true choice concerning union representation," and held that for a period of forty-five days following an employer's voluntary recognition of a bargaining agent, employees could file a decertification petition supported by a showing of interest by 30% of the bargaining unit. Four years later, the Board overruled *Dana Corp.*, based in part on the assertion that empirical evidence had demonstrated that the majority opinion in *Dana Corp.* regarding the reliability of card-check "was wrong." Lamons Gasket Co., 357 N.L.R.B. No. 72, 2011 WL 3916075 (N.L.R.B.) *6 (Aug. 26, 2011).

15 29 U.S.C. § 159(c)(1).

16 419 U.S. 301, 310 (1974).

17 Proposed Opposition, *supra* note 3, at 2 (citing NLRB v. Creative Food Design, 852 F.2d 1295, 1297-98 (D.C. Cir. 1988); NLRB v. CAM Indus., Inc., 666 F.2d 411, 414 (9th Cir. 1982). The difference between "evidence" and "convincing evidence" is particularly important because "evidence" is a term of art referring to anything that tends to make an assertion more likely or less likely to be true, i.e., to make the probability of the proposition's accuracy different to any degree whatsoever. *Cf., e.g.,* Fed. R. Evid. 401 (defining relevant evidence). Despite appearing to say in its Opposition that mere evidence can support a determination of majority status, the Board obviously recognizes that evidence of "majority support" must be "reliable." Complaint, ¶ XII.

18 Joseph Z. Fleming & Daniel B. Pasternak, *Mutuality Agreements: Innovative Approaches to the Use of Neutrality Agreements—A Unique Proposal for Compromise*, *in* Employment and Labor Relations Law for the Corporate Counsel and the General Practitioner 5 (Am. Law Inst. ed. 2007); *see* Kroger Co., 219 N.L.R.B. 388 (1975) (seminal case concerning card-check recognition agreements); SL DC Mgmt., LLC d/b/a Hotel Del Coronado, 345 N.L.R.B. 306 (2005); *In re* Central Parking System, Inc., 335 N.L.R.B. 390 (2001); Charles I. Cohen, Joseph E. Santucci, Jr., & Jonathan C. Fritts, *Resisting Its Own Obsolescence—How the National Labor Relations Board Is Questioning the Law of Neutrality Agreements*, 20 Notre Dame J.L. Ethics & Pub. Pol'y 521, 521 (2006) ("The NLRB . . . has been willing to defer to private agreements that resolve union representation matters, rather than deciding the representation question through a Board-supervised secret ballot election.").

19 Cohen, Santucci & Fritts, *supra* note 18, at 522.

20 Ariz. Const. art. 2, § 37. The amendments adopted by voters in South Carolina, South Dakota, and Utah each were worded somewhat differently. *See* S.C. Const. art. 2, § 12 ("The fundamental right of an individual to vote by secret ballot is guaranteed for a designation, a selection, or an authorization for employee representation by a labor organization."); S.D. Const. art. 6, § 38 ("The rights of individuals to vote by secret ballot is fundamental. If any

state or federal law requires or permits an election for public office, for any initiative or referendum, or for any designation or authorization of employee representation, the right of any individual to vote by secret ballot shall be guaranteed."); Utah Const. art. 4, § 8 ("All elections, including elections under state or federal law for public office, on an initiative or referendum, or to designate or authorize employee representation or individual representation, shall be by secret ballot.").

21 Letter from Hon. Lafe Solomon, Acting General Counsel, by Eric G Moskowitz, Assistant General Counsel, to Hon. Tom Horne, Attorney General of Arizona (Jan. 13, 2011), *available at* https://www.nlrb.gov/sites/default/files/documents/234/letter_az.pdf; Letter from Hon. Lafe Solomon, Acting General Counsel, by Eric G Moskowitz, Assistant General Counsel, to Hon. Alan Wilson, Attorney General of South Carolina (Jan. 13, 2011), *available at* https://www.nlrb.gov/sites/default/files/documents/234/letter_sc.pdf; Letter from Hon. Lafe Solomon, Acting General Counsel, by Eric G Moskowitz, Assistant General Counsel, to Hon. Marty Jackley, Attorney General of South Dakota (Jan. 13, 2011), *available at* https://www.nlrb.gov/sites/default/files/documents/234/letter_sd.pdf; Letter from Hon. Lafe Solomon, Acting General Counsel, by Eric G Moskowitz, Assistant General Counsel, to Hon. Mark Shurtleff, Attorney General of Utah (Jan. 13, 2011), *available at* https://www.nlrb.gov/sites/default/files/documents/234/letter_ut.pdf.

22 Nat'l Labor Relations Bd., State Constitutional Amendments Conflict With the NLRA (2011).

23 Letter from Hon. Alan Wilson, Attorney General of South Carolina; Hon. Mark L. Shurtleff, Attorney General of Utah; Hon. Tom Horne, Attorney General of Arizona; and Hon. Marty J. Jackley, Attorney General of South Dakota to Hon. Lafe Solomon, Acting General Counsel, National Labor Relations Board (Jan. 27, 2011) (citing, *inter alia,* Abood v. Detroit Bd. Of Educ., 431 U.S. 209, 233-35 (1977)), *available at* http://attorneygeneral.utah.gov/cmsdocuments/nlrb012711.sol.pdf.

24 Letter from Hon. Lafe Solomon to Hon. Alan Wilson, Attorney General of South Carolina; Hon. Mark L. Shurtleff, Attorney General of Utah; Hon. Tom Horne, Attorney General of Arizona; and Hon. Marty J. Jackley, Attorney General of South Dakota (Feb. 2, 2011), *available at* https://www.nlrb.gov/sites/default/files/documents/234/feb_2_letter.pdf ("As you have unanimously expressed the opinion that the State Amendments can be construed in a manner consistent with federal law, I believe your letter may provide a basis upon which this matter can be resolved without the necessity of costly litigation. My staff will shortly be in contact . . . to explore this issue further.").

25 David Montgomery, *Feds May Sue over Secret Ballot Vote*, Rapid City J., Mar. 18, 2011, *available at* http://rapidcityjournal.com/news/article_7bdaec7a-50f6-11e0-83c7-001cc4c002e0.html.

26 Letter from Hon. Alan Wilson, Attorney General of South Carolina; Hon. Mark L. Shurtleff, Attorney General of Utah; Hon. Tom Horne, Attorney General of Arizona; and Marty J. Jackley, Attorney General of South Dakota to Hon. Lafe Solomon, Acting General Counsel, National Labor Relations Board (Mar. 4, 2011), *available at* http://www.scag.gov/wp-content/uploads/2011/03/AGs-03-04-11-letter-to-Solomon.pdf.

27 Letter from Hon. Lafe Solomon to Hon. Tom Horne, Attorney General of Arizona; Hon. Alan Wilson, Attorney General of South Carolina; Hon. Marty J. Jackley, Attorney General of South Dakota; and Hon. Mark L. Shurtleff, Attorney General of Utah (Apr. 22, 2011), *available at* https://www.nlrb.gov/sites/default/files/documents/234/april_22_letter_from_gc_to_states.pdf.

28 Complaint at 1, Nat'l Labor Relations Bd. v. Arizona (D. Ariz. May 6, 2011).

29 *Cf.* Massachusetts v. Mellon, 262 U.S. 447, 480 (1923) (holding that a proceeding does not fall within the Supreme Court's original jurisdiction merely because a state is a party to the proceeding, but only where the state is a party to "a proceeding of judicial cognizance"); Raines v. Byrd, 521 U.S. 811 (1997) (holding that despite alleging an institutional injury (namely a loss of the legislature's power), Members of Congress as such lack standing to challenge the constitutionality of a federal statute to which they were opposed).

30 29 U.S.C. § 152(2).

31 Order, *supra*, Dkt. No. 18, at 6.

32 Parklane Hosiery Co., Inc. v. Shore, 439 U.S. 322, 327 n.7 (1979) (citing Blonder-Tongue Laboratories, Inc. v. Univ. of Ill. Found., 402 U.S. 313, 329 (1971) and Hansberry v. Lee, 311 U. S. 32, 40 (1940)).

33 As discussed in more detail below, the Board asserts that the very existence of the amendment has a chilling effect on the exercise of workers' rights under the Act. In this case, it may be immaterial whether the Board can marshal admissible evidence to prove this contention; since the Board did not name the Arizona Secretary of State as a defendant, the district court cannot effectively order that records of the amendment's due adoption be burned or that its text be obliterated. *See* Ariz. Const. art. IV, pt. I, § 1(13) (setting out governor's duty to proclaim the results of voting on proposed amendment to constitution, declaring such amendments as approved if voted for by a majority of electors voting); Ariz. Rev. St. 41-121(A)(2) (requiring the Secretary of State to "keep a register of and attest the official acts of the governor."). Thus, in a very important sense, Arizona's secret ballot amendment will remain "on the books" after the lawsuit to precisely the same extent as before the lawsuit.

34 Georgia v. Stanton, 73 U.S. 50, 75 (1867) ("This court can have no right to pronounce an abstract opinion upon the constitutionality of a state law. Such law must be brought into actual, or threatened operation upon rights properly falling under judicial cognizance, or a remedy is not to be had here.").

35 Cherokee Nation v. Georgia, 30 U.S. 1, 20 (1831).

36 In its analysis of the standing question, the district court invoked the three-part test in *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560-61 (1992), i.e., according to which a plaintiff must show (1) that it suffered an injury in fact that is concrete and particularized and actual or imminent; (2) that there is a causal connection between the injury and the conduct complained of that is fairly traceable to defendant's action; and (3) the likelihood that the plaintiff's injury can be redressed by a decision in its favor. Order, *supra*, Dkt. 18, at 2. The district court found that the amendment to Arizona's constitution caused an "injury" to the Board, and "affects plaintiff's legal rights by undermining its exclusive authority to administer the NLRA." *Id.*, at 6. This finding appears to extend the *Lujan* test for standing beyond its original scope, because *Lujan* defines an injury for purposes of that test as "an invasion of a legally protected interest." *Lujan*, 504 U.S. at 561. An adjudicatory body such as the Board has jurisdiction to resolve disputes, but that jurisdiction is more properly described as a power than an interest. Moreover, the delegation by Congress to an administrative agency of exclusive authority to enforce a federal law does not create an "interest" in that authority in the traditional sense of that word, for example, the way that a federal land grant creates a property interest on the part of a grantee. If Congress subsequently narrowed a prior delegation of enforcement authority, a court would be unlikely to decide that the agency as such had standing to seek a declaratory judgment questioning the constitutionality of the cut-back.

It is interesting to note that the Board's complaint is not predicated on an injury to anything it identifies as its own legally-protected interest. In its Opposition, the Board raised for the first time an argument that Arizona's secret ballot amendment "interferes with the Board's own activities or operations," but this argument amounts to suggesting that uncertainty will generate litigation or prolong disputes, which in turn will diminish the Board's ability "to meet its performance goals." Proposed Opposition, *supra* note 3, at 12. There is a large element of bootstrapping inherent in this argument, since the "performance goals" the Board refers to are ones the Board sets for itself. *See* Nat'l Labor Relations Bd., Fiscal Year 2010 Performance and Accountability Report, page prior to page 1, *available at* http://www.nlrb.gov/sites/default/files/documents/189/par2010.pdf. ("The Performance section compares the NLRB's performance to its annual performance goals. . . .").

37 *See* Alfred L. Snapp & Son, Inc. v. Puerto Rico *ex rel.* Barez, 458 U.S. 592, 600-07 (1982) (explaining the *parens patriae* doctrine).

38 NLRB's Opposition to Motion to Intervene of Save Our Secret Ballot and 34 Individuals at 3, Nat'l Labor Relations Bd. v. Arizona (D. Ariz. 2011) (citing and quoting Arakaki v. Cayetano, 324 F.3d 1078, 1086 (9th Cir. 2003).

39 29 U.S.C. § 157.

40 *Id.*

41 Six weeks after filing its Proposed Opposition Brief, the Board issued

a decision denying the legitimacy of its favoring one set of Section 7 rights over another. *See* Lamons Gasket Co., 2011 WL 3916075 (N.L.R.B.) at *8-9 (reasoning that because the Board's decision in *Dana Corp.* required posting notices of decertification petition rights, it "placed the Board's thumb decidedly on one side of what should be a neutral scale," and holding that "[s]uch administrative action is not appropriate under the Act").

42 *See* 29 U.S.C. §§ 160-61.

43 The Board contends, and the district court agreed, that the Supreme Court's decision in *Nat'l Labor Relations Bd. v. Nash-Finch Co.*, 404 U.S. 138, 144-47 (1971), provides that authority. It does not. The case simply holds that in an action *against an employer* involving a specific unfair labor practice, the Board is authorized to seek a remedy—even if the remedy is prohibited by ostensibly applicable state law—if the state law is in turn preempted by the NLRA. The invalidation of the state law (actually, state court injunction) in *Nash-Finch* was merely a necessary step in the enforcement of Section 8 of the Act in a case the NLRB was statutorily authorized to bring and over which the federal courts had a basis for jurisdiction. *Nash-Finch* does not stand for the proposition that the NLRB has the authority to sue a state seeking a declaration of a law's invalidity as an end in itself, and certainly does not stand for the proposition that the Board can pick and choose which of various actually- or supposedly-preempted statutes to sue over and which to leave be.

44 Georgia v. Penn. R.R. Co., 324 U.S. 439, 474 (1945) (Stone, C.J., dissenting).

45 Free Enter. Fund v. Pub. Co. Accounting Oversight Bd., 130 S. Ct. 3138, 3153 (2010); *see also* Myers v. United States, 272 U.S. 52, 164 (1926).

46 Humphrey's Executor v. United States, 295 U.S. 602, 620 (1935).

47 *See generally Free Enter. Fund*, 130 S. Ct. 3138.

48 Proposed Opposition, *supra* note 3, at 24.

49 Letter from Hon. Lafe Solomon, Acting General Counsel, by Eric G Moskowitz, Assistant General Counsel, to Hon. Tom Horne, Attorney General of Arizona, at 2-3, 3 n.1 (Jan. 13, 2011) (citations omitted), *available at* https://www.nlrb.gov/sites/default/files/documents/234/letter_az.pdf.

50 The attributes of the states' residual sovereignty trump the exercise of congressional power. Alden v. Maine, 527 U.S. 706, 713 (1999) (Congress may not require the courts of a state to hear claims against the state arising under a federal statute otherwise within its enumerated powers because "the States' immunity from suit is a fundamental aspect of [their] sovereignty."). While this paper is not the appropriate place for an elaborate discussion of the residual sovereign immunities of the states from federal inference in their internal politics, it is worth noting that it was already well-established before the Revolutionary War that, as an attribute of its sovereignty, a nation-state is by right immune from another sovereign's attempt to interfere in its internal governmental affairs. "It clearly follows from the liberty and independence of Nations that each has the right to govern itself as it thinks proper and that no one of them has the least right to interfere in the government of another." Emmerich de Vattel, The Law of Nations or the Principles of Natural Law Applied to the Conduct and to the Affairs of Nations and of Sovereigns 131 (Charles G. Fenwick trans., Carnegie Inst. 1916) (1758) (Book II, Ch. iv, §54).

51 Deborah Jones Merritt, *The Guarantee Clause and State Autonomy: Federalism for a Third Century*, 88 Colum. L. Rev. 1, 41 (1988) ("A republican government . . . is responsible to its voters rather than to any outside agency."); *cf.* Taylor v. Beckham, 178 U.S. 548, 571, 580 (1900) (holding that a state's electoral process should be "free from external interference" and treating "interference of the general government," i.e., from the national government, as a type of "external interference").

52 No. 11-CV-506 (W.D. Okla., July 11, 2011), appeal pending, No. 11–6209 (10th Cir., Aug. 11, 2011).

53 *Id.* at 2.

54 *Id.* at 3.

55 *Id.* at 4.

56 *Id.* at 6.

57 *Id.*

58 *Id.* at 5 (quoting 29 U.S.C. § 160(a)).

59 *Id.* at 6-7.

60 *Id.* at 9.

61 *Id.* at 7-8.

Two Guiding Trends in Contemporary Labor and Employment Law: Technology and Fairness

*By Daniel Morton-Bentley**

There are two primary trends guiding contemporary labor and employment law. The first is the recognition and incorporation of technology into existing law. Labor law has led the way with the National Labor Relations Board (NLRB or Board)'s increased focus on social media firings. The second is increased fairness measures at the expense of legal certainty. Employment law has led the way here, with recent regulations interpreting the Genetic Information Non-Discrimination Act (GINA) as well as judicial expansion of Title VII to include discrimination based on sexual orientation.

I. Recognition and Incorporation of Technology

Perhaps the greatest challenge to unionized and non-unionized workplaces alike is how to best adapt to technological change. While employees have had access to the internet for a long time, employee use of social media is a relatively new phenomenon. The term "social media" encompasses a broad range of online communication programs. The two key social media programs are Facebook and Twitter. These programs allow employers and employees to instantly transform their thoughts into text that the whole world can read.

This development has its advantages and disadvantages. A large disadvantage for employers is that employees are more likely to air grievances online. Employees' candid comments, in turn, often lead to termination. Thus, the NLRB has recently had several opportunities to indicate how the National Labor Relations Act (NLRA) applies to online conduct. In short, online protected activity is treated largely the same as in-person protected activity. The Board has recognized that social media is a new, but no less legitimate, form of human communication.

Technology can also be used to expedite and aid in the enforcement of existing laws. The Department of Labor has found a unique way to do so. It recently introduced two mobile phone applications that will help workers prove the hours they have worked and the temperature on any given day. These applications are intended to aid the enforcement of wage and employment laws.

A. The NLRB

The NLRB has been extremely active as of late, proposing several new rules and rendering significant decisions. Some of this represents the push and pull of the political system: the current Democratic Board has reversed several policies initiated by the Republican Board. However, other changes reflect the Board's engagement with technological change. Specifically, the Board has issued several important decisions on online protected activity. These decisions: (1) confirm that the Board's protected activity inquiry is the same in person or online; and (2) indicate that employers must carefully draft their social media policies to avoid conflict with employees' right to organize.

Also, the Board has proposed new election rules that would allow for the electronic service of documents. This is a marked change from the approach the Board took to technology in its 2007 decision in *The Register Guard*[1] prohibiting employees from using employers' computers for union solicitation purposes.

B. Social Media and the NLRA

The Acting General Counsel of the NLRB released a report on August 18, 2011 summarizing the results of fourteen recent Board decisions involving employee use of social media.[2] The decisions primarily involved one of two legal issues: (1) whether online employee interaction constituted protected activity; and (2) whether employers' social networking policies infringed on employees' rights.

Under the NLRA, most employees in the private sector have a right to communicate with fellow employees about job-related concerns such as wages, hours, and workplace conditions. The theory behind this is that employee discussions about job conditions may germinate into unionization efforts.

1. Online Protected Activity

While the summarized decisions involved online and/or social media communications, the Board's protected activity inquiry proceeded as if the conversations took place in person. In order to be protected, employee activity must be concerted; that is, employees must act with or on the authority of other employees, and not solely by and on behalf of the employee him or herself. The summarized decisions held that online employee interaction constituted protected activity when employees expressed group concerns about wages, hours, and other work conditions. Therefore, comments, responses, and clicks of the "Like" button on Facebook all qualified as protected activity.

An illustrative case involved a sports bar's alleged mismanagement of its books. In early 2011, several employees of the bar discovered, much to their chagrin, that they owed a considerable amount of state taxes for 2010. The employees suspected it was due to employer error. The employer was informed of the employees' dissatisfaction, and one employee requested that the matter be discussed at an upcoming company meeting. Meanwhile, a former employee posted a comment on her Facebook Wall complaining about the issue. Another employee clicked "Like" underneath the comment. Then, another employee made a comment that she too owed state taxes and that one of the owners was "[s]uch an asshole." In response, the sports bar fired the employees. The employees then contacted the NLRB, which filed unfair labor practice charges against the bar.

In deciding this case, the Board found that the employees' activity was protected because it referred to group concerns

** Daniel Morton-Bentley is an Associate Attorney at Royal, LLP, a law firm in Northampton, Massachusetts that exclusively practices management-side labor and employment law.* *He sincerely thanks Karina Schrengohst, Esq. for her careful editing and drafting assistance.*

about a work-related issue (the employer's administration of income tax records). The employer's firing of the employees based on the online discussion was deemed unlawful despite the fact that the employees made disparaging remarks about the owner. It is long-standing Board policy that an employee does not lose the protection of the NLRA by resorting to swearing, name-calling, and/or sarcasm. In a similar case, an employee's reference to her supervisor as a "scumbag" in the context of a Facebook discussion of supervisory action was protected. Important to the Board was the fact that the Facebook postings occurred outside the workplace and thus did not disrupt the work of any employee or undermine supervisory authority.

However, online individual gripes in the absence of concerted activity are not protected under the NLRA. The Board found that the firing of a bartender after he posted a message on Facebook complaining about his employer's tipping policy was lawful. Although the employee's Facebook posting involved the terms and conditions of his employment, it was not protected because there was no discussion about the posting with his co-workers. In addition, there had been no employee meeting or attempt to initiate group action regarding the tipping policy. Similarly, the firing of a customer service employee after he posted a profane message on Facebook complaining about the "tyranny" of store management was found to be lawful. Although several of the employee's co-workers responded to his post, the Board found that the employee expressed an individual gripe about an individual dispute rather than an intention to initiate group action.

2. Employers' Social Media Policies

The more novel legal issue discussed in the cases is the permissible scope of employers' online and social media policies. The Board found that employers' social networking policies infringe on employees' rights when they are broad in scope. Employers must walk a fine line, as broad policies that could be interpreted as discouraging employees from discussing work conditions will likely be found illegal.

An instructive case involved an ambulance company's blogging and internet posting policy. The Board found that language prohibiting "employees from making disparaging remarks when discussing the company or supervisors" violated the NLRA because it impliedly encompassed their right to concerted activity under the NLRA. In addition, the Board found that a prohibition on "depicting the company in any media" violated the NLRA because it would prohibit an employee from engaging in protected activity such as posting "a picture of employees carrying a picket sign depicting the company's name."

Similarly, a hospital's social media, blogging, and social networking policy, which banned employee use of "any social media that may violate, compromise, or disregard the rights and reasonable expectations as to privacy and confidentiality of any person or entity," was struck down as overly broad. The Board found that the policy lacked limiting language and provided no definition or guidance of what was considered to be private or confidential. Other cases disapproved of similarly broad policies that lacked limiting language and specific examples of what was covered under the policy.

In contrast, the Board found that a provision in a supermarket chain's social media policy, which prevented "employees from pressuring their coworkers to connect or communicate with them via social media," was lawful. This part of the policy passed the Board's scrutiny because employees have the right to refrain from organizing under the NLRA. Thus, the policy "was narrowly drawn to restricted harassing conduct and could not reasonably be construed to interfere with protected activity."

Two important things can be taken away from these case summaries. First, the NLRB does not treat online employee interaction differently from personal interaction. Employees are allowed to discuss work-related grievances whether their communications are in person or via social media, and employers cannot punish them for doing so despite how disparaging their comments may be. Second, overly-broad online or social media policies are likely to be struck down.

C. Department of Labor Mobile Phone Apps

The Department of Labor has launched two applications ("apps") for mobile phones that will assist workers in proving wage discrimination and hazardous working conditions. The first is a "timesheet" app that allows workers to "independently track the hours they work and determine the wages they are owed."[3] It is currently available in both English and Spanish for iPhone and iPod Touch. This application is significant because worker-generated records will stand as definitive proof of an employee's hours if his or her employer does not have adequate records.

Also available for download is a Heat Index application that records outdoor temperature, calculates a heat index, and recommends "protective measures that should be taken . . . to protect workers from heat-related illness."[4] While this application will be relevant to a more specialized audience (outdoor workers), its import is broad. For example, suppose an employer refused to grant breaks or provide water to employees working in extremely hot weather. The application's record of the outside temperature—along with its unheeded recommendations—could support a civil or criminal suit against the employer.

While the Department of Labor's initial foray into the mobile application market may be modest, it illustrates a movement towards the use of technology as a tool to enforce existing laws.

II. Fairness Measures in Employment Law

Recent developments in employment law have emphasized fairness at the expense of legal certainty. A perennial dilemma for policy makers is whether they should adopt hard-line rules or a flexible balancing test. Hard-line rules have the advantage of legal certainty, while flexible balancing tests are better suited towards individuals' unique situations. Employment law as of late has favored fairness measures that entail individualized consideration. For example, the Equal Employment Opportunity Commission's (EEOC) regulations interpreting GINA confirm that it is difficult for an employer to easily prove the lawful acquisition of genetic information. Additionally, courts have expanded Title VII's prohibition on

gender discrimination to encompass discrimination based on sexual orientation as well as "sex plus discrimination."

A. GINA Regulations

GINA's application is quite broad: it prohibits employers from "requesting, requiring, or purchasing genetic information" as well as making any employment decisions based on an individual or his or her family member's genetic information. "Family member" includes "a person who is a dependent . . . as the result of marriage, birth, adoption, or placement for adoption" as well as relatives of the first, second, third, and fourth degree.[5] "Genetic information" is defined as an individual or his or her family member's genetic tests, "the manifestation of disease or disorder" in family members of the individual, and any request for, or participation in, genetic testing.[6] There are six statutory exceptions where an employer may legally acquire genetic information. Most relevant are exceptions for "inadvertent acquisition" and for acquisition in the course of processing an employee's Family Medical Leave Act request.[7]

GINA's prohibition on genetic-based employment discrimination is grounded in fairness: because employees cannot control their genetic information, treating employees differently because of their genetic information is unfair. Like many other employment laws, GINA attempts to force employers to treat employees equitably.

However, it will not always be clear when an employer violated GINA due to the subtle distinction between permissible and non-permissible acquisition of genetic information. For example, the regulations indicate that overhearing a conversation about genetic information does not violate GINA unless the employer "actively listen[s]." Similarly, "a casual question between colleagues . . . concerning the general well-being of a parent or child would not violate GINA," but a "follow [ ] up . . . question concerning a family member's general health with questions that are probing in nature" would. This situation could be especially tricky as co-workers engaged in conversation about medical issues are not likely to be considering the niceties of GINA.

An employer who has sufficient proof that its manager or managers acted with good intentions will eventually be able to refute a GINA claim. However, prior to that point, GINA's ambiguities will allow employees to bring claims and require employers to defend against these claims. Judicial disposition of GINA claims thus far shows that employees have had little luck stating, let alone proving, GINA claims.[8] Nevertheless, GINA's breadth and subtle distinctions will surely engender uncertainty and give rise to more claims in the future.

B. Title VII

Several recent decisions have expanded the literal language of Title VII to include discrimination that does not fit the familiar gender discrimination paradigm.[9] Two examples are: (1) discrimination based on sexual stereotypes; and (2) discrimination based on a particular subclass of men or women ("sex plus" discrimination).

1. Discrimination Based on Sexual Stereotypes

Courts have stretched the literal language of Title VII's prohibition on gender discrimination for some time. The most notable example is the U.S. Supreme Court's holding in *Price Waterhouse v. Hopkins* that sexual stereotypes could give rise to Title VII gender discrimination.[10] The plaintiff in *Hopkins* was a well-qualified manager who was repeatedly put down for her failure to adhere to feminine stereotypes. She was once referred to her by a colleague as "macho."[11] A second co-worker claimed she "overcompensated for being a woman," and a third recommended that she take "a course at charm school."[12] The Supreme Court held that this discrimination was actionable under Title VII.

The Third Circuit Court of Appeals considered a more tenuous claim of gender discrimination in the recent case of *Prowel v. Wise Business Forms, Inc.*[13] The plaintiff in *Prowel*, a gay male, was subjected to cruel and pervasive harassment. He was called "Princess" and referred to as a "fag."[14] Co-workers wrote graffiti in the men's bathroom "claiming Prowel had AIDS and engaged in sexual relations with male co-workers."[15] An unidentified co-worker left "a pink, light-up, feather tiara with a package of lubricant jelly" at his work station.[16]

As a result of this harassment, Prowel brought a Title VII action claiming that he was discriminated against because he did not fit his co-workers' definitions of a stereotypical male. The Third Circuit first admitted that the line between gender discrimination and sexual orientation discrimination "can be difficult to draw."[17] Nevertheless, it found sufficient evidence that Prowel was discriminated against based on gender stereotypes. The evidence established that Prowel, among other things "did not curse . . . was very well-groomed . . . [and] discussed things like art, music, interior design, and décor."[18] This distinguished Prowel from his male factory colleagues who liked to "hunt [ ] . . . fish [ ] . . . dr[i]nk beer .. .[and watch] football [and other] sports."[19]

Some would argue that the Third Circuit's holding impermissibly stretched the language of Title VII. Title VII prohibits "discriminat[ion] against any individual . . . because of such individual's . . . sex."[20] It is more accurate to say that Mr. Prowel was discriminated because of his sexual orientation and not because of his sex. This is an important distinction, as Congress has considered, and repeatedly rejected, proposals to add sexual orientation as a protected category under Title VII.[21] *Prowel* illustrates that many courts are not willing to wait for Congress. Therefore, employers must take immediate action to prevent workplace discrimination based on sexual orientation.[22]

2. "Sex Plus" Discrimination

Many courts have held that "sex plus" discrimination is illegal under Title VII. "Sex plus" discrimination refers to discrimination based on an employee's sex plus an additional characteristic. Thus, the alleged victims are a subclass within the larger categories of male and female.

The First Circuit recently considered a sex plus discrimination claim in *Chadwick v. WellPoint, Inc.*[23] The plaintiff was a well-respected employee of the defendant. In 2006, the plaintiff applied for a promotion at her supervisor's urging. The company named two finalists for this position: the plaintiff and another woman. The plaintiff was better-qualified, having received a superior performance review and possessing

greater work experience. Nevertheless, the plaintiff did not receive the promotion.[24]

The plaintiff's interviewers made several comments that suggested that the plaintiff's status as a mother cost her the promotion. For example, one interviewer sent the plaintiff an e-mail two months before the decision that said: "Oh my - I did not know you had triplets. Bless you!"[25] Also, in response to a hypothetical question about disciplining an associate, an interviewer asked the plaintiff: "[Y]ou are a mother[.] [W]ould you let your kids off the hook that easy if they made a mess in [their] room[?] [W]ould you clean it or hold them accountable?"[26] Finally, after being denied the position, an interviewer told the plaintiff: "It was nothing you did or didn't do. It was just that you're going to school, you have the kids and you just have a lot on your plate right now."[27]

The First Circuit held that the plaintiff alleged sufficient facts to survive a motion for summary judgment because "an employer is not free to assume that a woman, because she is a woman, will necessarily be a poor worker because of family responsibilities."[28] The district court granted summary judgment in favor of the defendant because "nothing in [the interviewer's] words showed that the decision was based on a stereotype about female caregivers, not about caregivers generally."[29] The First Circuit disagreed, explaining that plaintiffs may prove their case through circumstantial evidence, and there was sufficient evidence to show that the employer acted based on stereotypical notions about working women with children.[30]

In the strictest sense, *Chadwick* was a case about working-parent discrimination. However, the First Circuit peered beneath the surface and surmised that the case was really about discrimination based on sexual (female) stereotypes. *Chadwick*, like *Prowel*, illustrates that many courts take an expansive view of Title VII's protections.

III. Conclusion

In the labor and employment law universe, there are two things that we can be certain of. The first is that new technology will continue to impact the way that people work and interact. The second is that new technological and societal developments will necessitate laws (or interpretations of laws) designed to promote fairness. This article has outlined how agencies, courts, and Congress have reacted to recent technological developments and calls for fairness. These trends will surely continue into the future, and forward-looking employers can be ready for them.

Endnotes

1 The Guard Publ'g Co., 351 N.L.R.B. No. 70 (Dec. 16, 2007).

2 These decisions were rendered by lower NLRB tribunals, and not the official National Labor Relations Board. For simplicity's sake, the decision-makers are referred to as "the Board."

3 Press Release, U.S. Department of Labor, Wage and Hour Division, Keeping Track of Wages: The US Labor Department Has an App for That! (May 9, 2011), *available at* http://www.dol.gov/whd/media/press/whdpressVB3.asp?pressdoc=national/20110509_1.xml.

4 Android Market, OSHA Heat Safety Tool, https://market.android.com/details?id=com.erg.heatindex (last visited Oct. 23, 2011).

5 *See* 29 C.F.R. §1635.3 (2011). Fourth-degree relatives include one's "great-great-grandparents, great-great-grandchildren, and first cousins once-removed." 29 C.F.R. §1635.3(iv).

6 *See* 29 C.F.R. §1635.3(c).

7 § 1635.8(b).

8 *See, e.g.*, Dumas v. Hurley Medical Center, 2011 WL 3112882 (E.D. Mich. 2011) (The plaintiff failed to state a claim); Dodrill v. Alpharma, Inc., 2011 WL 3877076 (N.D.W. Va. 2011) (same); Bullock v. Spherion, 2011 WL 1869933 (W.D.N.C. 2011) (same). *But see* Tate v. Quad/Graphics Inc., 2011 WL 4352301, at *2 (E.D. Ark. 2011) ("[The plaintiff] alleges that defendants have requested medical information with respect to him or his family members . . . . It is unclear whether [the plaintiff] has a GINA claim . . . . [Therefore,] he is ordered to file a second amended complaint correcting the deficiencies in his GINA claim within 30 days of this order.").

9 For an excellent discussion of the topic, see Susan Fahey Desmond, *Employment Law: Why It Frequently Changes and What Has Happened Recently*, *in* Complying with Employment Regulations, Leading Lawyers on Analyzing Legislation and Adapting to the Changing State of Employment Law (Inside the Minds ed., 2011).

10 490 U.S. 228 (1989).

11 *Id.* at 235.

12 *Id.*

13 579 F.3d 285 (3d Cir. 2009).

14 *Id.* at 287.

15 *Id.*

16 *Id.*

17 *Id.* at 291.

18 *Id.*

19 *Id.* at 287.

20 42 U.S.C. 2000 e-2(a).

21 Desmond, *supra* note 9, at *8.

22 That is, if state law does not already prohibit it. *See, e.g.,* Mass. Gen. Law. c. 151b ("It shall be an unlawful practice . . . [f]or an employer . . . because of. . . sexual orientation . . . to refuse to hire or employ or to bar or to discharge from employment [an] individual or to discriminate against such individual . . . .").

23 561 F.3d 38 (1st Cir. 2009).

24 *Id.* at 41.

25 *Id.* at 42.

26 *Id.*

27 *Id.*

28 *Id.* at 45.

29 *Id.* at 46 (quoting Chadwick v. WellPoint, Inc., 550 F. Supp. 2d 140 (D. Me. 2008)).

30 *Id.* at 46-47.

Litigation

Business Cases and the Roberts Supreme Court

*By Martin J. Newhouse**

The statement that the Supreme Court under Chief Justice Roberts, and more specifically the Court majority of five Republican-appointed Justices, has been unusually favorable, even biased, toward business interests is a familiar one in the media and much-repeated among liberal legal commentators (including, with respect to the 2010 *Citizens United* decision, the President of the United States).[1] But is this true? Have the Roberts Court's rulings in cases affecting business interests actually been especially favorable to those interests? This article seeks to answer this question.

Not surprisingly, the issue of pro-business bias is complicated. To begin with, it is clear beyond dispute that none of the Justices generally identified as conservative—specifically, Chief Justice Roberts and Associate Justices Alito, Kennedy, Scalia, and Thomas—is reflexively pro-business.[2] In numerous cases these Justices have cast their votes for, and even written the majority opinions in, decisions in which business parties have lost and investors, consumers, or employees have won.

Most recently, for example, Chief Justice Roberts, writing for a unanimous Court in *Erica P. John Fund, Inc. v. Halliburton Co.*,[3] issued a decision that makes it easier for plaintiffs to certify class actions in securities fraud cases, by holding that they are not required to prove loss causation at the certification stage. Justice Scalia similarly delivered the decision for a unanimous Court in January 2011 in *Thompson v. North American Stainless, LP*[4] holding that the plaintiff Thompson could maintain his claim for retaliation under Title VII even though he had not himself engaged in protected activity, because he alleged that he had been terminated in retaliation for the fact that his fiancée had filed a charge of sex discrimination against their common employer. Yet another recent unanimous decision by the Supreme Court that arguably was anti-business was *Matrixx Intiatives, Inc. v. Siracusano*, in which the five Republican-appointed Justices joined a majority opinion by Justice Sotomayor holding that plaintiffs could bring a securities fraud case based on "a pharmaceutical company's failure to disclose reports of adverse events associated with a product [where] the reports do not disclose a statistically significant number of adverse events."[5]

And lest one think that the allegedly pro-business Justices only join in decisions against business parties that are unanimous, therefore arguably only in cases in which the result is so obvious that even a judge with pro-business leanings could not hold for the business party in the case,[6] there have also been business-related decisions issued by the Roberts Court in which the five Republican-appointed Justices have split their votes, with some joining the majority ruling against corporate interests. A recent example is *Kasten v. Saint-Gobain Performance Plastics Corp.*,[7] which was another case dealing with an anti-retaliation provision, this time a provision of the Fair Labor Standards Act of 1938 ("Act").[8] Section 215(a)(3) of the Act makes it illegal for an employer, *inter alia,* "to discharge or in any other manner discriminate against any employee because such employee has filed any complaint . . . under or related to the Act. . . ." Here the employee had complained to his employer orally about certain work conditions. The issue decided in the case was whether "filed any complaint" in the provision included those oral complaints. The Supreme Court found that it did, in a majority decision by Justice Breyer, in which the Chief Justice and Justices Kennedy, Ginsburg, Alito, and Sotomayor joined. Justice Scalia, joined by Justice Thomas,[9] dissented, arguing in essence that even if "filed any complaint" included oral complaints the employer should still have prevailed because, in his view, § 215(a)(c) does not cover complaints to the employer at all, but only complaints made to a government agency. Interestingly, the employer had raised this issue below, but never mentioned it in its petition for certiorari and the majority, including three of the Justices often considered pro-business, deemed it to have been waived.[10] Two of the more conservative Justices, Kennedy and Thomas (concurring in the judgment), joined their more liberal colleagues in deciding in *Wyeth v. Levine* that federal law did not preempt a state law failure-to-warn claim with respect to Wyeth's anti-nausea drug Phenergan.[11]

Just as in cases such as *Kasten* and *Wyeth* allegedly pro-business Justices have ruled against the business party, so in other cases some of the so-called liberal Justices have joined with some of their conservative colleagues to support a result favoring a business party. Look, for example, at the constellation of Justices in *Watters v. Wachovia*.[12] The question before the Court was whether a wholly-owned mortgage lending subsidiary of a national bank could be regulated by state banking authorities. The answer depended on the enforceability of an Office of the Comptroller of the Currency regulation preempting state regulation of national bank subsidiaries. Preemption would have been the pro-business position, since parallel regulation by federal and state authorities would likely result in inefficiency, waste, and higher costs for the businesses involved (and ultimately, perhaps, for consumers). Justice Ginsburg's majority opinion rejecting the state's claim of parallel regulatory authority was joined by both conservatives and liberals: Justices Kennedy, Souter, Breyer, and Alito. Justice Stevens argued in dissent that preemption should be based on an explicit federal statute, not a mere OCC regulation, and that the majority's decision imperiled the delicate balance between federal and state authority in the banking field. Joining Justice Stevens in defense of federalism were two "conservative" Justices, Chief Justice Roberts and Justice Scalia. Could there be a clearer demonstration than this case that liberalism and conservatism do not automatically align with or against business's perceived interests?[13]

** President, New England Legal Foundation*

Nevertheless, while the business decisions by the Roberts Court, when taken as a whole, demonstrate that there is no automatic or reflexive alignment by the Court's more conservative Justices with business interests, it is also true that in a number of recent high-visibility business cases, the Justices have divided along what might be termed political lines (Republican-appointed Justices on one side, Democrat-appointed Justices on the other) with regularity, resulting in a number of 5-4 decisions in favor of business parties. Does this 5-4 split support the claim of a pro-business bias, at least in these cases? Perhaps it would if all one looked at were the results and the identity of the prevailing parties—as the media and commentators often seem to do (see, for example, the editorial in *The New York Times* greeting the Supreme Court's recent decision in *Wal-Mart Stores, Inc. v. Dukes*,[14] which was entitled "Wal-Mart Wins. Workers Lose."[15]). However, closer study of the decisions reveals that what is at issue in each case is not a simple matter of slant or bias (either pro- or anti-business), but rather a struggle with close questions and cutting-edge legal issues that, in all fairness, were evidently decided by each of the Justices based on their honest views of the law. A brief review of four recent allegedly pro-business decisions—three quite high-profile, one less so—will demonstrate this to be the case.

For this analysis there is perhaps no better place to begin than with the *Citizens United* decision, which appears for many liberal commentators to epitomize the alleged pro-business bias of the five conservative Justices.[16] At issue in the case was the constitutionality of the McCain-Feingold campaign finance reform that made it illegal for all corporations (including nonprofits) and labor unions to use money from their general funds for advocacy for or against the election of a candidate in a federal election within thirty days before a primary election and sixty days before a general election.[17] Often seemingly overlooked by liberal commentators is that the Supreme Court's decision that this provision was unconstitutional benefited not only business corporations, but also nonprofits (indeed, the case was brought by a non-profit), and unions. Indeed, one would seek in vain to find any recognition, in mainstream media and commentary at least, that this decision by the Court's Republican-appointed majority was, among other things, a First Amendment victory for nonprofit liberal advocacy groups and organized labor.[18]

Also overlooked is the extent to which both the majority and dissent in the case agreed about fundamentals. For example, many, if not most, critics of *Citizens United* seem unaware of the Supreme Court's prior decisions holding that the First Amendment applies to corporations; *Citizens United* broke no new ground in this respect (although it is routinely criticized for having done so).[19] Even Justice Stevens, in his eloquent dissent, agreed that corporations enjoy First Amendment protection.[20] Not only did both the majority and the dissent agree that corporate speech is protected by the First Amendment, they both also agreed that such protection is not absolute.[21] Additionally, with only Justice Thomas disagreeing, both the majority and the dissent agreed that the statute's disclosure requirements did not violate the First Amendment.[22]

Finally, it is clear from their opinions that all of the Justices, both those in the majority and in the dissent, agreed that the First Amendment serves a crucially important role in our democracy, namely to insure that the people have access to all the information they need in order to exercise their sovereignty as informed citizens. It was their answers to the question whether the limitations on speech at issue served this goal that divided the Justices. Such a question is always very difficult and calls for the most careful consideration and balancing.

In *Citizens United,* the dissenters clearly believed that the restrictions were justified based on a historical record that, for them, demonstrated the tendency of the for-profit corporate form to corrupt political debate. In contrast, the majority did not see this tendency as a proven fact and, perhaps, also did not think that even historical instances of corruption warranted a blanket limitation on all corporate speech, which can, in its own right, be informative.

Unfolding events will no doubt demonstrate whether the fears of the dissent or the hopes of the majority are justified. But for our purposes, it is sufficient to note that the majority's rejection of a limitation on speech that swept broadly enough to include within it not only large for-profit corporations (whose potentially malign influence on federal elections is the primary focus of critics of the decision[23]), but corporations of all sizes and descriptions, including nonprofit corporations and labor unions, can hardly be considered as simply a pro-business decision.

Turning from *Citizens United,* which was decided in 2010, to a more recent example: One of the most publicized and complained-about allegedly pro-business decisions was the Supreme Court's ruling in June 2011 in *Wal-Mart Stores, Inc. v. Dukes*,[24] in which the Court reversed the certification of a class of 1.5 million current and former female employees of the retailer in a sex discrimination suit.

For critics of the holding, the Court's decision appears to be yet again an automatic 5-4 ruling in favor of business, with the Republican-appointed Justices joining in a majority opinion written by Justice Scalia. This view ignores, however, the fact that the Court's reversal of the class certification was in one important aspect unanimous: all of the Justices, both conservative and liberal, agreed that class certification in the case had been sought and granted under the wrong provision of Rule 23 of the Federal Rules of Civil Procedure.[25] Specifically, the Court unanimously held that the plaintiffs' attempt to bring their class action under Fed. R. Civ. P. 23(b)(2) was improper because the plaintiffs were seeking individualized monetary relief, such as back pay. As Justice Scalia's opinion pointed out, Rule 23(b)(2) by its terms applies "only when a single injunction or declaratory judgment would provide relief to each member of the class," and "does not authorize class certification when each class member would be entitled to an individualized award of monetary damages."[26] Since the plaintiffs had relied solely on Rule 23(b)(2) for their claims, the certification had to be reversed—a conclusion with which the Justices dissenting as to other aspects of the decision expressly agreed.[27]

Clearly, at most, this unanimous ruling by the Court was a setback for the plaintiffs, since at the least it would require a new attempt to certify a class under a different part of Rule 23 (specifically Rule 23(b)(3)). Yet, most of the commentary on

the decision has paid little attention to the fact that the liberal Justices joined in this defeat for the Wal-Mart workers who had brought the suit.[28]

To be fair, the lack of focus on the unanimous part of the decision in *Wal-Mart* is no doubt due to the less technical aspect of the case over which the liberal and conservative Justices did differ, i.e., whether the evidence presented by the plaintiffs was sufficient to demonstrate the commonality needed for class certification under any provision of Rule 23. The liberal Justices plainly thought that the plaintiffs' evidence was sufficient for certification; the conservatives, in an opinion authored by Justice Scalia, clearly thought that it was not.[29] This is not the place to delve into the specifics of the *Wal-Mart* case, which would require an entirely separate article, but the most basic aspects of the litigation—including that the putative class composed of around 1.5 million women concerning employment decisions made by managers in each of Wal-Mart's approximately 3400 stores throughout the country, who were given discretion in employment matters by Wal-Mart (which had and has an official policy against sex discrimination); that some of the 1.5 million putative class members were themselves managers who arguably might have made some of the decisions complained of by other plaintiffs; that the expert testimony introduced by the plaintiffs included a sociological expert who "could not calculate whether 0.5 percent or 95 percent of the employment decisions at Wal-Mart might be determined by stereotyped thinking"—indicate the questionable, or from a different point of view the innovative (and, therefore, inherently risky), nature of the argument for class certification in the case.[30] It is precisely because the basis for certification was questionable at best, that *The New York Times*'s longtime Supreme Court commentator Linda Greenhouse, in her review of the Court's recent decisions, described the *Wal-Mart* ruling as the Court's "[l]east surprising decision" of the 2010 Term.[31]

Importantly for our purposes, while the decision on commonality that generated a 5-4 split among the Justices has also generated the most heat in the commentary about *Wal-Mart v. Dukes*, it is the more technical, unanimous ruling on the requirements of Rule 23(b) that may have a more lasting impact going forward. This is because, although all those seeking class certification must fulfill Rule 23's commonality requirement, the majority's ruling on that issue was based on the fairly unique facts of Wal-Mart's operations, while the ruling on the scope of Rule 23(b)(2) will govern future class actions in all factual contexts. And, given that this unanimous decision on the Court's part will limit the ability of all potential plaintiffs to seek certification under Rule 23(b)(2), including business plaintiffs (who might seek class certification in a variety of commercial contexts), it is hard to see how it could be characterized as the product of a pro-business bias either by the conservative majority or by the Court as a whole.

Yet another high-visibility recent business decision that has been criticized as one more indication of the conservative Justices' alleged anti-consumer and reflexively pro-business bias is *AT&T Mobility LLC v. Concepcion*,[32] which dealt with the much-litigated question of the enforceability of class-arbitration waivers, this time in a consumer mobile phone service contract. Again, as in *Wal-Mart,* the five conservative Justices ruled in favor of the business party, with a majority opinion by Justice Scalia overruling the Ninth Circuit's application of a California rule under which the waiver was automatically unconscionable. The four liberal Justices joined in a dissenting opinion by Justice Breyer, arguing that the per se state rule at issue did not, in fact, violate the Federal Arbitration Act ("FAA").[33] Despite the familiar 5-4 division of the Justices, and as with the other decisions under discussion, a closer look at the Court's decision erodes any simplistic view that would label it as decisively pro-business or anti-consumer.

The specific legal question before the Court in *AT&T Mobility* was whether the California federal courts' application of a state court rule that operated, in effect, to invalidate such waivers in consumer contracts as per se unconscionable violated the FAA.[34] By its terms, the FAA requires courts to treat arbitration agreements on an equal footing with all other contracts and bars courts from singling out arbitration agreements for suspect status.[35] In other words, the question for the Court was whether the FAA requires that courts, when confronted with a challenge to a class arbitration waiver provision on the grounds that it is unconscionable, treat the question as they would any other contract, i.e., as a fact-intensive inquiry into the nature and circumstances of the waiver and the arbitration provisions themselves.[36]

Thus, the basic issue in *AT&T Mobility* was purely legal, and both the majority opinion by Justice Scalia and the dissent by Justice Breyer agreed on what that legal issue was, coming as noted above to diametrically-opposed answers as to whether California's per se rule singled out arbitration agreements for the type of special treatment forbidden by the FAA, each side bolstering its arguments with discussions about the suitability of an arbitral forum for class action proceedings.[37] One would be hard-pressed to see in the Justice's opinions any sign that the identity of the parties involved—i.e., that it was a consumer case against a business—had anything at all to do with their analysis. Indeed, as recent major holdings in the area of arbitration have shown, the Supreme Court has routinely been indifferent to whether the parties involved were commercial enterprises or individuals.[38] Moreover, the notion that *AT&T Mobility* was an anti-consumer decision is belied by the fact that it did not remove the plaintiffs' ability to challenge the class arbitration waiver at issue; it only removed their ability to rely on a per se rule. The plaintiffs remained free to allege and prove unconscionability as it has traditionally been proven, through a close examination of the contract at issue and the circumstances surrounding its execution.[39]

Finally a brief analysis of one more, somewhat lower-profile business-related decision from the Supreme Court's 2010 Term will hopefully underscore the lack of foundation for claims that the Court's decisions in business cases are the product of bias, rather than the results of honest decision-making about issues that are, by their very nature, matters of first impression and not easy cases. This is another 5-4 decision, *Pliva v. Mensing,* in which the conservative majority held for the business defendant in an opinion written by Justice Thomas. Justice Sotomayor wrote a stinging dissent in favor of the consumer plaintiffs, in which all of her liberal colleagues joined.[40] The issue in the case was essentially the same as had been presented in *Wyeth*

v. Levine in 2009—i.e., whether federal law preempted a state law failure-to-warn claim brought against a drug manufacturer based on allegedly defective labeling—with this important, and, it turned out, decisive difference: in *Wyeth* the defendant was the brand-name manufacturer of the drug at issue; in *Pliva* the defendant was the generic manufacturer.

For the majority this distinction was crucial because of the different federal drug labeling duties of generic and brand-name manufacturers. Specifically, while, under federal law and regulation, the brand-name manufacturer has the ability to enhance its labeling without seeking FDA approval first, a generic manufacturer is not free to change its label on its own (federal law requires generic manufacturers to use the labeling approved for brand-name manufacturers). If a generic manufacturer wishes to strengthen warnings on its label, it can propose the change to the FDA, which, if it approves, will then work "with the brand-name manufacturer to create a new label for both the brand-name and generic drug."[41] Noting that federal preemption will be found where it is impossible for a private party to comply with both state and federal requirements, the majority ruled that impossibility was present in this case because (a) the generic manufacturer could not have changed its label to comply with state law without violating its federal duty to keep the label the same as the brand-name label, and (b) even if the generic manufacturer requested a change from the FDA (which the agency might refuse), it still would not have satisfied state law which demanded a safer label, not that the manufacturer ask the FDA for a labeling change.[42] As Justice Thomas's opinion puts it: "The question for 'impossibility' is whether the private party could independently do under federal law what state law requires of it."[43] On this basis, Justice Thomas, who had agreed in *Wyeth v. Levine* that the state-law claim against the brand-name manufacturer was not preempted, found preemption here, even as he recognized, and regretted, that this result would leave millions of generic drug users without a state law remedy.[44]

Writing for the dissent in *Pliva*, Justice Sotomayor, while she strongly disagreed with the majority's ultimate decision, did not dispute the existence of impossibility as a ground for finding federal preemption. She (and her liberal colleagues) only disputed that impossibility, and therefore preemption, had been established in this case, because there was no evidence the generic manufacturers had ever approached the FDA about a label change to bring them into compliance with state law. Rejecting the majority's view that impossibility exists if the private party cannot take action independently, Justice Sotomayor nevertheless described several scenarios that she believed would satisfy the impossibility standard:

> This is not to say that generic manufacturers could never show impossibility. If a generic-manufacturer defendant proposed a label change to the FDA but the FDA rejected the proposal, it would be impossible for that defendant to comply with a state-law duty to warn. Likewise, impossibility would be established if the FDA had not yet responded to a generic manufacturer's request for a label change at the time a plaintiff's injuries arose. A generic manufacturer might also show that the FDA had itself considered whether to request enhanced warnings in light of the evidence on which the plaintiff's claim rests but had decided to leave the warnings as is. . . . But these are questions of fact to be established through discovery.[45]

As this passage reveals, Justice Sotomayor's opinion, while expressing concern at the outset about the impact of the Court's decision on the "75 percent of all prescription drugs dispersed in this country," which are generics, and while, in other places, dismissing the majority's definition of impossibility as illogical, cannot be said to be reflexively anti-business or reflexively pro-consumer in her approach, since she does not foreclose preemption as a defense for generic drug manufacturers.

In truth, both the majority's and the dissent's positions in *Pliva* are reasonable interpretations of the law of preemption in light of the novel question presented to the Court, namely "whether conflict pre-emption should take into account [the] possible actions by the FDA and the brand-name manufacturer.[46] For the majority, the fact that the generic manufacturers were legally powerless to take remedial action on their own, but rather were dependent on a string of possibilities, was enough to establish impossibility in this instance.[47] On the other hand, the dissent's position that, at the very least, to establish impossibility the generic manufacturers should have been required to show that they had at least requested a change to bring their labels into compliance with state law, does not appear terribly unreasonable.[48] The point here is that neither position can fairly be described as pro-business or anti-business, much less were the Justices' positions plausibly based on any such biases. Rather, their disagreements were over legal doctrine, regulatory impact, and even the meaning of the Constitution's Supremacy Clause.[49] In short, once again, as in *AT&T Mobility,* the identity of the parties had very little to do with the outcome of the case.

To sum up, the above analysis of some of the major business decisions by the Roberts Court indicates that claims of an automatic or even a general pro-business bias are not well-founded, either with respect to the five more conservative Justices or with respect to the Court as a whole. That the Roberts Court has granted certiorari in more business cases than its predecessors is often pointed out, but as the cases above indicate, this may well be the result of a recognition that there are important and outstanding issues in this area that need to be resolved. For those who represent business interests, the Supreme Court's more hospitable attitude toward business cases is welcome. However, as the above analysis demonstrates, business parties should expect in the Supreme Court as elsewhere that, if they are to prevail, they must rely on the strength and cogency of their arguments and not the makeup of the bench.[50]

Endnotes

1 Citizens United v. Fed. Election Comm'n, 130 S. Ct. 876 (2010). For a recent, somewhat balanced media discussion of the issue of alleged business bias in the Supreme Court's business decisions see Richard McGregor, *Walmart Discrimination Win Highlights Claims of Court Bias*, Fin. Times, June 21, 2011, *available at* http://www.ft.com/intl/cms/s/0/40f8f9b4-9c3f-11e0-acbc-

00144feabdc0.html#axzz1UZF6FCx4. As is well-known, President Obama's criticism of the Supreme Court's decisions in *Citizens United* was made before the Congress and the Nation during his State of the Union address on January 27, 2010.

2 Although such categories are objectionable because of their over-simplification and imprecision, for ease of reference in what follows the term "conservative Justices" will be used to identify collectively the five Republican-appointed Justices (Chief Justice Roberts and Associate Justices Alito, Kennedy, Scalia, and Thomas). The term "liberal Justices" will similarly be used to refer to the four Democrat-appointed Justices (Associate Justices Breyer, Ginsburg, Kagan, and Sotomayor).

3 131 S. Ct. 2179 (2011).

4 131 S. Ct. 863 (2011).

5 131 S. Ct. 1309 (2011).

6 In fact, that the Supreme Court's decision is unanimous does not mean that the result was obvious from the start. For example, with respect to the issue presented in *Thompson v. North American Stainless, LP*, 131 S. Ct. 2179, as legal commentator Edward Whelan has pointed out, those courts of appeal that had earlier dealt with the question of such third-party retaliation claims (including the decision under review) had decided in the employer's favor that such claims could not be brought under Title VII. *See* Those Sneaky Corporatist Justices, Posting of Ed Whelan to Bench Memos, http://www.nationalreview.com/bench-memos/257788/those-sneaky-corporatist-justices-ed-whelan (Jan. 24, 2011, 13:36 EST).

7 131 S. Ct. 1325 (2011).

8 52 Stat. 1060, 29 U.S.C. §201 *et seq.*

9 Justice Kagan did not participate in the consideration or decision of the case.

10 *Kasten*, 131 S. Ct. at 1336. In his dissent, Justice Scalia takes strong issue with the Court's refusal to consider the question whether intra-company complaints are even covered by the anti-retaliation provision.

11 555 U.S. 555 (2009). The *Wyeth* decision, while benefiting consumers/patients, illustrates how ambiguous such benefits may be, and raises the question whether the case is ultimately actually pro-patient. For example, anyone concerned about the availability of much-needed drugs should be disturbed that, as the dissent in Wyeth correctly points out, the majority's decision makes state court juries, rather than the FDA, ultimately responsible for regulating warning labels for prescription drugs (at least with respect to brand-name manufacturers, see discussion below of the Court's 2011 holding in *Pliva v. Messing*, which came to a different conclusion with respect to generic manufacturers). It is an understatement to say that lay juries, with their exclusively retrospective and case-specific vision and authority, are ill-equipped to perform the important function of regulating warning labels. Furthermore, it seems obvious that, for the vast majority of consumers, the Court's decision will mean higher costs and less availability as pharmaceutical companies deal with the uncertainty that has been created and the litigation that will result. How is a drug company ever to know whether its drug labeling is sufficient? Even after one jury answers that question, another can decide it differently. And the wider economic impacts of this decision, as other regulated industries determine the extent to which the Supreme Court majority's reasoning may apply to them, remain to be seen. Such considerations were evidently not on the majority Justices' minds in *Wyeth* (nor were they decisive for the majority in the 2011 decision *Pliva v. Mensing*, discussed in note 33 and the text below).

12 550 U.S. 1 (2007).

13 Justice Thomas took no part in the consideration or decision of the case. More recently, in another example of crossing the liberal/conservative line, Justice Sotomayor joined her five more conservative colleagues for a 6-3 decision in *Sorrell v. IMS Health Inc.*, 131 S. Ct. 2653 (2011), which struck down a Vermont statute restricting the sale of pharmacy records showing individual physicians' prescribing practices as in violation of the First Amendment.

14 131 S. Ct. 2541 (2011).

15 Editorial, *Wal-Mart Wins. Workers Lose.*, N.Y. Times, June 20, 2011.

16 Citizens United v. Fed. Election Comm'n, 130 S. Ct. 876 (2010). While *Citizens United* was a First Amendment case, as opposed to the commercial cases generally discussed in this article, it and the Court's subsequent First Amendment decision in *Sorrell v. IMS Health*, 131 S. Ct. 2653 (2011), (see note 13 above) are discussed because they both dealt with issues of great importance to business interests.

17 2 U.S.C. § 441b(b)(2), § 434(f)(3)(A).

18 That such groups, and not simply large for-profit corporations, were on the Justices' minds is revealed by Justice Kennedy's description of the sweep of the provision in his majority decision for the Court:

> Thus, the following acts would all be felonies under §441b: The Sierra Club runs an ad, within the crucial phase of 60 days before the general election, that exhorts the public to disapprove of a Congressman who favors logging in national forests; the National Rifle Association publishes a book urging the public to vote for the challenger because the incumbent U.S. Senator supports a handgun ban; and the American Civil Liberties Union creates a Web site telling the public to vote for a Presidential candidate in light of that candidate's defense of free speech.

Citizens United, 130 S. Ct. at 897. One point of contention between the majority and dissent was whether the statute at issue could penalize a corporation or labor union for publishing a book within the proscribed period; Justice Kennedy obviously thought it did (see quote above), while Justice Stevens strongly disagreed. *See id.* at 944 n.31. Justice Stevens, however, based his disagreement primarily on a statement in 11 CFR §100.29(c)(1) that "electioneering communication does not include communications appearing in print media," which does not self-evidently include books, since the term "print media" is generally associated with newspapers and/or news magazines. Moreover, the government apparently conceded in oral argument that a book advocating the election or defeat of a federal election candidate would be covered by the statute and could potentially be banned during the blackout period, especially if the book were transmitted in electronic form (such as via satellite to an electronic reading device). *See, e.g.*, 2009 WL 760811, *29. It therefore does not seem to be quite so clear, as Justice Stevens would have it, that by its plain terms, § 203 does not apply to books, and the majority's concern about the statute's over-breadth, to say the least, was not implausible.

19 For earlier precedents establishing that speech by corporations is entitled to First Amendment protection, *see, e.g.,* First Nat. Bank of Boston v. Bellotti*,* 435 U.S. 765, 778 n.14 (1978) (listing numerous prior decisions invalidating laws infringing protected speech by corporate bodies). As was recognized already in *Bellotti*, the express language of the First Amendment does not protect speakers, but speech itself; if the speech at issue is protected (and there could be no doubt that, in the abstract at least, the type of political speech at issue in *Citizens United* would be protected under the First Amendment), then the identity and nature of the speaker is irrelevant under the First Amendment. As noted in the text, despite the pre-*Citizens United* Supreme Court cases holding that corporate speech is protected by the First Amendment, critics of the decision often appeared to characterize it as originating this view. *See, e.g.,* Editorial, *The Court's Blow to Democracy*, N.Y. Times, Jan. 21, 2010 ("The majority is deeply wrong on the law. Most wrongheaded of all is its insistence that corporations are just like people and entitled to the same First Amendment rights."); Ronald Dworkin, *The "Devastating" Decision,* N.Y. Rev. Books, Feb. 25, 2010 ("The nerve of [Justice Kennedy's] argument—that corporations must be treated like real people under the First Amendment—is in my view preposterous. Corporations are legal fictions. They have no opinions of their own to contribute and no rights to participate with equal voice or vote in politics.").

20 130 S. Ct. at 960 (Stevens, J., dissenting) (acknowledging that "speech does not fall entirely outside the protection of the First Amendment merely because it comes from a corporation," and noting that "no one suggests the contrary and that neither *Austin* nor *McConnell* held otherwise").

21 For dissent's view, see, e.g., *id.* For majority's view, see, e.g., 130 S. Ct. at 898.

22 See Part IV of Justice Scalia's majority opinion. 130 S. Ct. at 913-917.

23 *See, e.g.*, Ronald Dworkin, *The "Devastating" Decision*, N.Y. Rev. Books, Feb. 25, 2010 (The "five right-wing Supreme Court justices have now guaranteed that big corporations can spend unlimited funds on political advertising in any political election."); *see also The "Devastating" Decisions: An*

Exchange, N.Y. Rev. Books, Apr. 20, 2010; Ronald Dworkin, The Court's Embarrassingly Bad Decisions, N.Y. Rev. Books, May 26, 2011.

24 131 S. Ct. 2541 (2011).

25 As discussed below, the Rule 23 aspect of the decision is probably more consequential because it will govern all class actions going forward, while the ruling on commonality was tied to the specific facts of Wal-Mart's operations.

26 *Dukes*, 131 S. Ct. at 2557.

27 *See id.* at 2561. In her opinion, Justice Ginsburg, joined by Justices Breyer, Sotomayor, and Kagan, makes clear her agreement with the majority on this point:

> The class in this case, I agree with the Court, should not have been certified under Federal Rule of Civil Procedure 23(b)(2). The plaintiffs, alleging discrimination in violation of Title VII, 42 U.S.C. § 2000 *et seq.,* seek monetary relief that is not merely incidental to any injunctive or declaratory relief that might be available.

28 While the liberals' decision in this regard has not been the focus of great criticism, it has not been ignored. *The New York Times,* in its editorial decrying the decision, *see supra* note 15, does point out that the Court unanimously dismissed "the suit as the plaintiffs presented it," thus giving "Wal-Mart what it wanted from the court." The *Times*, primarily focusing on where the majority and liberal Justices differed, refers to the liberals as "the four moderates on the court." However, as Justice Scalia pointed out in his opinion, the plaintiffs' decision to proceed under Rule 23(b)(2) was not only harmful to Wal-Mart, by denying it the ability to defend itself against individual claims, *Dukes*, 131 S. Ct. at 2560-61, but was also harmful to the potential class members (i.e., Wal-Mart employees) themselves by limiting the types of claims they could bring, *id.* at 2559, and by divesting them of the ability to opt-out of the class action. *Id.* at 2558 (noting that a (b)(2) class is mandatory and that the rule "does not even oblige the District Court to afford [putative class members] notice of the action").

29 Although Justice Ginsburg stated at the outset of her opinion that whether the plaintiffs satisfied the requirements of Rule 23(b)(3) was not before the Court and should be considered on remand, *id.* at 2561 (Ginsburg, J., concurring in part and dissenting in part), she clearly considered that the evidence met the requirements of 23(b)(3). *Id.* at 2565.

30 For basic facts of the case, see *id.* at 2547-2549 (majority opinion). The putative class included "*all* Wal-Mart's female employees," and, therefore, the class included women in management positions (not to mention women who may have prospered at Wal-Mart). *See id.* at 2549 (emphasis in original). For the plaintiffs' sociological expert's inability to tell what percentage employment decisions were determined by stereotypical thinking, *see id.* at 2553.

31 A Supreme Court Scorecard, Posting of Linda Greenhouse to Opinionator, http://opinionator.blogs.nytimes.com/2011/07/13/a-supreme-court-scorecard/ (July 13, 2011, 21:30 EDT). As Ms. Greenhouse put it:

> Whatever its merits, the nationwide class action of 1.5 million women (representing "all women employed at any Wal-Mart domestic retail store at any time since Dec. 26, 1998") was an accident waiting to happen from the minute it showed up on the radar screen of a Supreme Court that is deeply skeptical of litigation, particularly of lawsuits that appear designed to achieve broad policy aims.
>
> One indispensable test of whether a lawsuit may proceed as a class action is whether question of law or fact are common to all plaintiffs. In addition to the huge size of the class in Wal-Mart Stores v. Dukes, the essence of the plaintiffs' complaint made the case vulnerable. The claim was not that Wal-Mart's policies actively discriminate against women in pay and promotion, but rather that headquarters leaves local managers with too much discretion on those matters. In other words, the problem was said to be not the existence but the absence of a uniform company-wide employment policy.

Id. (In fact, of course, the company did have a uniform, company-wide policy against sex-discrimination.) Ms. Greenhouse, who appears to be favorable to the dissent's view that certification could have been ordered based on the fact that women at Wal-Mart faced "unsupervised bosses who *might* base personnel judgments on unconscious stereotypes as well as conscious prejudices," *id.* (emphasis added), nonetheless conceded that while this was a "strong argument," "it would have been a surprise had it prevailed." *Id.*

32 131 S.Ct. 1740 (2011).

33 In addition to Justice Scalia's majority opinion and Justice Breyer's dissent, Justice Thomas filed a concurring opinion joining his conservative colleagues in reversing the Ninth Circuit decision in the case.

34 The rule invalidating class action waivers in consumer arbitration agreements was announced by the California Supreme Court in *Discover Bank v. Superior Court,* 113 P.3d 1100 (Cal. 2005). In that case, the California Supreme Court held that a class action waiver in a consumer arbitration agreement is unconscionable as a matter of law if it is contained within a "consumer contract of adhesion," if the claim is of a kind that "predictably involve[s] small amounts of damages," and "when it is alleged that the party with the superior bargaining power has carried out a scheme to deliberately cheat large numbers of consumers out of individual small sums of money . . . ." *Id.* at 1110. (California broadly defines a contract of adhesion as any unilateral, "take it or leave it" contract, in which the "part of superior bargaining strength[] relegates to the subscribing party only the opportunity to adhere to the contract or reject it." *Id.* at 1108 (internal quotations and citations omitted).)

35 *See* 9 U.S.C. § 2; Rent-A-Center, West, Inc. v. Jackson, 130 S. Ct. 2772, 2776 (2010) ("The FAA . . . places arbitration agreements on an equal footing with other contracts . . . and requires courts to enforce them according to their terms . . . . Like other contracts, however, they may be invalidated by *generally applicable contract defenses,* such as fraud, duress, or unconscionability." (emphasis added)).

36 Unlike the rigid rule applied by the lower federal courts in *AT&T Mobility*, the generally-applicable unconscionability standard under California law involves a "sliding scale" approach, under which courts weigh varying degrees of procedural and substantive factors to determine a particular contract's overall fairness at the time of its formation. *See* Armendariz v. Found. Health Psychcare Servs., Inc., 6 P.3d 669, 690 (Cal. 2000); Cal. Civ. Code § 1670.5(A) ("If the court as a matter of law finds the contract or any clause of the contract to have been unconscionable at the time it was made, the court may refuse to enforce the contract.").

37 While Justice Scalia's view was that, contrary to the California rule, "[r]equiring the availability of class wide arbitration interferes with fundamental attributes of arbitration and thus creates a scheme inconsistent with the FAA," *Dukes*, 131 S. Ct. at 1748, he acknowledged, of course, that if the parties to an arbitration agreement agree to a class proceeding, the arbitration must go forward on that basis. *See, e.g.*, *id.* at 1751. For the majority, the California rule operated, in effect, to permit consumers to require class arbitration in consumer contracts irrespective of the parties' agreement. "Although the [*Discover Bank*] rule does not *require* arbitration, it allows any party to a consumer contract to demand it *ex post.*" *Id.* at 1750.

38 Thus, for example, in this case, the Court applied the conclusions regarding the need for consent to class action arbitration that it had reached in *Stolt-Nielsen S.A. v. AnimalFeeds Int'l Corp.*, 130 S. Ct. 1758 (2010), even though the latter case was purely a dispute between businesses.

39 It appears that one of the reasons the plaintiffs in *AT&T Mobility* wished to take advantage of the *Discover Bank* per se rule was that the agreement at issue contained unusually consumer-friendly terms and was therefore unlikely to be found to be unconscionable under the traditional tests. Indeed, in this very case, the Ninth Circuit itself expressly acknowledged that AT&T Mobility's agreement was consumer-friendly. Laster v. AT&T Mobility LLC, 584 F.3d 849, 855-56, 856 n.9, n.10 (9th Cir. 2009). In an earlier case that reviewed the same agreement and rejected an unconscionability challenge, the federal district court noted that AT&T Mobility's agreement "contains perhaps the most fair and consumer-friendly provisions this Court has ever seen." Makarowski v. AT&T Mobility, LLC, 2009 WL 1765661, at *3 (C.D. Cal. June 18, 2009).

40 131 S. Ct. 2567 (2011).

41 *Id.* at 2576.

42 *Id.* at 2577-78.

43 *Id.,* at 2579.

44 As Justice Thomas, writing for the Court, put it:

> We recognize that from the perspective of [the plaintiffs], finding pre-emption here but not in *Wyeth* makes little sense. . . . We acknowledge

the unfortunate hand that federal drug regulation has dealt [the plaintiffs] and others similarly situated. . . . But it is not this Court's task to decide whether the statutory scheme established by Congress is unusual or even bizarre.

Id. at 2581-82.

45 *Id.* at 2588-2589.

46 *Id.* at 2578.

47 From the majority's point of view, taking into account the process the generic manufacturers could have initiated to strengthen their labels would ultimately "render conflict analysis largely meaningless because it would make most conflicts between state and federal law illusory," since it would subject conflict preemption to a conjectural analysis of what *might* have happened had a request to change the federal requirements been made. *Id.* at 2579.

48 Even the majority describes this position as a "fair argument," although it rejects it. *Id.*

49 See, e.g., Justice Thomas's extended discussion of the Supremacy Clause, in which he argues that "the phrase 'any [*state* law] to the Contrary notwithstanding' is a *non obstante* provision," suggesting that "federal law should be understood to impliedly repeal conflicting state law." *Id.* at 2579-80.

50 It should be mentioned, as pertinent to the subject of this article, that a report was issued in December 2010 entitled "Is the Roberts Court Pro-Business?" The report, which is available online at epstein.usc.edu/research/RobertsBusiness.pdf, was authored by three prominent scholars, Lee Epstein, William M. Landes, and Richard A. Posner, who analyzed those cases in the U.S. Supreme Court Database that are categorized as dealing with "Economic Activity." The report's conclusion was that, based on the data reviewed, "it might be reasonable to conclude that the current Court is distinctly favorable toward business interests." Although dealing with the same topic under discussion here, in fact no reliance was placed upon the Epstein, Landes, and Posner report or its conclusion during the preparation of this article for the simple reason that their report utilizes categories and definitions, as well as a methodology, that, to the author of this article, seem flawed. For example, the report is based upon the view that pro-business decisions are always conservative and anti-business decisions are always liberal (even though as demonstrated above the liberal/conservative division does not consistently match pro/anti-business results). Furthermore, even the definitions of liberal and conservative can be problematic. Thus, for example, Epstein, Landes, and Posner use a definition of liberal decisions as "anti-business, anti-employer, pro-liability, pro-competition, pro-consumer, etc.," though they admit in a footnote that the definition is imperfect. The most serious flaw of the report, however, is that its methodology appears to ignore not only the substantive issues in the cases it reviews, but the statutory and regulatory background of each matter. A decision that might be counted as conservative under the Epstein, Landes, and Posner rubric might be compelled by a statutory provision enacted by Congress or, as in the unanimous Fed. R. Civ. P. 23 holding in *Wal-Mart Sores v. Dukes*, the result of the wording of a rule, rather than the result of the Justices' liberal or conservative outlooks. Without looking more deeply into the cases, any analysis undertaken to demonstrate the existence or lack of bias will reveal little beyond the identity of the prevailing parties, which as shown above is very far from the whole story.

Florida Constitutional Challenge to Obamacare: It All Comes Down to Broccoli

*By Karen R. Harned**

Note from the Editor:

This paper provides an update on the litigation dealing with the Patient Protection and Affordable Care Act signed into law in 2010, and examines some of the arguments made in the case, which is now headed to the U.S. Supreme Court. As always, The Federalist Society takes no position on particular legal or public policy initiatives. Any expressions of opinion are those of the author. The Federalist Society seeks to foster further discussion and debate about the health care litigation. To this end, we offer links below to briefs and court opinions on various sides of this issue and invite responses from our audience. To join the debate, you can e-mail us at info@fed-soc.org.

Related Links:

- Florida *ex rel.* Attorney General v. U.S. Department of Health & Human Services, Nos. 11-11021 & 11-11067 (11th Cir. Aug. 12, 2011): http://www.ca11.uscourts.gov/opinions/ops/201111021.pdf
- Virginia *ex rel.* Cuccinelli v. Sebelius, Nos. 11-1057 & 11-1058 (4th Cir. Sep. 8, 2011): http://pacer.ca4.uscourts.gov/opinion.pdf/111057.P.pdf
- Seven-Sky v. Holder, No. 1:10-cv-00950 (D.C. Cir. Nov. 8, 2011): http://www.cadc.uscourts.gov/internet/opinions.nsf/055C0349A6E85D7A8525794200579735/$file/11-5047-1340594.pdf
- Thomas More Law Center v. Obama, No. 10-2388 (6th Cir. June 29, 2011): http://www.ca6.uscourts.gov/opinions.pdf/11a0168p-06.pdf
- Petition for Certiorari, U.S. Department of Health & Human Services v. Florida, No. 11-398 (U.S. Sept. 2011): http://www.justice.gov/osg/briefs/2011/2pet/7pet/2011-0398.pet.aa.pdf
- Brief of State Respondents, U.S. Department of Health & Human Services v. Florida, No. 11-398 (U.S. Oct. 17, 2011): http://www.supremecourt.gov/docket/PDFs/11-398%20BIO%20States.pdf
- Reply Brief for Petitioners, U.S. Department of Health & Human Services v. Florida, No. 11-398 (U.S. Oct. 2011): http://www.justice.gov/osg/briefs/2011/2pet/7pet/2011-0398.pet.rep.pdf

On August 12, 2011, the Eleventh Circuit Court of Appeals ruled that Congress exceeded its authority by forcing all Americans to purchase health insurance through the health care law's "individual mandate." A 2-1 majority held that enacting the individual mandate was beyond Congress's power under the Commerce Clause. However, the court held that while the individual mandate was unconstitutional, it was severable from the law as a whole, and the rest of the law could stand.

Just a year earlier, then-U.S. House of Representatives Speaker Nancy Pelosi dismissed the idea that there was even a question regarding the constitutionality of this law, which dramatically changes 17.6% of the nation's economy.[1] As of the writing of this article, thirty-one challenges to the health care law have been filed in federal courts across the country, with mixed results. The Eleventh Circuit was the first appellate court to find the law unconstitutional—both the Sixth Circuit and District of Columbia Circuit previously upheld the law. On November 14, 2011, the U.S. Supreme Court announced it would hear the case in its 2011-2012 Term. As a result, a final decision on the law's constitutionality is expected in June 2012, just as the presidential election is in full swing.

** Karen R. Harned is the Executive Director of the National Federation of Independent Business (NFIB) Small Business Legal Center. NFIB is one of the parties in the health care litigation discussed in this article.*

The Individual Mandate and the Commerce Clause

"The powers delegated by the . . . Constitution to the federal government are few and defined," whereas "[t]hose which . . . remain in the State governments are numerous and indefinite."[2] These principles must apply to the interpretation of the Commerce Clause, which bestows upon Congress the power to regulate commerce with foreign nations, among the several states, and with the Indian tribes.[3] For example, the Commerce Clause was intended to be a "negative and preventative provision against injustice among the states themselves, rather than as a power to be used for the positive purposes of the General Government."[4] Here, the ACA's mandate invades the states' traditional protection of their citizens' health and welfare by compelling individuals to enter into contracts to subsidize the insurance industry and its voluntary customers.[5] This exercise of plenary police power is not authorized by the Constitution.

The notion that the Commerce Clause permits Congress only to regulate interstate activities was quashed in modern-day Commerce Clause jurisprudence. In *Wickard v. Filburn*, the Supreme Court held that even where a farmer, Filburn, was growing wheat on his own farm for personal consumption with no intent to sell it, Filburn's wheat-growing activities reduced the amount of wheat he would buy. Wheat was traded nationally, and, as a result, Filburn's production of excess wheat affected interstate commerce. Therefore, it could be regulated

by the federal government under the government's Commerce Clause powers.[6]

Similarly, in *Gonzales v. Raich*, the Supreme Court held that the Commerce Clause permitted the government to criminalize the production and use of homegrown cannabis, even when permitted for medicinal use.[7] No longer was interstate activity required for Congress to have regulatory power. Instead, Congress was empowered to regulate business activity that was purely local, if the aggregate activity had a substantial effect on interstate commerce. As a result, there are essentially no limits to Congress's ability to regulate as long as the commerce being regulated constitutes an "activity."

In *United States v. Lopez*, the Supreme Court set limits to Congress's Commerce Clause power, holding that Congress did not have the power to regulate the carrying of handguns in school zones without providing a sufficient link to interstate commerce.[8] The *Lopez* Court set out three categories of activity that Congress was empowered to regulate: "the *channels* of interstate commerce, the *instrumentalities* of interstate commerce, or *persons or things* in interstate commerce, and activities that *substantially affect* or *substantially relate to* interstate commerce."[9] Relying on the *Lopez* decision, the Supreme Court held in *United States v. Morrison* that Congress did not have the authority to enact the Violence Against Women Act (VAWA) because the acts of violence that the VAWA dealt with did not have a substantial enough effect on interstate commerce.[10]

In summary, Commerce Clause jurisprudence to date has focused exclusively on activities. The main debate in Commerce Clause cases to date is over whether or not an activity "substantially affects" interstate commerce and not whether the subject of regulation constitutes activity in the first place. Thus, the decisions to grow wheat and produce cannabis are activities that substantially affect interstate commerce and can therefore be regulated. Conversely, the decisions to carry a handgun or commit violence, while activities, do not substantially affect interstate commerce and cannot be regulated. In none of these cases was inactivity or the decision not to engage in activity discussed. The Government's contention that a non-activity such as the decision not to buy health insurance can be regulated by Congress is a novel idea in Commerce Clause jurisprudence. Thus, even under the Supreme Court's broadest conception of the Commerce Clause, typified by *Raich* and *Gonzalez*, the individual mandate cannot be justified.

The Commerce Clause Does not Provide Constitutional Justification for the Mandate

Congress' power to regulate under the Commerce Clause broadly allows Congress to regulate interstate commerce as well as address other conduct that "substantially affect[s] interstate commerce."[11] All that needs be considered is the aggregate effect of particular conduct on interstate commerce—Congress need not consider whether and to what extent individual actors contribute to that conduct.[12] Moreover, the courts have a "modest" role in reviewing Commerce Clause litigation. All that is required is a finding that Congress had a "rational basis" for concluding conduct has a substantial impact on interstate commerce. The courts similarly should give broad deference to Congress regarding the means chosen to achieve a legitimate end.

The individual mandate is not justified by the Commerce Clause because forcing individuals to buy health insurance is not a regulation of commerce. Under controlling precedent, Congress may regulate under its commerce power: (1) "the use of the channels of interstate commerce"; (2) the operation of "the instrumentalities of interstate commerce"; and (3) "those activities that substantially affect interstate commerce."[13] Yet none of these "categories of activity" encompasses the inactivity regulated by the mandate—i.e., the failure to purchase health insurance.

The Mandate Does Not Regulate Commerce

The Government justified Congress's passage of the ACA by arguing that Congress intended to regulate the health insurance and health care markets to ameliorate the cost-shifting problem brought about by individuals who do not have insurance but at some time seek medical care for which they cannot pay.[14] Furthermore, the Government defended the individual mandate as constitutional because it regulates "quintessentially economic" activity related to an industry of near universal participation and does not compare to the regulations in *Lopez* and *Morrison*, which only touched on criminal conduct, not economic activity. Congress only mandated how Americans finance their inevitable healthcare needs.[15] Embedded in the Commerce Clause, the Government contended, is the power to override ordinary economic decisions and redirect funds Americans would spend anyway to other purposes.[16] Thus, the Commerce Clause gives Congress the power to direct and compel an individual's spending in order to further its regulatory goals.[17]

The plaintiffs argued that the mandate does not regulate commerce itself, in either its interstate or intrastate channels or instrumentalities. Rather the mandate compels the uninsured to participate in the health insurance market. The Commerce Clause gives Congress the power to regulate interstate commerce, but not the power to force individuals to enter into commerce. For example, while Congress may regulate the terms of voluntary contracts between General Motors and its customers, it may not compel individuals to enter into purchase contracts with GM because there is no pre-existing "commerce" to regulate. Otherwise, Congress could force individuals to purchase any product, from GM cars to Citibank mortgages to broccoli. Compelling commerce, here the ACA's punishment of individuals for not buying insurance, is not *regulating* commerce.[18] Since the "regulated" individuals have not entered the insurance market, there is no "commerce" for the ACA to regulate.[19]

The Individual Mandate Regulates the Non-Purchase of Health Insurance, Not the Non-Payment for Healthcare

One defense of the individual mandate proposed by Congress was that it ameliorated the cost-shifting problem caused when people sought medical care they could not pay for. This is the so-called "free rider" problem. The mandate, the Government argued, only directs how Americans would finance

their inevitable health care needs. The plaintiffs countered that the mandate does not regulate how individuals pay for health care, but instead only their failure to obtain health insurance.[20] The mandate penalizes them for not having health insurance in a given month, even if they obtained no medical care that month.[21] Conversely, it does not penalize those who do have insurance but have failed to pay their medical bills.[22]

The Eleventh Circuit's ruling sided with the plaintiffs. The majority found that the language of the individual mandate does not really regulate "how and when health care is paid for," but instead regulates those who have not entered the health care market at all.[23] Indeed, the majority found that the mandate "does not even require those who consume healthcare to pay for it with insurance."[24] Thus, the mandate actually regulates the non-purchase of health insurance, a totally different subject matter than that proposed by the Government.[25]

The majority also observed that the primary targets of the individual mandate were not the so-called "free riders," who obtain medical care without paying for it, but are actually individuals who are relatively healthy and wealthy but were compelled to enter into contracts to subsidize insurance companies.[26] The plaintiffs argued that the mandate was not crafted to regulate how people pay for their health care, but was instead a tactic to subsidize the insurance industry by forcing healthy individuals to enter into insurance contracts.[27] The majority agreed: the mandate forces non-free riders to "pay insurance premiums now to subsidize the private insurers' cost in covering more unhealthy individuals under the Act's reforms."[28]

The Mandate and Regulation of a Class of Economic Activities that Substantially Affects Interstate Commerce

Substantial Effects Doctrine

Modern Commerce Clause jurisprudence holds that only an economic activity that substantially affects interstate commerce may be regulated by Congress. The Government claimed that the mandate meets this test because the status of being uninsured "substantially affects" interstate commerce and thus falls within Congress's commerce power.[29] However, the plaintiffs argued that the Government's line of reasoning is at odds with the Supreme Court's "substantial effects" precedent and would eviscerate the entire concept of enumerated powers.

Congress's enumerated power to regulate "interstate commerce" does not necessarily confer power to regulate "things affecting interstate commerce."[30] The "substantial effects" doctrine allows Congress to regulate intrastate activities affecting interstate commerce only to effectuate the execution of its enumerated power to regulate interstate commerce.[31] A common issue under which courts have allowed Congress to exercise its Commerce Clause powers occurs when the aggregate effect or a product's local use adversely "influences the prices and market conditions" desired by Congress.[32] Since local and national products are fungible, the substantial effects doctrine eliminates the distinction between intrastate and interstate commerce.[33]

Supreme Court precedent demonstrates this point. In a line of cases illustrated by *Wickard* and *Raich*, the Court allowed for a broader interpretation of the "substantial effects" doctrine. For example, in *Wickard*, while aiming to increase wheat prices nationally, Congress restricted the amount of wheat that farmers could grow, even if for personal use.[34] The Court upheld the restriction because local wheat production would obstruct Congress's goal of raising interstate prices[35] because local wheat production both increased the supply of wheat that could be sold interstate and decreased demand for purchasing wheat intrastate.[36] Similarly, in *Raich*, Congress's attempt to eliminate the interstate market for marijuana was undercut by intrastate manufacture and possession of state-law-authorized medical marijuana.[37] While numerous regulations have been upheld under the "substantial effects" doctrine, it is critical that none have involved regulation of individuals who neither participate in commerce nor pose barriers to commerce.[38]

Contrarily, in *Lopez* and *Morrison*, the Court demonstrated that the doctrine has limits when it clarified that some barriers to commerce may *not* be regulated under the "substantial effects" doctrine. For instance, *Lopez* invalidated a law banning gun possession near schools, and *Morrison* invalidated a law providing civil remedies for violence against women.[39] Even though these activities certainly had substantial negative effects on the United States' commercial productivity, the regulated individuals had not engaged in any "economic activity" resembling the type of "commerce" that Congress could regulate at the interstate level.[40]

Accordingly, the plaintiffs argued that there are three reasons why the "substantial effects" doctrine does not confer upon Congress the power to compel individuals to purchase health insurance. First, the "substantial effect" of not participating in commerce is not a barrier to commerce.[41] Second, not participating in the insurance market is *not* "economic activity."[42] Finally, expanding the doctrine to include *not* purchasing a product would create plenary federal power.[43]

A majority of the Eleventh Circuit panel agreed with the plaintiffs when it found that regardless of the uninsured's effects on interstate commerce, the uninsured lacked a sufficient connection, or nexus, to interstate commerce.[44] The majority stressed that what matters is the regulated subject matter's connection to interstate commerce; that connection is lacking in the case of the individual mandate.[45] "Under any framing, the regulated conduct is defined by the *absence* of both commerce or even 'the production, distribution, and consumption of commodities'—the broad definition of economics in *Raich*."[46]

Non-Participation in the Health Insurance Market Is not Economic Activity

The plaintiffs argued that inactivity in the health insurance market is non-economic activity and thus not reachable by the government through the Commerce Clause. The non-purchase of health insurance is not "economic activity" according to the Supreme Court because it is not "the production, distribution, [or] consumption of commodities."[47] In fact, the non-purchase of insurance is even less connected with commerce than gun possession in *Lopez* since guns can only be possessed after being produced, distributed and acquired through commercial

transactions.[48] Since Lopez's possession of a gun in a school zone was not economic activity, the same conclusion must follow for the uninsured's inactivity, which "continues their estrangement from the insurance market and thus leaves them even more remote from commerce than was Lopez."[49]

Surprisingly, the Eleventh Circuit quickly dispensed with the activity/inactivity debate in this case. Even though it recognized that the Supreme Court's Commerce Clause jurisprudence uniformly involved Congress attempting to regulate pre-existing activities, it found this formalistic dichotomy insufficient to decide this case.[50] The majority observed that to the extent that uninsured Americans can be described as "active" in the insurance market, their activity is the absence of their participation in the market.[51] Thus, the majority concluded, the individual mandate could not be clearly classified as regulating either economic activity or noneconomic activity.[52]

Non-Participation in the Insurance Market Does not Burden or Obstruct Commerce

Additionally, the plaintiffs argued that non-participation in the health insurance market is not an activity that obstructs or burdens commerce. Persons whose intrastate activities adversely affect Congress' preferred market conditions can affect interstate commerce by "imposing burdens and obstructions" or by creating "potential stimulants."[53] However, individuals without health insurance are not involved in the health insurance market at all and thus do not stimulate or obstruct its operation.[54] Because market non-participants, like the uninsured, impose no barriers to interstate commerce, regulating them is not justified as a prophylactic execution of Congress' commerce power.[55] Analogously, Congress could not force urban pedestrians to purchase car insurance on the theory that their refusal to do so burdens the car insurance market because they were selecting out of the "risk pool."[56] In fact, the Government conceded that individuals do not engage in commerce when declining to purchase insurance; they also do not affect commerce when making that same decision.[57] The majority observed that because of the realities of the modern marketplace, the decision not to buy insurance, when aggregated, would substantially affect the insurance market.[58] However, the majority found that it would need to apply the aggregation principle to citizens *outside* the stream of commerce.[59] This application, it said, had the danger of making Congress's power to regulate unlimited.[60]

The "Substantial Effects" Doctrine Cannot Be Expanded to Cover Non-Participation in the Health Insurance Market without Federal Power Becoming Plenary.

Finally, the plaintiffs contended that if the "substantial effects" doctrine is used to uphold the individual mandate, the government would be granted plenary regulatory power and the concept of limited federal government would be eviscerated. Since all inactivity could be deemed to substantially affect interstate commerce, Congress could require *any* purchasing decision. If not purchasing health insurance is reachable under the Commerce Clause, one would be "hard pressed to posit any [in]activity by an individual that Congress [would be] without the power to regulate."[61] In fact, the Congressional Budget Office already recognized that the individual mandate could lead to a "command economy, in which the President and Congress dictated how much each individual and family spent on all goods and services."[62] This, noted the majority, underscored the necessity of a strong nexus between the regulated activity and interstate commerce. Otherwise, the government could require the purchase of any product, given that the aggregated effect of the non-purchase of any good will always have a substantial effect on commerce.[63]

The Government appreciated the far-reaching consequences of the ACA, but argued that the individual mandate did not involve the creation of a plenary power for Congress. In its explanation, the Government essentially argued that the mandate is an emergency tool for use in an extreme and unique situation. The Government argued that health care and the health insurance industry are unique. Therefore, the mandate is not likely to lead to a plenary power because of the inevitability of the need for health care, the unpredictabilitiy of that need, the high costs associated with health care, the federal requirement that hospitals treat uninsured individuals, and the resulting cost-shifting.[64]

The majority rejected this line of reasoning, however, because from a doctrinal standpoint, there is no way to limit the Government's theory to decisions not to purchase health insurance.[65] The five factual criteria posited by the Government to make the health care market appear "unique" are not "limiting principles rooted in any constitutional understanding of the commerce power."[66] Thus, if the Government's position was adopted, Congress could have plenary power over all economic decisions because there is no way to cabin the Government's rationale whether or not the health care market is unique.[67]

The Mandate and the Necessary and Proper Clause

Congress may "make all Laws which shall be necessary and proper for carrying into Execution [its enumerated] Powers . . . ."[68] The Necessary and Proper Clause confirms that Congress has "incidental power" to further the legitimate end of executing its enumerated powers through appropriate means that are plainly adapted and consistent with the spirit of the Constitution.[69] The plaintiffs maintained that the mandate is neither necessary nor proper and thus that the Government's reliance on the clause is nothing more than a last-ditch effort to "defend ultra vires congressional action."[70]

The Mandate Is Not a Necessary Means of Carrying the ACA's Commercial Regulations into Execution

The Government relies heavily on the breadth of the Necessary and Proper Clause to ultimately save the individual mandate. The individual mandate is not necessary to serve the end of carrying the ACA into execution. The Government contended that "Congress's power extends to the regulation of even 'noneconomic local activity' otherwise beyond the reach of the commerce power" where "needed to make [a] regulation [of interstate commerce] effective" because "failure to regulate [the uninsured] would undercut the regulation of the [insurance] market."[71] Under this "effective regulation" doctrine, Congress may regulate economic and noneconomic activity, but only if doing so is a "necessary part of a more

general regulation of interstate commerce" because the activity interferes with, obstructs, or undercuts the regulatory scheme.[72] Here, the uninsured's activity does not affect congressional regulation of the interstate health insurance market because the uninsured neither impede nor frustrate regulation of market participants.[73]

The Government responded that Congress found that eliminating the uninsured was essential to cure the problem that insurers would lose money due to individuals who postpone purchasing insurance until the need for it arose.[74] However, the plaintiffs argued, and the 11th Circuit agreed, that the uninsured are not interfering with Congress's efforts to regulate insurers.[75] The majority noted that the individual mandate is not designed to enable the execution of the ACA's regulations, but is actually designed to "counteract the significant regulatory costs on insurance companies and adverse consequences stemming from the fully executed reforms."[76]

The Necessary and Proper Clause does not authorize Congress to pursue ends outside of its legitimate, enumerated powers. Yet, it is an illegitimate end to offset a regulatory scheme's costs for market participants by compelling third parties who are not part of the scheme to participate.[77] This is especially true where, as here, those third parties are not barriers to the effective execution of the regulatory scheme.[78] There is no regulation being affected, or commerce being regulated, by forcing uninsured individuals to participate in the health insurance market in order to subsidize participants in the market.[79] Congress's powers cannot be enhanced solely because it created costs that need to be offset.[80]

The Necessary and Proper Clause does not permit Congress to reduce a regulatory scheme's cost for market participants by regulating strangers to the scheme.[81] For example, in *United States v. DeWitt*, the Government defended a federal ban on the intrastate sale of certain illuminating oils.[82] The ban was defended on the basis that it aided and supported the federal tax imposed on other illuminating oils because eliminating competition from the banned oils would increase production of the taxed oils and therefore increase tax revenue.[83] Similar to the individual mandate, the regulation of some individuals was justified because it would support others that were burdened by the government's regulation and would make the scheme more effective.[84] The Court found that the ban was not permissible because it was not an appropriate and plainly-adopted means for carrying out Congress's power to lay and collect taxes.[85]

The plaintiffs contended that any contrary conclusion would convert the Necessary and Proper Clause into a "vehicle for Congress to pursue 'ends' beyond its enumerated powers."[86] The plaintiffs further contended that the mandate is not a means to accomplishing the end of regulating commerce in insurance but is instead imposed to counteract the costs imposed by the ACA.[87] This means that the mandate's justification actually arises from the ACA itself.[88] However, Congress's powers are derived from the Constitution, not from the statutes it passes.[89] "While Congress may broadly regulate interstate commerce under the Necessary and Proper Clause, it cannot use such regulation *to bootstrap* additional regulatory powers otherwise beyond Congress' reach."[90]

The Mandate Is Not a Proper Means of Carrying the ACA's Commercial Regulations into Execution

The individual mandate is not a proper means of carrying the ACA into execution. Laws are only "proper" under the Necessary and Proper Clause if they employ means that are "plainly adapted to [the legitimate] end, which are not prohibited, but consist[ent] with the letter and spirit of the constitution."[91] Under *McCulloch v. Maryland*, a "proper" law is (1) an ordinary method of execution that respects (2) the states' sovereignty and (3) the People's liberty. The plaintiffs contend that the individual mandate fails each of these factors.

First, the individual mandate is not plainly adapted. A regulation is "plainly adapted" if it invokes the "ordinary means of execution."[92] The Necessary and Proper Clause merely confirms the existence of "incidental or implied powers" to execute Congress's stated authority; the powers most readily "deduced from the nature of the objects themselves" are the "ordinary means of execution."[93] The plaintiffs argued that the mandate is far from being "incidental," "implied," and certainly not "plainly adapted" since it "plows thoroughly new ground and represents a sharp break with the longstanding pattern of federal . . . legislation."[94] The mandate is entirely unprecedented. In fact, not even the Congress that passed the New Deal thought of supporting wheat prices by forcing Americans to purchase wheat.[95]

Second, the individual mandate does not properly account for state interests and in fact tramples upon the states. The plaintiffs contend that the mandate demonstrates an acute disrespect for state interests by "foreclosing the states from experimenting and exercising their own judgment in an area which states lay claim by right of history and expertise."[96] The mandate invades the areas of health insurances and citizens' health and safety, which are typically left to the control of states.[97] The mandate contravenes at least the twenty-six state plaintiffs and represents a considerable federal intrusion into states' traditional authority to regulate for the health of their citizens.[98]

Finally, the individual mandate is not a proper means of executing Congress's enumerated powers because it unduly infringes on the liberty of individuals. The plaintiffs argue that one of the most fundamental rights enjoyed by Americans is their "freedom from being forced to give their property to, or contract with, other private parties."[99] The Supreme Court has observed that Congress should not be able to force a contract on an individual without his consent because the consent of the parties to be bound is the essence of a contract.[100] The plaintiffs contend that these interests are "gravely threatened" by the ACA's individual mandate, which compels citizens "to contract on disadvantageous terms with wealthier insurers to reduce costs" that are not related to any injury sustained or caused by those individuals.[101] The mandate tramples on the rights of the affected individuals and thus is not appropriately in the reach of the Congress's power under the Necessary and Proper Clause.[102]

The Mandate as a Tax

In addition to its Commerce Clause justification for the ACA, the Government argued that the individual mandate

could be sustained under Congress's broad taxing power. The Taxing and Spending Clause provides that "Congress shall have the Power to lay and collect Taxes, duties, Imposts and Excises, to pay the Debts and provide for the Common Defence and general Welfare of the United States."[103] This power is plenary and comprehensive, and the fact that the individual mandate has a regulatory purpose is irrelevant because "a tax 'does not cease to be valid merely because it regulates, discourages, or even definitely deters the activities taxed.'"[104] So long as a statute produces revenue, Congress may enact it under the taxing power.[105]

Though Congress's taxing power is indeed plenary, the majority rejected the Government's contention that the individual mandate was a tax. First, the plain language of the statute establishes that the individual mandate is a penalty, not a tax. The majority noted that the language of the ACA repeatedly states that a "penalty" will be imposed on individuals for failing to obtain health insurance.[106] The majority would not construe Congress's choice of language as a careless one-time invocation of "penalty" because the other relevant provisions also describe the mandate as imposing a penalty, not as a tax.[107] The unambiguous text of the individual mandate provides that it imposes a penalty that aims to encourage compliance with the ACA's insurance requirement "by imposing a monetary sanction on conduct that violates that requirement."[108]

Second, even if the text of the ACA were unclear, the legislative history demonstrates that Congress intended to impose a penalty for failure to obtain and maintain health insurance. The majority observed that before the ACA was passed, earlier bills in both houses of Congress proposed an individual mandate that was accompanied by a tax.[109] For example, Section 401 of the "America's Affordable Choice Act of 2009," introduced in the House of Representatives, provided that "there is hereby imposed a tax" on any person who did not maintain minimum health insurance.[110] Furthermore, "America's Healthy Future Act," introduced in the Senate, also used the term "tax."[111] Thus, the majority found it notable that the final version of the ACA discontinued the use of "tax" in favor of "penalty."

The Government responded that the individual mandate is still "a tax in both administration and effect."[112] It claimed that in the process of evaluating the constitutionality of a tax law, the court should only be concerned with the law's practical operation, not its definition or the precise form of descriptive words that may be applied to it.[113] Since the individual mandate will produce revenue and be enforced by the Internal Revenue Service and is collected through taxpayers' annual tax returns, the Government argued that it is a tax.

The majority remained unpersuaded. The Government's claim that the mandate is a tax simply because it has a revenue-raising function was not dispositive because it did little to address the distinction between a tax and a penalty. The majority noted that criminal fines, civil penalties, and taxes all share the same features: "they generate government revenues, impose fiscal burdens on individuals, and deter certain behavior."[114] Furthermore, the majority observed that the individual mandate operates as a monetary sanction for an individual who has failed to obtain insurance.[115] In the majority's view, "such an exaction appears in every important respect to be punishment for an unlawful act or omission," which is a penalty.[116] Finally, the fact that the individual mandate is housed in the Internal Revenue Code and is collected through taxpayers' annual returns is also not dispositive. The Code itself makes clear that Congress's choice of where to place a provision has no interpretive value[117] and not every provision in the Code is a tax.[118] Thus, after review, the majority found that the individual mandate was a regulatory penalty, not a tax, and must find justification in a different enumerated power.

Severability

The district court found that the individual mandate was not severable from the ACA as a whole. The plaintiffs contended that it is well-established that once a court strikes a statute's unconstitutional provisions, the provisions remaining must be invalidated where Congress "would not have enacted those provisions . . . independently of that which is invalid."[119] The plaintiffs argued that in this analysis, courts must inquire whether the severed statute would "function in a manner consistent with . . . the original legislative bargain."[120] The plaintiffs asserted that under these principles the mandate cannot be severed from the ACA as a whole because the mandate "so affects the dominant aim of the statute" that it is inconceivable that Congress would have enacted the ACA without it.[121]

The Government argued that the district court erred when it found the Act to be non-severable.[122] This argument was curious in light of the fact that the Government recognized that the mandate is "integral" to the ACA's regulation of insurance.[123] Nevertheless, the Government maintained that while the mandate is integral to the ACA's operation, certain other provisions are not.[124] The Government proffered various examples: employer-provided rooms for nursing mothers, nondiscrimination protection for providers refusing to furnish assisted suicide services, and Medicare reimbursements for bone-marrow density tests.[125] The plaintiffs countered that the Government "cannot seriously claim that Congress 'would have been satisfied' with this menagerie of tag-along provisions."[126]

The Eleventh Circuit reversed the decision of the district court and found that individual mandate could be severed from the rest of the ACA. The majority began its application with the Supreme Court presumption in favor of severability: courts must "strive to salvage" acts of Congress by severing any constitutionally infirm provisions "while leaving the remainder intact."[127] The Supreme Court's test for severability is as follows: "Unless it is *evident* that the Legislature would not have enacted those provisions which are within its power, independently of that which is not, the invalid part may be dropped if what is left is *fully operative as a law.*"[128]

In its analysis, the majority offered several reasons why the district court erred in finding the individual mandate not severable from the rest of the ACA. At the outset, the court stated that "a lion's share" of the ACA has nothing to do with private insurance, let alone the individual mandate.[129] The majority found that representative samples of such provisions

included establishing reasonable break time for nursing mothers, an HHS study on urban Medicare-dependent hospitals, restoration of funding for abstinence education, and an excise tax on indoor tanning salons.[130]

Furthermore, the majority found that the district court placed "undue emphasis" on the ACA's lack of a severability clause. The majority noted that in light of the Supreme Court's precedent that "the ultimate determination of severability will rarely turn on the presence or absence of such a clause," Congress's silence in that regard must not raise a presumption against severability.[131] The majority also observed that both Senate and House legislative drafting materials state that as a result of the Supreme Court's presumption of severability, severability clauses are "unnecessary unless they specifically state that all or some portions of a statute should *not* be severed."[132] Thus, the majority found that in light of controlling precedent, and Congress's own drafting materials, the plaintiffs did not meet the heavy burden required to rebut the presumption of severability.

Endnotes

1 *See* Ctrs. for Medicare & Medicaid Servs., National Health Expenditure 2009 Highlights 1 (2011).

2 United States v. Lopez*,* 514 U.S. 549, 552 (1995).

3 *See* U.S. Const. art. I, § 8.

4 W. Lynn Creamery, Inc. v. Healy, 512 U.S. 186, 193 n.9 (1994).

5 *See* Brief For Private Plaintiffs-Appellees National Federation of Independent Business, Kaj Ahlburg, and Mary Brown at 13, Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., (11th Cir. Aug. 12, 2011) (No. 11–11021) [hereinafter "Plaintiff's Brief"].

6 Wickard v. Filburn, 317 U.S. 111 (1942).

7 Gonzales v. Raich, 545 U.S. 1 (2005).

8 United States v. Lopez, 514 U.S. 549 (1994).

9 *Id.* at 559-560.

10 United States v. Morrison, 529 U.S. 598 (2000).

11 *Raich*, 545 U.S. at 17.

12 *See id*. at 22; United States v. Maxwell, 446 F.3d 1210, 1215 (11th Cir. 2006).

13 United States v. Lopez, 514 U.S. 549, 558-59; *accord* Gonzales v. Raich, 545 U.S. 1, 16-17 (2005).

14 *See* 42 U.S.C. § 18901(a)(1)(A), (H).

15 *See* Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 106 (11th Cir. Aug. 12, 2011).

16 *See id.* at 113.

17 *See id.*

18 *See* Plaintiff's Brief, *supra* note 5, at 15.

19 *See id.*

20 *Id.* at 50.

21 *See id.*

22 *See id.*

23 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 130 (11th Cir. Aug. 12, 2011).

24 *Id.*

25 *See id.*

26 *See id.* at 140-41.

27 Plaintiff's Brief, *supra* note 5, at 51.

28 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 140-41 (11th Cir. Aug. 12, 2011).

29 Brief for Appellants at 16, 24, Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., (11th Cir. Aug. 12, 2011) (No. 11–11021) [hereinafter "Government's Brief"].

30 *See* Plaintiff's Brief, *supra* note 5, at 16.

31 United States v. Darby, 312 U.S. 100, 118-120 & n.3 (1941).

32 *See* Plaintiff's Brief, *supra* note 5, at 17.

33 *See id.* at 18.

34 Wickard v. Filburn, 317 U.S. 111, 113-15, 127-29 (1942).

35 *Id.* at 127-29.

36 *See id.*

37 Gonzales v. Raich, 545 U.S. 1, 17-22, 25-32 (2005).

38 *See* Plaintiff's Brief, *supra* note 5, at 19.

39 United States v. Lopez, 514 U.S. 549, 551, 563-64 (1995); United States v. Morrison, 529 U.S. 598, 601-02, 614-15 (2000).

40 *See* Plaintiff's Brief, *supra* note 5, at 20 (citing *Lopez*, 514 U.S. at 611; *Morrison*, 529 U.S. at 561).

41 *Id.*

42 *Id.*

43 *Id.*

44 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 125-26 (11th Cir. Aug. 12, 2011).

45 *Id.*

46 *Id*.

47 Gonzales v. Raich, 545 U.S. 1, 25-26 (2005).

48 *See* Plaintiff's Brief, *supra* note 5, at 25 (citing United States v. Lopez, 514 U.S. 549, 561 (1995)).

49 *Id.* at 25-26.

50 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 109 (11th Cir. Aug. 12, 2011).

51 *Id.* at 111.

52 *Id.*

53 Gonzales v. Raich, 545 U.S. 1, 35 (2005).

54 *See* Plaintiff's Brief, *supra* note 5, at 21.

55 *See id.* at 21.

56 *See id.* at 23.

57 *Id*. at 24-25.

58 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 124 (11th Cir. Aug. 12, 2011).

59 *Id.*

60 *Id*.

61 *See* United States v. Lopez, 514 U.S. 549, 564 (1995).

62 Cong. Budget Office, The Budgetary Treatment of an Individual Mandate to Buy Health Insurance (1994), *available at* http://www.cbo.gov/ftpdocs/48xx/doc4816/doc38.pdf.

63 *See* Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 124-25 (11th Cir. Aug. 12, 2011).

64 *See id.* at 131-32.

65 *Id.* at 125, 131-32.

66 *Id.*

67 *See* Plaintiff's Brief, *supra* note 5, at 27.

68 U.S. Const. art. I, § 8, cl.18.

69 *See* McCulloch v. Maryland, 17 U.S. 316, 421 (1819).

70 Plaintiff's Brief, *supra* note 5, at 30; Printz v. United States, 521 U.S. 898, 923 (1997).

71 Government's Brief, *supra* note 29, at 18, 28, 33.

72 Gonzales v. Raich, 545 U.S. 1, 36-37 (2005).

73 *See* Plaintiff's Brief, *supra* note 5, at 33-34.

74 Government's Brief, *supra* note 29, at 28-32.

75 Plaintiff's Brief, *supra* note 5, at 34; Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 164-65 (11th Cir. Aug. 12, 2011).

76 *Florida ex rel. Attorney Gen.*, Nos. 11-11021 & 11-11067 at 164-65.

77 *See* Plaintiff's Brief, *supra* note 5, at 34.

78 *See id.*

79 *See id.*

80 *See id.*

81 *See id.* at 35-36; United States v. DeWitt, 76 U.S. 41 (1869).

82 *DeWitt*, 76 U.S. at 41.

83 *Id.* at 44.

84 *See* Plaintiff's Brief, *supra* note 5, at 36.

85 *DeWitt*, 76 U.S. at 44.

86 Plaintiff's Brief, *supra* note 5, at 37.

87 *Id.*

88 *See id.*

89 *See id.*

90 *Id.*

91 McCulloch v. Maryland, 17 U.S. 316, 421 (1819).

92 *Id*. at 409.

93 *Id* at 406-09, 420-21.

94 Plaintiff's Brief, *supra* note 5, at 43; *see* United States v. Lopez, 514 U.S. 549, 563 (1994).

95 *See* Wickard v. Filburn, 317 U.S. 111, 113-15, 127-29 (1942).

96 *See* United States v. Lopez, 514 U.S. 564, 583 (1995).

97 *See* Plaintiff's Brief, *supra* note 5, at 45.

98 *See id.* at 46 (citing City of Boerne v. Flores, 521 U.S. 507, 534 (1997)).

99 *Id.* at 47.

100 *Cf.* Trustees of Dartmouth College v. Woodward, 17 U.S. (4. Wheat) 518, 662-63 (1819) (opinion of Washington, J.)

101 Plaintiff's Brief, *supra* note 5, at 49.

102 *See* United States v. Comstock, 130 S. Ct., 1949, 1967 (2010) (Kennedy, J., concurring in the judgment).

103 U.S. Const. art. I, § 8, cl. 1.

104 Government's Brief, *supra* note 29, at 50.

105 *See id*.

106 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 175 (11th Cir. Aug. 12, 2011).

107 *Id.* at 176.

108 *Id.* at 176-77.

109 *Id.* at 181.

110 H.R. 3200, 111th Cong. (2009).

111 S. 1796, 111th Cong. § 1301 (2009).

112 Government's Brief, *supra* note 29, at 54.

113 *Id.*

114 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 184 (11th Cir. Aug. 12, 2011), *quoting* Dep't. of Rev. of Mont. v. Kurth Ranch, 511 U.S. 767, 778 (1994).

115 *See* 26 U.S.C. § 5000A(b)(1).

116 *Florida ex rel. Attorney Gen.*, Nos. 11-11021 & 11-11067 at 186.

117 *See* 26 U.S.C. § 7806(b).

118 *See Florida ex rel. Attorney Gen.*, Nos. 11-11021 & 11-11067 at 187. Indeed, Congress placed numerous civil penalties within the Internal Revenue Code. These civil penalties run the gamut from broadly-applicable penalties, such as filing frivolous tax returns, to highly industry-specific penalties, such as selling or reselling adulterated diesel fuel that violates environmental standards. *See id.*

119 *See* Free Enter. Fund v. Pub. Co. Accounting Oversight Bd., 130 S. Ct. 3138, 3161 (2010).

120 *See* Alaska Airlines, Inc. v. Brock, 480 U.S. 678, 685 (1987).

121 Plaintiff's Brief, *supra* note 5, at 60 (quoting R.R. Ret. Bd. v. Alton R. Co., 295 U.S. 330, 362 (1935)).

122 Government's Brief, *supra* note 29, at 55.

123 *Id.* at 59.

124 *Id.*

125 *Id.* at 56-57.

126 Plaintiff's Brief, *supra* note 5, at 61 (quoting Williams v. Standard Oil Co. of La., 278 U.S. 235, 242 (1929)).

127 Ayott v. Planned Parenthood of N. New Eng., 546 U.S. 320, 329-30 (2006).

128 Alaska Airlines, Inc. v. Brock, 480 U.S. 678, 684 (1987).

129 Florida *ex rel.* Attorney Gen. v. U.S. Dep't of Health & Human Servs., Nos. 11-11021 & 11-11067, slip op. at 192 (11th Cir. Aug. 12, 2011).

130 *Id.*

131 *Id.* 192-93.

132 *Id.* at 193 (citing Office of Legislative Counsel, U.S. Senate, Legislative Drafting Manual § 131 (1997); Office of Legislative Counsel, U.S. House of Representatives, House Legislative Counsel's Manual on Drafting Style § 328 (1995)).

How to Think About Errors, Costs, and Their Allocation

*By Ronald J. Allen**

Note from the Editor:

In December 2010, the Federalist Society heard from a number of federal judges and civil procedure experts about amendments to the Federal Rules of Civil Procedure, including the process that would be undertaken to amend the rules and some proposed amendments that might be offered. Based on the comments and perspectives received, the Federalist Society determined that it could add value to the broader discussion over amending the rules by asking experts to flag issues or perceived problems with the rules as they currently exist, and to identify the range of solutions that are being offered to address these problems. This back-and-forth culminated in four papers, one of which follows. A version of these papers will appear in the Florida Law Review, and they are published here with permission.

There is an ongoing robust debate about the structure of litigation in general, and in particular, about access to the courts. For a considerable period of time, the mantra that the courts should be readily available to all to present claims that their rights have been violated has dominated both academic discourse and perhaps significantly influenced the structure of litigation.[1] The conventional view that the courts should be freely open to all was dealt a blow by the *Iqbal*[2] and *Twombly*[3] decisions, which imposed greater gatekeeping responsibilities on the federal district courts. Predictably these decisions provoked a storm of protest, in large measure because they may indeed make it more difficult for many petitioners to have their petitions considered on the merits.[4] However, whether that result is a social harm or a positive good depends on matters in addition to simply winnowing the field of potential disputants, a point neglected by much of contemporary scholarship in civil procedure. That scholarship has had a laser-like focus on facilitating the bringing of claims, and in doing so makes two serious errors. It neglects that litigation is one small part of a larger social optimization problem, and has a peculiar conception of errors and costs and how they should be allocated. In this brief paper, I provide the analytical background to these assertions.

Primary and litigation behavior are conventionally conceived of as distinct spheres with internal logics of their own, the former articulating rules governing everyday actions, from social interaction to structuring efficient economic behavior and the latter governing that peculiar set of actions involved in litigation. Facilitating appropriate primary behavior is the overriding goal of social organization, and one of its main tools is the substantive law. Litigation behavior is the effort to resolve disputes about inappropriate primary behavior or to reestablish the status quo following disruptions of the social fabric.

Resources devoted to litigation appear to most legal commentators as wasted resources, adding no value to society. Since litigation itself does not produce useful good, litigation should obtain to correct results as efficiently as possible. These aspirations are reflected in Rule 1 of the Federal Rules of Civil Procedure that the rules of civil procedure "should be construed and administered to secure the just, speedy, and inexpensive determination of every action and proceeding," and Rule 2 of the Federal Rules of Evidence: "These rules shall be construed to secure fairness in administration, elimination of unjustifiable expense and delay, promotion of growth and development of the law of evidence to the end that the truth may be ascertained and proceedings justly determined." The principle animating these provisions is access to justice, in particular the principle that even the indigent should be not disadvantaged when the adversary is wealthy. Regrettably, life, as always, is complicated. Costs cannot be eliminated, and thus the most important question is their allocation.

Primary behavior does not produce goods cost-free. There are both waste products produced and the risk of harm to others. Lowering the cost of the production of an item encourages its production, as raising the cost has the opposite effect. Consequently, if the producer can externalize some of its cost (dumping waste in the river or on a neighbor's property), the cost of the good will not reflect its true social cost, which means that there will likely be over-production of the good in question. By contrast, optimal production of social goods is facilitated by having them produced at their true social cost. This is why it is important for the substantive law to align costs with behavior.

Litigation costs are generally believed to be socially perverse, as they act as a tax on productive behavior. To some extent this is true, but a costless legal regime would stimulate the production of its product like any "manufacturer," and the result could be overproduction of this good, as well. Although this may appear counter-intuitive, remember that decisions must be made as to how to dispute—in simple terms whether to sue or negotiate. Everyone comes into contact with numerous instances in which this decision must be made. Perhaps a neighbor plays music too loudly or neglects to dispose of trash correctly. If litigation were costless, rather than negotiate, one could simply sue. The costs of litigation affect the manner in which people relate, and those effects can be beneficial or perverse. The costs of litigation, in short, may counter-intuitively produce social goods through the incentive effects they create for modes of disputing.

The precise policy prescriptions following a deeper understanding of the problem of social cost are ambiguous because they depend in part on the relative values of resolving different kinds of disputes in different ways. It may be sensible to nudge certain kinds of disputes toward formal dispute resolution and others away from it. Maybe commercial disputes differ from family disputes, and maybe discrete commercial

** John Henry Wigmore Professor of Law, Northwestern University; President, Board of Foreign Advisors, Evidence Law and Forensics Sciences Institute, China University of Political Science and Law.*

transactions differ from antitrust actions systematically. Life, in short, is complicated, and one of the tasks for the legal system is sorting out that complexity.

I believe the history of the federal rules of procedure and evidence at least implicitly reflects these analytical points; they were enacted in part to offset what were believed to be distorting aspects of the systems that they replaced.[5] The previous systems were believed to disadvantage plaintiffs by raising their costs much too high. The solution to this was to simplify pleading requirements and allow cases to get on to what was believed to be low-cost discovery, followed by low-cost trials. Discovery costs would be low because the assumption was that knowledge of the typical cases was shared by both parties and thus a substantial investment in discovery would not be required. In addition, both parties would have the incentive to keep costs of discovery to their necessary minimum. It is immediately obvious how these conceptions map onto the previous analytical points. In a world of symmetrical information and low transaction costs, the federal rules perhaps accomplished the goal of facilitating the accurate and efficient resolution of disputes without distorting the underlying substantive law, values that the procedural regime the federal rules replaced did not adequately secure. If the original assumptions about litigation are true, procedural wrangling serves no purpose. Moreover, costs were not and could not be lowered to zero, so there remained reciprocal incentives to avoid litigation through other means of resolving disputes.

Note the historical contingency of the era that adopted the Federal Rules of Civil Procedure. It involved substantive assumptions about the relative positions of plaintiffs and defendants that were empirically true but not logically entailed. Thus, changes in the relative positions of plaintiff and defendant from the pre-Rules situation may justify changes in the procedural context, which could entail among other things a reallocation of costs. Perhaps originally the procedural regime favored defendants and thus subsidized socially wasteful activity; perhaps now in some set of cases it favors plaintiffs with the opposite effect. In such cases, defendants will be deterred from productive activities not by the law but by litigation costs that increase the *in terrorem* value of even meritless suits that put pressure on a defendant to settle and burden otherwise lawful conduct. Potential defendants will engage in litigation avoidance tactics that are likely to be socially wasteful, and they will settle to avoid litigation costs rather than risk liability on the merits. This increases the cost of socially useful activity that cannot be distinguished from socially costly activity at an acceptable price through litigation. The alternative is to buy peace through settlements even though the underlying primary behavior is perfectly acceptable. The effect is a tax on useful behavior.

To generalize, the interactive effects of primary and litigation behavior must be taken into account by the legal system. The effect or consequences of primary behavior on litigation behavior is often noted, but litigation behavior affects primary behavior as well. This means that the regulatory problem is unlikely to be solved by simple slogans such as those concerning access to court. Before addressing how to approach regulating such a complex problem, another issue involving the inadequacy of the conventional understandings of the litigation matrix needs to be addressed. In addition to inadequately considering the relationship between primary and litigation behavior, the conventional conception of an error is inadequate.

The conventional conception of an error is composed of two parts: denying a petitioner access to an adjudication on the merits (through narrowing the court house door)[6] and a belief that Type I (false positive) and Type II (false negative) errors are roughly equivalent and that the procedural goals should be to treat parties roughly equally and to minimize total errors.[7] Although these ways of thinking have been around for a considerable period of time, it is plain that they suffer from serious defects.

First, each time an undeserving litigant imposes costs on an adversary, an error has been made, a point that seems rather remarkably to have been neglected by those who have complained of the recent Supreme Court forays into procedural matters. The image of the federal or any other court system being constantly open and easily accessible for all neglects that a plaintiff walking through the courthouse door imposes costs on a defendant. If the defendant has behaved inappropriately by reference to the substantive law, these are costs the defendant should bear. But, as elaborated above, if the defendant has not behaved inappropriately by reference to that same substantive law—if a plaintiff's claim is unjustified, in other words—the costs imposed on defendants are errors that impose taxes on productive behavior, and thus likely socially perverse. The point is so obvious as to need little further elaboration. An undeserving plaintiff deprives a deserving defendant of its assets, and the best-case scenario is that the deserving defendant passes those costs on to a hapless public. The best-case scenario, in short, is decidedly unappealing. The point, of course, is that the conventional view seems dominated by the belief that there are no wrongful complaints filed, which is ludicrous. More importantly, in an era of asymmetrical costs, where filing a complaint can generate enormous costs on the part of the defendant, the defendant will be consistently in the position discussed above of having to minimize extra costs attached to socially useful behavior and passing whatever costs cannot be avoided on to someone else if possible.

There is a second fundamental error in the conventional thinking about errors. It focuses on just two of the decisions that can be reached at trial—an error for one side or the other—but there are four decisions that can be made at trial, and all have social benefits or costs. In addition to errors, correct decisions can be made. Neglecting correct decisions is peculiar. For example, in civil cases, the error equalization policy is satisfied by making errors in every single case, so long as the base rates of cases that go to trial include roughly the same number of deserving plaintiffs and defendants.

The relationship between the four possible outcomes at trial and procedural regulation is itself more complicated than it appears on the surface. In general, without knowledge of the base rates of deserving parties that go to trial and the relationship between the assessments of fact finders and true states of affairs, there is literally no way to predict the effect of procedural regulation on correct or incorrect decisions. For example, implicit in the conventional discourse is that a finding

that the probability of liability is .8 means that in eight out of ten similar cases, the true facts are consistent with liability. However, there could be any relationship at all between fact finders' findings of probability and true states of affairs. In the set of all cases where fact finders find there to be a .8 probability of liability, it could be true that all cases in that subset are cases where no liability should be found. Similarly, if everyone who goes to trial is guilty or liable, there can be no convictions of the innocent or mistakes against deserving plaintiffs, no matter how low the standard of proof, and vice versa.

The conventional discourse on procedural regulation also assumes a static system, whereas in fact it is dynamic. One aspect of this dynamism is that parties decide which cases to take further into the procedural system, and can adjust their decision in light of changes in the rules. Thus, the simple assumption that changing the burden of pleading or persuasion, or whatever, causes more errors of one kind than another, or any other suggested cause and effect relationship between regulations and outcomes, is obviously not analytically true; it depends on how the system responds to the change.

The combined effect of the neglect of the interactive relationship between primary and litigation behavior and the curious conception of an error is obvious. The result is to obscure that trial decisions are only one part of the output of the legal system. Parties negotiate outcomes in both civil and criminal cases. They do so in the shadow of trials, among other things, but the outcomes in those cases are part of the total social welfare effects of a legal system. In addition, those decisions are made in a dynamic not static environment, which leads to the question how to most effectively regulate such complex processes.

In the abstract, I think the answer is clear. How to translate the abstractions into feasible regulation is another matter. First, the abstract answer is addressed in the quote below from my recent Meador Lecture, which is followed by my further reflections on social optimization of the procedural system:

> [T]he reality of the legal system is not nice, tidy, simple, and static, contexts but instead bubbling cauldrons of messy, complicated, organic, evolutionary processes. The standard tool used to regulate this bubbling mess is rules, and it is the friction between that tool and many of the uses to which it is put that explains in general why fact finding and legal regulation are viewed as so often problematic. This same relationship is explanatory of many legal puzzles, such as, in ascending order of importance, the curious implications of standard legal error analysis, the rules v. standards debate, and the meaning of "law."
>
> The simple concept of a rule as setting necessary or sufficient conditions from which outcomes may be deduced is an example of monotonic logic, in which the addition of postulates or assumptions simply adds to what may be deduced from the previous assumptions. Monotonic logics are powerful tools, as the rise of modern mathematics and the success of many scientific fields demonstrate. They work best when their operant assumptions accurately capture their domains, which means they work quite well, in Hayek's famous dichotomy, in made systems such as games and less well in grown or organic systems, which typifies much of the human condition.[8] A large part of debate over rules and their limits is often implicitly about the complexity of the relevant domain and one's tolerance for mistakes of different kinds. As the number of pertinent variables increases or when some of them are continuous rather than discrete, the deductive problem quickly becomes computationally intractable, even for computers let alone humans. And of course if a new variable pops up that was not previously anticipated, all deductive bets are off, as it were. In either case (computational intractability or failure of imagination), algorithmic approaches that rely on extant rules generate the standard critiques of the indeterminate nature of rules. In reality, it is not that rules are indeterminate but that they are being put to a task for which they are not optimal.[9]
>
> I suggested in that lecture that the central problem of the legal system is similar to the central problem of rationality, which is the taming of complexity. In both cases, simple deductive tools were being put to uses that were suboptimal. That raises the important question what other approaches may be more fruitful. Inspired by a brilliant article by an artificial intelligence researcher, Tim Van Gelder, that I came across many years ago, one possible answer I ventured was that the struggle of rationality to tame complexity may be less like digital computation and more akin to a dynamic regulator, such as the Watt Centrifugal Governor that was a critical part of the industrial revolution.[10] Analogously, legal analysis may need to evolve to deal with the complexities of systems. Van Gelder's example is a metaphor rather than an argument for my purposes, for it provides just the suggestion of possibilities rather than a defined research program, but it is nonetheless interesting.

Static

1. Measure the speed of the flywheel.
2. Compare the actual speed against the desired speed.
3. If there is no discrepancy, return to step 1. Otherwise,
   a. Measure the current steam pressure;
   b. Calculate the desired alteration in steam pressure;
   c. Calculate the necessary throttle valve adjustment.
4. Make the throttle valve adjustment

Return to step 1

Tim Van Gelder, WHAT MIGHT COGNITION BE, IF NOT COMPUTATION, The Journal of Philosophy, Vol. 92, No. 7 (Jul., 1995), p. 348

Here is the problem the dynamic regulator solved. The growth of the textile industry in England depended upon a consistent energy source with very limited variability. The steam engine provided the energy, but its pistons provided episodic bursts of energy rather than a smooth, continuous stream. Fly wheels were helpful, but still not adequate. As Van Gelder pointed out, one potential solution to this problem is computational (see the figure at the bottom of the previous page).

Unfortunately, this computational solution requires a costly person doing it, and it will rarely produce a smooth enough source of energy. James Watt solved this problem by placing movable arms on a spindle at the center of the flywheel, whose motion was transmitted instantaneously to the valve regulating the flow of steam. As the rotation of the fly wheel speeds up, the arms extend, which transmits to the valve and closes it until the proper equilibrium is reached, and vice versa (see the figure at the top of this page).

Dynamic

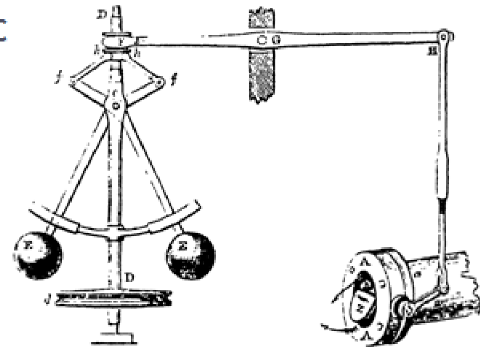

Figure 1[3]

[3] The Watt centrifugal governor for controlling the speed of a steam engine—from J. Farey, *A Treatise on the Steam Engine: Historical, Practical, and Descriptive* (London: Longman, Rees, Orme, Brown, and Green, 1827).

Tim Van Gelder, What Might Cognition Be, If Not Computation, The Journal of Philosophy, Vol. 92, No. 7 (Jul., 1995), p. 349

Regardless whether the centrifugal regulator captures something important about rationality, viewing the legal system with this metaphor in mind may be fruitful. The most dramatic point is that some problems can be solved other than through deductive arguments or simple rules; the contrary belief is a consistent constraint on legal scholarship generally. It is undoubtedly useful to break problems down into smaller parts, and so on, but at the same time that process can be counterproductive, disguising rather than highlighting the nature of the entity under examination. The alternative is to think of the legal system more, perhaps, like fluid dynamics treats the flow of liquids and gases, to embrace, in other words, the messiness of real life rather than abstract it away.

How does this apply in the procedural context? Telling trial judges to behave as centrifugal regulators in order to optimize social productivity is probably not likely to yield satisfactory results. The second-best solution would be to assign the true costs of parties' actions to them. However, it is impossible to determine, practically and maybe theoretically, the "true" costs of litigation behavior. For example, when I ask for discovery, I may be trying to build my case or respond to the opponent's case. I should be responsible for building my case, but responding to my opponent's case perhaps is a cost that he should bear. When a lawyer cross-examines, whose costs are those? If it is pointing out the limits of the adversary's case, he should bear those costs; but if through cross-examination I am building my case, I should bear those costs. How could these different effects be sorted out into the categories of useful for one side or the other? A crude rule—opposing party pays for my costs of cross-examination—leads to potential manipulation. Nor is adopting a British-style loser pays system an obvious solution. Recent empirical work shows both that simple predictions about the effect of a "loser pays" system are likely false (can increase transaction costs), and people do not opt for the English rule in contract negotiations.[11]

Alternatively, the objective could be to structure the process so that the parties have the incentive to properly allocate costs, with when necessary the involvement of the trial judge. That would involve categorical cost allocation, with the possibility of relief from the trial judge. One category probably ripe for such treatment is discovery costs. Discovery costs generally benefit the party asking for the discovery, and also have been a cause of considerable injustice because of their increasingly asymmetric allocation. Plaintiffs simply by filing can impose enormous costs on defendants while bearing virtually none themselves. Note how far from the original conceptions giving rise to the Federal Rules of Civil Procedure the modern condition may be. If each side will have about the same amount of discovery costs, it makes perfect sense to let each side bear their own costs. That is identical to cost shifting, and any resources spent in shifting costs are simply wasted. Asymmetric costs, by contrast, cause skewed cost allocation and provide the opportunity for strategic exploitation. By contrast, placing the costs of discovery provisionally on the person asking for it, but allowing for judicial involvement to make adjustments, may both generally give incentives for the optimal production of information and permit a safety valve in the unusual case.

Although the possibilities are diverse, an example of an "unusual" case would be where there is good reason to believe that an adversary is acting strategically primarily in order to impose costs. In such a case, the "benefit" is to the adversary, and that is who should bear the costs. That would be accomplished by petitioning the trial judge for relief. In making such determinations, the judge's decision will be constructed by the adversarial process, and the parties will have the correct incentives to educate the trial judge. That is not a guarantee of perfection, but it provides some hope for reasonable outcomes. It exploits the advantages of both an initial "bear your own costs" scheme with the apparent inertia of trial courts that do not want to get involved with cost allocation or discovery regulation unless forced to. They would be forced to only when the situation was egregious enough to justify a well-grounded petition for relief.

Endnotes

1 *See* Arthur Miller, *From* Conley *to* Twombly *to* Iqbal*: A Double Play on the Federal Rules of Civil Procedure*, 60 Duke L. Rev. 1 (2010).

2 Ashcroft v. Iqbal, 129 S. Ct. 1937 (2009).

3 Bell Atl. Corp. v. Twombly, 550 U.S. 544 (2007).

4 See, e.g., the symposium on the cases in the Penn. St. Law Rev. ***

5 *See* Ronald J. Allen & Alan E. Guy, *Conley* as a Special Case of *Twombly* and *Iqbal*: Exploring the Intersection of Evidence, Procedure, and the Nature of Rules, *** Penn. St. L. Rev. ***.

6 *See* Miller, *supra* note 1.

7 *See, e.g.*, David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 A.B.F.J. 487.

8 Friedrich A. Hayek, Law, Legislation, Liberty: Rules and Order 35-54 (1973).

9 Ronald J. Allen, Rationality and the Taming of Complexity, ** Ala. L. Rev. ***

10 Tim Van Gelder, *What Might Cognition Be, If Not Computation*, 92 J. Phil. 348 (1995).

11 Kong-Pin Chen & Jue-Shyan Wang, *Fee-Shifting Rules in Litigation with Contingency Fees*, 23 J.L. Econ. & Org. 519 (2007); Keith N. Hylton, *Rule 68, the Modified British Rule, and Civil Litigation Reform*, 1 Mich. L. & Pol. Rev. 73 (1996); Theodore Eisenberg & Geoffrey P. Miller, The English vs. the American Rule on Attorneys Fees: An Empirical Study of Attorney Fee Clauses in Publicly-Held Companies' Contracts, *available at* http://ssrn.com/abstract=1706054.

Twombly in Context: Or Why Federal Rule of Civil Procedure 4(b) Is Unconstitutional

*By E. Donald Elliott**

Note from the Editor:

In December 2010, the Federalist Society heard from a number of federal judges and civil procedure experts about amendments to the Federal Rules of Civil Procedure, including the process that would be undertaken to amend the rules and some proposed amendments that might be offered. Based on the comments and perspectives received, the Federalist Society determined that it could add value to the broader discussion over amending the rules by asking experts to flag issues or perceived problems with the rules as they currently exist, and to identify the range of solutions that are being offered to address these problems. This back-and-forth culminated in four papers, one of which follows. A version of these papers will appear in the Florida Law Review, and they are published here with permission.

> "Under the relaxed pleading standards of the Federal Rules, the idea was not to keep litigants out of court but rather to keep them in."

Justices Stevens and Ginsburg, *dissenting* in *Bell Atlantic Corp v. Twombly*, 550 U.S. 544, 570, 575 (2007).

> "Every reform, however necessary, will by weak minds be carried to an excess, that itself will need reforming."

Samuel Taylor Coleridge (1817)[1]

Viewed from the standpoint of strategic incentives, Rule 4(b) is the foundation of the Federal Rules of Civil Procedure: the state compels someone to appear in court and expend resources to move or answer without regard to the merit of the claims brought. Rule 4(b) is probably unconstitutional, but it is certainly bad policy and creates a distorted incentive structure. *Twombly*[2] is a well-intentioned but misdirected attempt to fix this fundamental problem in the incentives created by the Federal Rules of Civil Procedure, but it focuses in the wrong place. The problem is created pre-service, and that is where it should be fixed.

I. The Fatal Flaw in Rule 4(b).

The fundamental flaw in Rule 4(b) of the Federal Rules of Civil Procedure[3] is delegating governmental power to a private individual to compel another to appear and defend at significant cost and inconvenience without either a preliminary inquiry by a judge that the imposition on the defendant is reasonable, or a reliable practice of assessing costs retroactively if it turns out that the interference with the time and money of the person sued was not reasonable. This delegation of state power to hale[4] people into court without safeguards is particularly anomalous because, as Judge Learned Hand famously reminds us, the greatest calamity that can befall a person, other than sickness or death, is to become involved in a lawsuit.[5]

The unsupervised power that Rule 4(b) delegates to private parties is incongruous. Many similar provisions under which the government summons someone to account for her actions are preceded by a preliminary judicial inquiry appropriate to the circumstances before the state intrudes on a citizen's most fundamental right: the right to be let alone.[6] For example, we require a preliminary judicial inquiry into the bona fides of claims before:

- Summoning someone to answer criminal charges;[7]
- Requiring someone to answer civil claims if brought *in forma pauperis*;[8]
- Requiring someone to answer civil claims that may be brought in retaliation for exercise of their First Amendment rights (a so-called "strategic suit against public participation" (SLAPP));[9]
- Requiring someone to produce documents or testimony in response to a government inquiry;[10]
- Requiring a government official to answer a petition for *habeas corpus*.[11]

A preliminary determination by a judicial official reviewing the grounds for summoning someone to civil court is required by our long-standing American legal tradition dating back to the Founding,[12] as well as by the more recent "due process revolution."[13] Rule 4 of 1938 is an isolated relic of the New Deal penchant for delegating governmental power to private actors[14] that resulted from an unholy compromise between the drafters of the rules and the practicing bar.[15] It is time to fix Rule 4 by requiring a magistrate judge or other judicial official to review the grounds proposed for suit before issuing an order of summons to determine that they are plausible enough to justify haling the persons named in the complaint into court to answer the charges. That is the central insight toward which the Supreme Court was reaching in *Twombly*.[16] Alternatively, if courts do not want to bother to assure themselves of the reasonableness of lawsuits before ordering people to spend time and resources answering them, we should routinely make whole

....

** Professor (adj) of Law, Yale Law School and Partner, Willkie Farr & Gallagher LLP, Washington, DC. e.donald.elliott @yale.edu and delliott@ willkie.com.* *I am grateful to Matthew Christensen, Yale Law School class of 2012, for his excellent research assistance.*

An earlier version of this article was presented at The Federalist Society's Litigation Practice Group Conference "Changing the Federal Rules of Civil Procedure: Has the Time Come?" at the National Press Club, December 9, 2010. http://www.youtube.com/watch?v=ZPsOcRf2zEk&feature=player_embedded#! I also benefitted from comments at workshops at Arizona State Sandra Day O'Connor College of Law and Yale Law School, and particularly the critical comments of my colleague and friend John Langbein. Of course, I alone am responsible for the errors that remain.

those who are sued without sufficient justification by awarding costs retroactively.[17]

The key concept now missing from Rule 4(b) is a requirement for a routine preliminary determination by the judiciary that the grounds proposed for a civil suit are sufficiently plausible that it is reasonable for the government to compel someone to come to court to answer. I call this a "Pre-Service Plausibility Determination" (PSPD) and argue that it is required by our Constitution and tradition as well as by good policy and common sense.

The most basic underpinning of due process of law has long been recognized to be that "The United States cannot . . . interfere with private rights, except for legitimate governmental purposes."[18] But under the current version of Civil Rule 4(b), no attempt whatsoever is made to determine that "a legitimate governmental purpose" is served by requiring someone to appear and answer in a civil case. The criminal rules, in contrast, already routinely require a PSPD, a probable cause determination "by the court" before an order of summons is issued requiring someone to answer charges.[19] In principle, there is little difference between the burdens that the government imposes on someone by issuing an order of summons requiring them to answer private charges in a civil as opposed to a criminal case, although the ultimate consequences may be different.

However, like the fish that does not see the water that surrounds it,[20] most courts and commentators[21] have overlooked Rule 4 and the potential for abuse that it creates. Many casebooks and courses in civil procedure give great emphasis to the general rules of pleading under Rule 8, and to motions to dismiss under Rule 12, but hardly mention Rule 4.[22] Those that do discuss Rule 4 focus almost entirely on the mechanics and territorial limits of service. [23] Scant attention is ever paid to the incentives that Rule 4 creates for nuisance settlements by requiring persons to expend resources to defend without a PSPD that it is reasonable to require them to do so.

The recent initiatives by the Supreme Court in *Twombly*[24] and *Iqbal*[25] to require that lawsuits must be "plausible" have also wrongly focused on the general rules of pleading and motions to dismiss. Many of the problems in the American litigation system have their roots in Rule 4(b), and its state cognates, because that is where the principle is laid down that someone may use government power to impose costs on others regardless of the merit of their claims. This principle creates distorted incentives for rent-seeking[26] and nuisance litigation that should be fixed either by providing a Pre-Service Plausibility Determination by the judiciary before the courts command someone to appear and answer, or a more reliable system for reimbursing persons wrongfully sued for their costs after the fact.[27]

This Article argues that *Twombly* and its progeny are ultimately grounded on values of constitutional dimension, not merely optional constructions of the language of Rule 8(a)(2) requiring "a short and plain statement of the claim showing that the pleader is entitled to relief." The *Twombly* Court put the problem succinctly:

> [S]omething beyond the mere possibility of loss causation must be alleged, lest a plaintiff with "a largely groundless claim" be allowed to "take up the time of a number of other people, with the right to do so representing an *in terrorem* increment of the settlement value." [28]

Motions to dismiss are decided too late to remedy the abuses at which *Twombly* and *Iqbal* were aimed. By the motion to dismiss stage, the persons sued have already been required to expend significant resources, and thus the "*in terrorem* increment of the settlement value" has already occurred, albeit the extent varies depending on how much motions practice and discovery has been allowed. But filing a motion to dismiss does not stay discovery or the costs that it imposes.[29] The Rules stipulate only that motions to dismiss "must be heard and decided before trial unless the court orders a deferral until trial."[30]

A good illustration that even successful motions to dismiss are granted too late to prevent significant harm is *Ward v. Arm & Hammer*.[31] In that 2004 federal district court case, an inmate serving a long sentence in federal prison for selling crack cocaine sued the manufacturer of baking soda for failing to warn on its package that it was illegal to use the product to cut crack cocaine. The federal district judge did eventually grant the defendant's motion to dismiss, pointing out among other things that the inmate had been sentenced in 1995 but had waited until 2003 to file the case. Thus, the claim on its face was barred by the two-year statute of limitations. Despite the tardy and patently implausible nature of the complaint, under the mandatory command of Rule 4(b), the summons and complaint were duly served on the defendants ordering them in the name of the court to answer these patently frivolous charges, thereby compelling them to go the time and expense of retaining counsel to move or answer frivolous charges. The case was filed December 18, 2003, but not dismissed until October 21, 2004, over ten months later. In the meantime, the defendant was required to spend tens of thousands of dollars[32] to defend against a totally bogus claim; if a claim is implausible under *Twombly*, as this one was, the defendants should not be ordered by the federal government to come to court to answer it in the first place. Under the procedures in effect from the Founding until 1938, the defendant in *Ward v. Arm & Hammer* would <u>not</u> have been ordered by the government to answer such patently frivolous claims.[33] But today, because we lack a Pre-Service Plausibility Determination as a regular part of our civil procedure, a federal district court has no mechanism to decline to issue a court order to appear and defend at the request of anyone able to pay the filing fee, no matter how frivolous or stale the charges.[34] Today no government official even reads the complaint before issuing an official court order requiring the persons sued to report to court and to answer civil as opposed to criminal charges. Issuing a governmental order without any attention to its underlying justification is a blueprint that virtually guarantees that government actions will be arbitrary. Moreover, it is an open invitation to "rent-seeking," the private use of governmental power to extort economic value from others.[35]

In addition to coming too late, by focusing on pleadings and motions to dismiss, *Twombly* and *Iqbal* are misdirected because the mechanism of detailed fact pleading is ill-suited to the task of screening claims, as opposed to testing theories for legal sufficiency.[36] No one has yet shown that rules requiring

more detailed fact pleading actually result in anything other than more detailed fact pleading.[37] A mechanism more tailored to the task of screening out cases that should not be served must be developed.[38] In appropriate cases, this preliminary process of screening complaints before service could include a checklist regarding key evidentiary support, as well as a conversation by judges or magistrate judges with the plaintiff's lawyer in which probing questions could be asked about what evidence is available to support certain key allegations or legal theories. I call these inquisitorial inquiries by the judge or magistrate judge before the adversary process begins "Pre-Service Plausibility Determinations." They would be a return to our historical practice, as well as our current practice in many other areas of our law, in which the plaintiff's lawyer appears in court to convince a judge or magistrate judge that the state should summon the persons that he wants to sue to answer his charges. Only in the misguided Rule 4(b) of 1938 did federal law first grant an absolute "right" of a private citizen to commandeer the power of the state to order someone else into federal court.

This strange departure from our usual approach of requiring safeguards against abuse of governmental power is sometimes justified by positing that the person suing is a "rights seeker,"[39] but the person being sued is also a "rights seeker": they just have different visions of their respective rights. The government has an obligation to treat both kinds of "rights seekers" neutrally unless and until it determines that there is a reasonable basis to favor the claims of one over the other.

The bizarre, albeit now familiar, governmental practice of issuing official court orders based solely upon the unverified claims of persons who wish to sue is an open invitation to abuse. It is costly to answer charges, even if they are baseless and are ultimately dismissed, as illustrated by *Ward v. Arm & Hammer*. The problem is exacerbated because of a strong policy in America—completely out of step with most of the rest of the world[40]—that our courts almost never impose costs on losing parties in litigation. Thus, someone can sue, whether or not they have a reasonable basis, and thereby impose costs on others with little or no risk that they will ever have to reimburse those injured by their actions. This is unfair, as well as an open invitation to strike suit arbitrage,[41] and it never should have happened.

The pivotal wrong turn in our law to hand over to private parties with a financial interest in coercing settlements the state's power to summon people to court was wrought in 1938 by what purported to be a merely technical change in an obscure rule governing service of process.[42] In fact, however, the 1938 change in Rule 4 was a fundamental policy shift that quietly gutted statutes that had been passed by the First Congress in 1789 and made permanent by the Second Congress in 1792 to maintain judicial control over the power to issue writs, including the writ of summons to appear in a civil case.[43]

Some might object that returning to the pre-1938 practice of Pre-Service Plausibility Determinations before issuing process is too fundamental a change to consider. But preliminary judicial screening to weed out "junk lawsuits" is no more politically implausible today than judicial screening to weed out "junk science" appeared only a few years ago prior to *Daubert*,[44] while imposing costs retroactively is arguably inconsistent with the American legal culture.[45] At base, the argument against screening cases by imposing costs retroactively is that the *in terrorem* effect of self-executing threats of economic consequences will over-deter some cases that should be brought to the overall detriment of society.[46] A Pre-Service Plausibility Determination by the judiciary, on the other hand, has the advantages that it is not economically punitive and that it is transparent. Judges must make and justify openly a determination that the claims are so implausible that the likely social benefit is not worth the cost, and this ruling is ultimately subject to the safeguard of review on appeal if they deny the right to go forward. A Pre-Service Plausibility Determination is analogous to the existing requirement that a judge, on his or her own motion as well as when requested, must restrict discovery if it appears that the likely benefits are outweighed by the costs,[47] or a decision by the Supreme Court to deny a request to issue a writ of certiorari to decide an issue that someone would like the Court to decide. We all understand why the Supreme Court's resources should not be wasted on cases that are not worth its time, but we have a blind spot when it comes to wasting the time and money of the persons sued in ordinary civil cases.

American judges and magistrate judges routinely screen many other kinds of requests for judicial orders for reasonableness before imposing burdens on private citizens in the name of the judiciary.[48] Reinstating judicial screening to prevent service of "junk complaints" by Pre-Service Plausibility Determinations in all civil cases, not just those brought *in forma pauperis*, would not be judicial activism, but rather a return to our long-standing Anglo-American traditions and the original understanding and practices of the Founders from which we have unwisely deviated.

The root of the incentive structure about which the *Twombly* Court rightly complained is not in Rule 8 regarding pleadings, but in Rule 4 regarding automatic issuance of a court order to appear and defend. That is what requires the person sued to expend resources regardless of the merits of the claim. Contrary to our long-standing traditions, Rule 4 now takes the judge completely out of the loop. The plaintiff's lawyer now controls who is ordered by the court to appear to answer charges in a civil case. Thereby, Rule 4 strikes a fundamentally unfair and unconstitutional imbalance between the rights of persons who wish to sue and the rights of the persons whom someone wishes to sue. The state imposes substantial burdens on the latter based only on the unverified say-so of the former. But both are entitled to equal dignity before the law. The fundamental constitutional norm of state neutrality unless and until a reasonable basis is shown to distinguish among classes of citizens requires that the judiciary must conduct a PSPD, a reasonable inquiry into the *bona fides* of a proposed lawsuit *before* it disrupts someone's right to be left alone. This is particularly true because the chances that anyone will actually be made whole if they are wrongfully sued are vanishingly small in our current system.

This Article makes the case that Civil Rule 4(b) is unconstitutional,[49] but the policy issues are even clearer and more important than the constitutional ones. Even if Rule 4(b) isn't technically unconstitutional, at least not in Holmes' sense of a bloodless prediction of "what the courts will do in fact,"[50] it certainly *should* be unconstitutional. Fundamental norms in

our law underlying several different constitutional provisions all dictate that the court must conduct an appropriate preliminary inquiry into the *bona fides* of claims that one citizen wishes to bring against another to determine that they are reasonably well-founded *before* the state imposes the burden of requiring those whom someone wishes to sue to expend resources to respond. It is important to embed the current debate about *Twombly* and *Iqbal* in this broader context of our constitutional values and traditions, which to date have generally been overlooked.[51]

Rule 4(b) is also badly out of step with what came afterward in constitutional law, as well as with long-standing Anglo-American tradition. In the years since 1938, Rule 4(b)'s approach of empowering creditors to commandeer state power to impose burdens on alleged debtors without appropriate due process protections has been repeatedly repudiated by a long line of Supreme Court cases.[52] Rule 4(b) was drafted before this "due process revolution" of the 1970's recognized that the state has obligations to conduct an inquiry, appropriate to the circumstances, *before* imposing burdens on alleged debtors.[53] However, Rule 4's delegation of unsupervised power to creditors to impose substantial costs on alleged debtors without any quality control by the state has never been seriously re-examined in light of these subsequent constitutional developments.

The term "alleged debtor" or "person someone wants to sue" rather than "defendant" is used advisedly in an attempt to liberate the reader from the social construction—dare I say, "narrative"—prevalent in our culture that "defendants" are always unscrupulous corporations and "plaintiffs" are all sick, impoverished, or injured workers or consumers who are seeking justice.[54] The defining feature of procedure is its potential for reciprocal application. Evil corporations may also sue crusading scientists to coerce their silence.[55] One cannot legitimately design rules of civil procedure by quietly assuming that plaintiffs are always the good guys and defendants are always the bad guys.[56]

Rule 4(b) is indefensible as a matter of public policy, and the public policy issues are even more important and clear-cut than the constitutional legalisms. Rule 4 not only allows unjustified impositions on individuals without a rational justification; at a systemic level, Rule 4 creates economic incentives to over-supply litigation by encouraging the filing of cases that are not cost-justified by either their probability of success, or their potential to develop law or facts in a socially-useful way. The policy and constitutional issues are particularly intense when private parties with a financial stake in the outcome are empowered by the state to impose substantial costs on others that are not justified under existing facts or law, but in the hope that something may turn up. For this narrow category of cases, the "reasonable but speculative" cases, I suggest that not only a preliminary determination of reasonableness by government should be required, but that the lawyer bringing the case should also be required routinely to pay for the costs of a venture from which he or she will profit if successful.[57]

II. Rule 4(b) Unconstitutionally Delegates State Power.

Federal Rule of Civil Procedure Rule 4(b) delegates to any person in the United States (not only attorneys as officers of the court), without any judicial supervision whatsoever, the inherently governmental power to order any other person to stop whatever they are doing and appear in court upon pain of substantial financial penalties. Incredibly, this fearsome state power to summon any person to court to answer to anything upon threat of harsh financial penalties may be exercised merely by filling in three pieces of information on a government form: the plaintiff's (or her attorney's) name and address, and the defendant's name. There is no reference at all in the current Rule 4 to the plausibility or legal sufficiency of the allegations of the complaint, nor is there any regular process for determining whether the grounds for suit are minimally sufficient on either the law or the facts. On the contrary, Rule 4(b) requires that the Clerk of Court "must" issue a summons, an official court

Form 3. Summons.

(Caption – See Form 1.)

To _name the defendant_:

A lawsuit has been filed against you.

Within 20 days after service of this summons on you (not counting the day you received it), you must serve on the plaintiff an answer to the attached complaint or a motion under Rule 12 of the Federal Rules of Civil Procedure. The answer or motion must be served on the plaintiff's attorney,__________, whose address is __________. If you fail to do so, judgment by default will be entered against you for the relief demanded in the complaint. You also must file your answer or motion with the court.

Date _______

____
Clerk of Court

(Court Seal)

order requiring the defendant to appear and answer upon pain of default, if two names and one address are filled in on a printed form that is available in the clerk's office and a minimal filing fee (currently $350[58]) is paid:

> *If the summons is properly completed, the clerk must sign, seal, and issue* it to the plaintiff for service on the defendant.[59]

This is not a drafting glitch. Both the courts and the commentators agree that under current law, issuing the summons is a purely ministerial act by the clerk's office that has no discretion to refuse to issue the summons.[60] The government takes the plaintiff at its word and automatically and without the regular exercise[61] of any government review or discretion issues a court order summoning the person designated by the plaintiff to expend his resources to answer.

As shown in the official appendix of forms, the federal form of summons used in every federal district court today is set out at the bottom of the previous page. The form summons is an official order from the court that states specifically that the defendant "must" answer the complaint. To emphasize its official character, it is signed by the Clerk of Court, a federal official, and bears the official seal of the court.[62] It also makes a stern threat that the government will impose financial sanctions if the recipient disobeys ("judgment by default will be entered against you for the relief demanded in the complaint").

Most American lawyers are so used to this system that it seems natural and they take it for granted. One enlightened exception, however, is Philip K. Howard, who rightly points out that "suing . . . is a use of government power against another free citizen . . . . Being sued is like being indicted for a crime, except that the penalty is money. Today in America, however, we let any self-interested person use that power without any significant check."[63]

Once that undeniable reality is made visible and we see the current Rule 4 system for what it is, we should recoil in horror and recognize that this practice, although so familiar in our legal culture that we may hardly be aware of it, [64] is completely contrary to our constitutional traditions and values. The federal government is *commanding* someone to appear in court[65] based merely on a form being "properly completed" with names and addresses by a private party! That is not the prevailing practice in most state courts, where the service of a summons is not a court order but a private act by the plaintiff's lawyer with no compulsory legal force or effect until a judge later decides whether to grant a default judgment based on the law and the facts.[66] The federal practice of ordering someone to court without any quality control is (1) an unwarranted departure from our historical tradition that the judge controls the basis upon which someone can be haled into court, as well as facially unconstitutional as (2) an unreasonable seizure; (3) a deprivation of private property without due process of law; and most clearly of all, (4) a standardless delegation of inherently governmental power to private individuals. For all of these reasons, Rule 4 should be revised to include a Pre-Service Plausibility Determination by the court prior to service of process, as is explained in the following sections.

A. Rule 4 Deviates from Our Historical Tradition that a Federal Judge Controls the Grounds upon Which Someone May Be Summoned by the Court.

Rule 4 is a sharp departure from our Anglo-American tradition that the court, not private parties, defines regular and predictable grounds upon which someone can be summoned by the government to answer at law.[67]

1. The Original Understanding of the Court Order of Summons.

It was clearly established in both England[68] and the Colonies[69] at the time of the Founding that common law courts had *discretion* to decline to issue a court order to summon the prospective defendant to court based on a PSPD review of the *bona fides* of the proposed lawsuit.

According to a book written by federal district judge Samuel Betts early in the 19th century, the practice in his court prior to the Revolution was for the lawyer for the plaintiff to appear in open court and state her case orally to the judge, who would then decide whether or not to summon the person whom they wished to sue to answer.[70] But even after the oral testing of the request for a writ of summons in open court fell into desuetude, there were still substantial safeguards in the form of a discretionary decision by either a judge or the clerk's office, not the plaintiff or plaintiff's lawyer, that process was warranted:

> *In some cases the judge still considers and determines preliminarily the right of the party to coercive process, and in others subrogates the clerk to that office. And in no instance is the actor permitted to use the process of the court to institute or forward an action at his own discretion*, nor without placing on the files a justificatory document (Rule 2). . . . When no order of the judge is filed, the clerk examines carefully the case made by the libel and the prayer of process, and gives the party such process as his libel will justify. . . . Although the process issues thus by act of court, yet it is taken out by the actor at his risk and responsibility.[71]

The key concept is not whether the preliminary screening before service was oral or written (although I argue later that oral is better, because it allows probing questions). The main point is that a private party was "in no instance" entitled to a summons "at his own discretion" (as is now routinely the case under Rule 4). Rather, as of 1838, either the judge or the clerk "examines carefully" the filing, and only gives the party an order of summons to serve on the proposed defendant if justified.

While Judge Betts was writing a treatise about admiralty, he was a federal district judge sitting in general jurisdiction, and throughout his treatise he routinely notes significant differences between the practices in ordinary civil cases as opposed to admiralty. No such differences are mentioned on this point, which strongly suggests that a similar practice under which judges or the clerk's office exercised discretion before issuing a writ of summons also applied in other civil cases. There is, moreover, no logical reason why the clerk's "duty" only to issue

such process as was justified (as Judge Betts puts it) would be restricted to admiralty cases only.[72]

Similarly, another federal district judge, Alfred Conkling,[73] writing a generation later shortly before the Civil War, also testifies that either the judge or the clerk's office made a substantive review before granting a request for a writ of summons to compel someone to appear and answer. After quoting portions of the passage from Judge Betts also quoted above, that "[w]hen no order of the judge is filed, the clerk examines carefully the case made by the libel and the prayer of process, and gives the party such process as his libel will justify," Judge Conkling goes on to observe:

> Such is the course of proceeding supposed to have been contemplated by the above recited [1844 Supreme Court Admiralty] rule. Except in those cases which require the previous order of the court directing the issue of process, the mere delivery or transmission of the libel to the clerk is all that the rule requires. *But the duty thus imposed upon this officer demands vigilance and intelligence on his part; for he cannot lawfully issue any process, until, by an examination of the libel, he has ascertained that the matter of complaint is in its nature cognizable in a court of admiralty; that the libellant is,* prima facie*, entitled to redress, and that the particular form of process prayed for in the libel is adapted to the case.*[74]

Judge Conkling's statement is even stronger than Judge Betts': he maintains that examining and testing the complaint was not only the prevailing practice, but that it is legally required before the clerk may "lawfully issue" process and therefore that it must be read into the rules. In addition, Judge Conkling makes clear that the review before issuance of the summons was not only for formal defects but must also confirm that the person suing is "*prima facie* entitled to redress."

This already-existing discretion to decline to issue a writ of summons was incorporated by reference into the procedures of the federal courts by the original 1789 Judiciary Act, which created the lower federal courts. Section 14 of the 1789 Judiciary Act authorized the federal courts to issue writs, including writs of summons, but only on terms "agreeable to the principles and usages of law."[75] This was understood to mean "those general principles, and those general usages, which are to be found, not in the legislative acts of any particular state, but in that generally recognised and long established law, (the common law,) which forms the substratum of the laws of every state."[76] In other words, existing English and Colonial practice, including preliminary review of complaints for plausibility before issuance of summons, was specifically incorporated by reference as a condition by the section of the 1789 Judiciary Act that authorized federal courts to issue writs of summons in the first place.

But the First and Second Congresses were not content with this indirect reference to existing understandings and practice. In the 1792 Process Act,[77] they specifically legislated that the federal judiciary must control the issuance of writs, including the writ of summons. On most procedural matters, the early Congresses simply mandated that the federal courts follow existing state procedures, but the Founding Generation thought this one thing important enough to impose it separately regardless of state practice: a federal judge had to "test" (certify), and the clerk had to sign every writ personally, not delegate that right to a plaintiff's lawyer, even though that was already the practice in some state systems.[78] As one of their first acts establishing the federal courts, the First and Second Congresses enacted the statute, set out below, requiring all processes issued by district courts, including writs of summons, to "bear test of the judge."[79] The statutory command of 1792 that the district judges "test" process before issuing writs gradually reified into a formal requirement to include a "teste," an attestation clause witnessing the document.[80] But the statutory requirement that the judge must sign off on process before it issued is still important,[81] just as signing a contract is important to signify that one has adopted its terms. The statutory requirement that

SECOND CONGRESS. Sess. I. Ch. 35, 36. 1792.

Chap. XXXVI.—*An Act for regulating Processes in the Courts of the United States, and providing Compensations for the Officers of the said Courts, and for Jurors and Witnesses.*(*a*)

Section 1. *Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled*, That all writs and processes issuing from the supreme or a circuit court, shall bear test of the chief justice of the supreme court (or if that office shall be vacant) of the associate justice next in precedence; and all writs and processes issuing from a district court, shall bear test of the judge of such court (or if that office shall be vacant) of the clerk thereof, which said writs and processes shall be under the seal of the court from whence they issue, and signed by the clerk thereof. The seals shall be provided at the expense of the United States.

the judge must test and the clerk must issue, seal, and sign signifies that issuing process, including a writ of summons, is a discretionary act by the United States,[82] not a power granted to the plaintiff's lawyer. The federal statute just cited was understood throughout the 19th and early 20th centuries to establish a federal policy to keep issuance of a summons to answer in court "under the immediate supervision and *control* of the court."[83] The clear understanding from the Founding until 1938 was that federal judges and court clerks had a responsibility to satisfy themselves that it was reasonable to order the proposed defendant to court to answer before doing so.[84]

It is true, sadly, however, that some federal judges wanted to avoid what they evidently considered the tedious work of reviewing complaints before service. Without the modern institution of magistrate judges[85] to assist them, the review of complaints to determine whether writs of summons should issue was delegated to the clerk's office and, because assistant court clerks (many of whom are not even lawyers) do not typically have the training or breadth of vision of federal district judges or magistrate judges, review of complaints before service gradually became more technical, formalistic, and less substantive. A late-19th-century treatise from 1895 devotes over sixty-four pages to considering various formal defects in issuing process, and whether they void the court's jurisdiction, or are merely voidable, and hence, subject to correction by amendment.[86]

One of the principal drafters of the Federal Rules of Civil Procedure, Edson Sunderland, notes in a 1909 article that review by the clerk's office was not limited to matters of form or whether proper allegations had been made in the complaint. Sunderland states, "[I]t is within the discretion of the court to allow or refuse the issuance of summons after a long delay."[87] In other words, where it was apparent from a preliminary review of the complaint that a long time had passed between the events forming the basis for suit and the filing of a case, the court in the 19th century and early 20th century had clear discretion to refuse to issue a summons. That now-"superseded" practice[88] compares favorably with the 2004 case of *Ward v. Arm & Hammer*,[89] in which the clerk's office, acting under the edict of "modern" Rule 4(b), mechanically issued a summons requiring a company to spend ten months defending against patently frivolous charges that they failed to warn that using their product to cut crack cocaine was illegal, despite it also being apparent on the face of the complaint that the statute of limitations had long since run.[90]

The practice of pre-service review of complaints described in the treatises is also confirmed by the few pre-1938 appellate decisions that discuss this issue. Historical records of the practices of courts in declining to issue writs of summons are not easily available. There would typically be no written record of these discretionary decisions by judges and clerks except in the rare instances in which a disappointed pleader whose papers had been rejected brought an appeal to a higher court and the appellate court wrote and published an opinion. Several such reported appellate decisions do confirm, however, that the prevailing practice prior to 1938 was for courts to reject requests for summons for a variety of deficiencies, both substantive and formal.

The 1913 decision by the First Circuit in *In re Kinney*[91] is illustrative. There a prominent Pennsylvania inventor, investor, and frequent *pro se* litigant brought a contract suit against a company in federal court in Massachusetts.[92] When his request for a writ of summons was rejected by the clerk of court, he requested the district judge to order the clerk to issue the summons. The district judge upheld the clerk's refusal to issue the summons in an unpublished opinion. The disappointed litigant then attempted to mandamus the district judge in the First Circuit, which also denied his request for a summons, "because the proposed writs 'were not made returnable at the proper return day.'"[93] However, the First Circuit's opinion strongly suggests that there were additional, more substantive reasons as well as formal defects: "It is not necessary for us to examine the reasons given by the judge of the District Court beyond this, because this was a sufficient reason for his refusal."[94]

Another route by which the practice of the clerk's office in declining to issue summonses could come to light was if a disappointed litigant sued the clerk for damages. The 1905 case of *United States ex rel. Kinney v. Bell*[95] illustrates this route. There the same *pro se* litigant referred to above, Robert D. Kinney, sued the clerk of the Circuit Court of the United States for the Eastern District of Pennsylvania, and his sureties, on his bond for refusing to issue a summons in a case that Mr. Kinney desired to bring against several state court judges who had ruled against him. In this instance, the refusal by the clerk's office to issue a summons was clearly because of a substantive defect: lack of federal jurisdiction. The Third Circuit held that Kinney had not suffered any legal damage because the clerk had properly refused to issue a writ of summons because there was no colorable allegation of federal jurisdiction. [96]

2. The "Reforms" of 1938.

The stern insistence in Rule 4 that the clerk "must" issue a court order to appear if a simple form is filled out correctly was no accident; it was an over-reaction by the drafters in 1938 against the then-prevailing practice of assistant clerks rejecting complaints for a variety of formal defects. But it threw out the baby with the bathwater by completely abrogating judicial control over the grounds for haling someone into court.

Charles Clark, then dean of Yale Law School and the principal drafter of the rules, wanted to go even farther. He originally proposed "the New York system," in which private attorneys serve the complaint on prospective defendants and only thereafter file it with the court. [97] Clark thought that this system "works quite satisfactorily," but according to him, the practicing bar objected that it "seemed undignified and over-simple."[98] They called it the "hip pocket system," in which attorneys could sue without filing anything with the court until later when some action was requested of the court.[99] The compromise that ultimately resulted required the complaint to be filed with the court, but removed the court's discretion not to issue the summons. It was a political compromise that combined aspects of the then-prevailing state and federal systems but in an untenable way. Like the then-prevailing federal practice, a lawsuit was initiated by filing a complaint with the federal court and the clerk's office would issue a summons in the form

of a federal court order. But as in the state systems, the clerk's office would issue the summons as a matter of course without any preliminary review by the court before an order to appear was issued.

In an article published a year after the new federal rules were adopted, Dean Clark described the new system succinctly but without any apparent awareness of the problems that this new hybrid had created: "You start a suit by taking your complaint to the clerk, and the clerk issues the summons and the summons and complaint are served by a marshal."[100] There was no attention at all to the incentive structure for strike suits created or the constitutional issues of ordering someone to report to court to answer even implausible charges. In fact, with evident impatience at what he evidently regarded as the unthinking conservatism of the bar over anything with which they were unfamiliar, Clark described the final compromise as "long on dignity" and adopting "the original procedure in the Federal Courts" merely because "that was the more familiar system throughout the country."[101] An outline found in the Clark papers at the Yale University library for a September 1937 speech to the ABA by the Chair of the Rules Committee, former Attorney General William D. Mitchell, tells essentially the same story.[102]

In fact, however, Clark was misstating the reasons that at least some members of the bar wanted to keep the court in the loop for reasons more substantive than mere "dignity." Irvin H. Fathchild, a prominent Chicago attorney, argued in his comments, also found in the Clark Papers at Yale, that requiring a summons to emanate from the court, rather than from a private party, would eliminate a lot of suits "which never would have been filed if the court filing was required as an official step in litigation."[103]

The drafters of Rule 4 were forced by opposition from the bar into a political compromise that amalgamated two different systems into a new hybrid that is constitutionally unsustainable. Under the option originally proposed and preferred by the drafters of the rules, Rule 4 would have incorporated the New York system for initiating a lawsuit. That private system, like that used by most states, is constitutional and does not involve the flaws in the current federal system identified in this Article. Under the New York system (both then and now), state power does NOT become involved in ordering someone to court without assessing the *bona fides* of a proposed lawsuit. Rather, the service of the complaint is a private act performed by an agent of the prospective plaintiff and merely *notifies* the prospective defendant that the action is about to be brought, how to appear to answer it, and what the potential consequences of failing to appear might be. As Clark correctly described it in his 1939 article, under the New York system, "the Court [is] not in the case until some action is asked of it."[104]

That fundamental difference between the federal practice of issuing a writ of summons as a court order, and the practice in many of the states of merely providing a private notice of suit from the plaintiff's attorney, was explained in 1904 in *Leas & McVitty v. Merriman*[105]:

> [T]he word "process," as used in Rev. St. § 911 [the successor to the 1792 Federal Process Act quoted above], *means an order of court*, although it may be issued by the clerk. *The summons in a common-law action, which is, I think, a "process," in the name of the court commands the sheriff or marshal to summon the defendant,* etc. Johnston's Forms, p. 1. The writs of scire facias, fieri facias, habeas corpus, subpoenas for witnesses, subpoenas duces tecum, writs of certiorari, supersedeas, attachments, and of venire facias are all commands or orders of court that something be done. In equity the writ of subpoena, and in criminal cases the bench warrant, command that something be done. Now, the notices under the Code are in no sense commands or orders of court. *They are mere notices that the plaintiff will on some specified rule day file the declaration, or make a motion in court*. . . .
>
> In several of the states a summons in an action may be issued by the plaintiff's attorney. See 19 Am. & Eng. Ency. (1st Ed.) p. 222, notes, and cases cited supra. And in at least the majority of such states it is held that a summons is not a process. *This conclusion is based on the fact that in such states the summons is not issued by the court, and is not an order of court*.

For example, you can find the current New York state form of summons, which is like that in many states, at the top of the next page.[106] Note that the NY summons, unlike the federal one, is not signed by the court, but merely by the attorney for the plaintiff. In addition, the summons is not served by an officer of the state such as a federal marshal, but rather may be served by any person over eighteen who is not a party of the action.[107] Most importantly, the New York form of summons is <u>NOT a court order to appear</u>. Rather, it is merely notice by the plaintiff's attorney that if the person sued fails to appear, the plaintiff intends to apply to the court for a judgment by default against them. Private notice of the general form "I am about to sue you and here's how that works" does not raise the federal constitutional issues of delegating state power to private individuals, or of the state seizing someone without a reasonable basis to do so, or of a deprivation of property by the state without due process of law, which are all raised by the federal system of a court issuing an order to someone to appear without determining that there is a reasonable basis to do so. All of these constitutional issues depend upon state action, which is not present in the typical state system for issuance of summons by the plaintiff's attorney because the court is not involved until later. By contrast, under the federal system, "behind that innocent-looking piece of paper titled 'Summons' stands the full coercive power of the State."[108]

After service in the typical state system, the complaint is "returned" to court, and the lawsuit and the state's involvement begins. If the defendant declines to appear and answer, the state may enter a default judgment against the defendant. But note that entering default judgment is a judicial act, performed by a judge or sometimes a clerk acting under judicial supervision. And most importantly for our purposes, a default judgment may NOT be entered without state scrutiny of the *bona fides* of both the law and the facts.[109] Thus, unlike the federal system created by the 1938 rules, the system of summons by private notice as opposed to court order in effect in New York and many

SUPREME COURT OF THE STATE OF NEW YORK
COUNTY OF [Type in County}

Type in Plaintiffs]	Index No. [type in Index No]
Plaintiff(s),	
-against-	**Summons**
[Type in Defendants]	
Defendant(s).	Date Index No. Purchased:

To the above named Defendant(s)

TYPE IN NAME AND ADDRESS OF DEFENDANTS BEING SUMMONED

You are hereby summoned to answer the complaint in this action and to serve a copy of your answer, or, if the complaint is not served with this summons, to serve a notice of appearance, on the Plaintiff's attorney within 20 days after the service of this summons, exclusive of the day of service (or within 30 days after the service is complete if this summons is not personally delivered to you within the State of New York); and in case of your failure to appear or answer, judgment will be taken against you by default for the relief demanded in the complaint.

The basis of venue is [Basis for choice of venue] which is [fact supporting venue, i.e. address of plaintiff]

Dated: [Type location & Date]

[Type in Law Firm]

by________________________________

[Type in name of signing attorney]

Attorneys for Plaintiff

[Type in name of plaintiff and address of law firm]

other states does NOT involve state power at the initial stage of serving a complaint, but rather only after the complaint is returned to court and the state decides whether or not to enter a default judgment.[110]

It has long been recognized that federal and state practices for commencing a lawsuit are fundamentally different. The leading case is *Dwight v. Merritt*.[111] In that case, a hapless New York lawyer attempted to initiate a lawsuit in federal court using the New York practice for private issuance of summons signed by the attorney rather than the court. The court held, however, that the federal statutory requirement for the court to issue an order of summons was a jurisdictional requirement:

In this case an attempt has been made to commence a suit at common law, in this [federal] court, by serving on the defendant a paper purporting to be a summons, in the form prescribed by the statute of New York, for commencing a civil action. It is signed by the plaintiffs' attorney, but is not under the seal of the court, nor is it signed by the clerk of the court. Section 911 of the Revised Statutes of the United States provides that 'all writs and processes issuing from the courts of the United States shall be under the seal of the court from which they issue, and shall be signed by the clerk thereof.' A summons, or notice to the defendant, for the commencement of a suit, is certainly

process, quite as much as a capias or a subpoena to appear and answer is process. The statute intends that all process shall issue from the court, where such process is to be held to be the action of the court, and that the evidence that it issues from the court and is the action of the court shall be the seal of the court and the signature of the clerk. It is clear that a signature by the plaintiffs' attorney, without a seal, and an issuing from the office of such attorney, cannot be substituted.[112]

For our purposes, the important point is that Rule 4 as it now exists is a sharp departure from the methods of initiating a lawsuit that prevailed historically in *both* the federal and state systems. In the federal system, summons was an official court order to appear, but it was preceded by a preliminary review by a court official to determine that it was justified. In the typical state system, summons was a private action by the plaintiff's lawyer merely to put the prospective defendant on notice.[113] The state did not become involved until later, when the state decided based on the facts and the law whether a default judgment was justified. The new federal system of 1938 in which the government MUST order the defendant to appear regardless of the merit or lack thereof of the plaintiff's claims was neither fish nor fowl.

The federal rules drafters in 1938 certainly must have known that they were abrogating a long tradition by making the issuance of a court order of summons automatic and nondiscretionary in Rule 4.[114] With cryptic understatement, the 1937 Advisory Committee Note to Rule 4(b) recites merely that "USC, Title 28, former § 721 (now § 1691) (Sealing and testing of writs) is substantially continued insofar as it applies to a summons, but its requirements as to *teste* of process are superseded."

One might question how honest a characterization it was for the drafters to say that Rule 4(b) "substantially continued" the provisions of the 1792 statute. Rule 4(b) actually totally abrogated long-standing judicial discretion not to issue a summons and delegated the decision to summon someone to court instead to the person suing (or more practically, the plaintiff's lawyer). This fundamental shift to put state power in private control was not even mentioned in the 1937 Advisory Committee note.

No one seems to have noticed or raised any controversy as this aspect of the new rules made their way through the process. Nor did anyone note the constitutional issues (which, in fairness, did not become prominent until the "due process revolution" of the 1970's). After the rules were adopted, several of the drafters wrote law review articles and delivered speeches describing the significant changes wrought by the new rules. None of these shows any awareness that a fundamental change had been made in the incentive structure for litigation by delegating the unsupervised power to private parties to issue court orders requiring others to appear in court to answer charges.

In a 1939 article provocatively titled, "Fundamental Changes Effected by the New Federal Rules,"[115] Charles E. Clark, then-Dean of Yale Law School and Reporter for the Rules Advisory Committee, began with a telling remark that reveals his general approach: "[P]rocedural rules are but means to an end, means to the enforcement of substantive justice . . . ."[116] Clark goes on to describe many aspects of the then-new rules in detail, but the process for issuing a writ of summons receives only the briefest passing mention: "You start a suit by taking your complaint to the clerk, and the clerk issues the summons and the summons and complaint are served by a marshal."[117] There is no intimation that the phrase "the clerk issues the summons" papered over a significant change or that the new process in any way impinged upon long-standing traditions and constitutional values.

Another academic who also served on the drafting committee, Professor Armistead M. Dobie of the University of Virginia Law School, later a judge on the Fourth Circuit, acknowledged at least obliquely that the court no longer had authority to review the complaint before issuing a summons: "Process, in the form of a summons, is issued by the clerk *as a matter of course* and is served on the defendant together with a copy of the summons."[118] The "as of course" language may have been drawn from former Equity Rule 12 of 1912,[119] which is cited in the 1937 Advisory Committee note to Rule 3.[120] However, it is clear that a subpoena issued under Equity Rule 12 still required "teste" by the district judge under the 1792 Process Act.[121] What was significant about the 1938 changes was the removal of review by the court before issuance of a court order to appear, and thereby the implicit repeal of the 1792 statute by the adoption of Rule 4.

In abolishing review of complaints by judges and the clerk's office prior to service in 1938, the drafters of the federal rules may have felt that they were striking a blow to reduce formalism and legal technicalities and to insure that cases would be decided on their merits. But this "reform" brings to mind Coleridge's admonition quoted in the epigraph that "[e]very reform, however necessary, will by weak minds be carried to an excess, that itself will need reforming."[122] It is one thing to say that the clerk's office should not reject complaints for formal defects that do not affect substantive rights, and quite another to provide that a court order of summons must be issued at the behest of a self-interested private party in every case without any regard to the merits of the claims presented. A more sensible, moderate amendment to Rule 5 in 1993 specifically prohibited the clerk's office from rejecting papers for formal defects.[123] But that moderate approach of overlooking formal defects was not the approach adopted in the 1938 rules, which instead completely eliminated judicial involvement in issuing court orders to appear and defend.

The change in attitude toward "largely groundless claims" (in the words of *Twombly*) before and after the 1938 rule changes is palpable. In a 1933 decision, the Tenth Circuit had proclaimed:

> A court has inherent power to determine whether its process is used for the purpose of vexation or fraud, instead of the single purpose for which it is intended-the adjudication of bona fide controversies. It is the duty of the court to prevent such abuse, and a dismissal of the cause is an appropriate way to discharge that duty.[124]

A few years later, however, under the aegis of new Rule 4, the focus had shifted away from preventing abuse of the court's processes to enforcing the newly-created "right" under Rule 4

for every plaintiff to have her complaint served on whatever persons she wished to sue, regardless of patent lack of merit or an evident purpose to harass. An illustrative case is *Dear v. Rathje*, a 1973 *per curiam* decision by the Seventh Circuit.[125] That case involved a vindictive ex-wife who filed numerous *pro se* cases against her ex-husband and his new wife, as well as picketing his place of employment. The immediate complaint in question was a civil rights claim in federal court against the state court judge who had previously enjoined her from picketing her former husband's place of employment, as well as the lawyer who had represented the husband in that prior case, and the new wife as well. The clerk's office referred Ms. Dear's complaint to a district judge, who after taking judicial notice of a "series" of Ms. Dear's numerous prior cases against her ex-husband and others allegedly acting in concert with him to conspire against her, dismissed the case *sua sponte* prior to service.[126]

The Seventh Circuit reversed, stating:

> It appears that a pattern of practice has developed in the Clerk's office in which summons are not issued [automatically without review] when a *pro se* complaint is filed. . . . We do not need to reach the issue of whether the practice is constitutional since it is possible to decide the appeal on other grounds. The practice here . . . is in clear conflict with Rule 4(a)[now (b)], Fed.R.Civ.P. which imposes a duty on the Clerk to issue the summons "forthwith." [citations omitted] We are not unsympathetic with the plight of the district courts as they face growing numbers of "professional litigants." We also understand the reluctance of its judges to have their courts used as a tool for harassment of public officials and others. But . . . it is not for a United States district court to resolve the problem by cutting off *pro se* litigation at the wellspring.[127]

While the Seventh Circuit may have had a good point about a local rule that singled out *pro se* cases for special review, the rest of its opinion is shallow and one-sided. The opinion only considers the "right" of the plaintiff under the language of Rule 4 to have a summons issued "forthwith," but fails to weigh in the constitutional balance the countervailing privacy interests of those being sued not to be harassed by being required by the state to answer baseless charges.

As a result of the appellate court decision enforcing the terms of Rule 4, Mr. Dear, his lawyer, his new wife, and the state court judge who had ruled in his favor in the prior injunction case were required to endure eighteen more months of litigation, from September 25, 1973 to March 17, 1975, when the district court finally granted summary judgment for all defendants.[128] There is no record of the expense involved, but we do know that two law firms and two lawyers from the Attorney General's office all appeared in the case, and that the ex-husband, Ralph Dear, was eventually forced into default because he lacked the financial resources to answer all of his ex-wife's numerous lawsuits.[129]

In granting summary judgment, the district judge observed that the suit against Mr. Dear's new wife was totally groundless: "This action is nothing more than an aftermath of a domestic controversy between plaintiff and her former husband. Plaintiff made Ralph C. Dear's new wife a defendant but made no allegations against her, merely charging that she was a conspirator."[130] Almost equally groundless was the claim that the attorney had acted under color of state law in representing Mr. Dear, or that the state court judge lacked judicial immunity for rulings made in the ordinary course of business, even if erroneous.[131]

All told, this totally groundless lawsuit by a vindictive ex-wife lasted over three-and-a-half years—from August 14, 1972, when the original complaint was filed, until March 16, 1976, when the Court of Appeals finally summarily affirmed the summary judgment.[132] And this case was merely one of a long "series" that she filed against her ex-husband and anyone unlucky enough to be associated with him. But under the rigid command of Rule 4, the Seventh Circuit held that a federal court was now powerless to prevent its processes from being used as an instrument of abuse by a woman scorned.

Neither Mitchell, Clark, Dobie, nor any of the others involved in drafting the 1938 rules gives any indication of any awareness that they had fundamentally altered the incentive structure of civil litigation, with far-reaching consequences of constitutional dimension. None of the drafters of the Federal Rules of Civil Procedure seems to have paid any attention to the economic *incentives* for the law business that their work was creating.[133] In particular, they seem totally unaware of the "increment of the settlement value" that they were creating in Rule 4 by giving "a plaintiff with a largely groundless claim" the "right . . . to take up the time of a number of other people" (in the words of the Supreme Court in *Twombly*[134]).

Sixty years after Rule 4 was adopted, in the 1998 re-codification of the *United States Code*, portions of the statute passed by the First and Second Congresses relating to court control over issuance of writs of summons were quietly deleted from the statute books on the grounds that they had been "superseded" by the adoption of Rule 4(b) in 1938.[135] The small portion of the original 1792 law about testing of process by the judge before issuance that is still on the books today[136] is a pale shadow on the original passed by the first two Congresses; today the requirement for *teste* of process is formalistic[137] and performed as a ministerial act by the clerk's office without any judicial involvement or discretion; instead, the operative rule that the clerk "must" issue a summons at the behest of a private party is provided by Rule 4(b) rather than the statutes passed by the first two Congresses.

Whether the adoption of the Federal Rules of Civil Procedure by the Supreme Court in 1938 was actually effective to "repeal" the provisions of the 1792 statute requiring test of process by the judge before issuance depends upon whether its provisions are deemed to have provided persons sued with "a substantive right," such as a substantive right to be free from being required to answer implausible lawsuits. Under the Rules Enabling Act, procedural rules may not modify "any substantive right," but laws in conflict with the rules are "of no further force or effect."[138]

In addition, it might be argued that the 1938 rule was ineffective to repeal the 1792 Process Act because it did not go through the constitutional procedures for amending a

statute required by *INS v. Chadha*,[139] the legislative veto case. Several courts and commentators have noted the apparent inconsistency between the Rules Enabling Act provisions for invalidating inconsistent statutes and *Chadha*. In 1988, when the Rules Enabling Act was last reauthorized by Congress, the House questioned including the provision about superseding inconsistent statutes on the grounds that it violated *Chadha*'s requirements for bicameral passage and presentation to the President for a possible veto.[140] The Senate did not concur, however, and the provision was restored.[141] Subsequently, a unanimous panel of the Fifth Circuit Court of Appeals noted but did not reach the issue, stating as an alternative rationale for its statutory construction that the Rules Enabling Act's provision for superseding statutes "approaches a violation" of *Chadha* and "would strain the Constitution's limits on the exercise of the legislative power."[142] The issue has never been squarely decided by the Supreme Court. It was noted in passing, however, in *Clinton v. City of New York*,[143] the line item veto case. There the federal government argued unsuccessfully that the line item veto should be constitutional by analogy to the Rules Enabling Act, but the Supreme Court distinguished the two situations, albeit not altogether persuasively.[144]

It may be that the Supreme Court might not apply the principles of *Chadha* full-force to the repeal of statutes by procedural rules because of the Court's own role in promulgating rules of procedure for the lower federal courts. A full exploration of that interesting issue would require an article at least as long as this one. But for present purposes it is enough to indicate that the issues raised by this Article could be raised in litigation as well as through the rules amendment process. A person summoned to appear in court pursuant to Rule 4 by a summons that had not been tested pre-service for plausibility by a judge could argue that the 1792 Process Act requiring all writs including the writ of summons to bear the "test" of a judge remains in effect, both because it created a "substantive right" not to be required to come to court to answer patently frivolous claims, but also because the purported nullification of this portion of the 1792 Process Act by Rule 4 did not go through the constitutional process required by *Chadha* for amending or repealing a statute.

In the next three sections, I argue that the 1938 change to eliminate Pre-Service Plausibility Determinations, even if superficially legal under the Rules Enabling Act, was unconstitutional as well as unwise.

B. Rule 4(b) Unconstitutionally Seizes Persons and Property.

The Fourth Amendment guarantees, "The right of the people to be secure in their persons, houses, papers, and effects, against *unreasonable* searches and *seizures,* shall not be violated . . . ."[145] This is an important protection for that most fundamental of all rights: the right of privacy; the right to be left alone without intrusion by the government except when reasonably justified.

It is a basic requirement imposed by the Fourth Amendment that, absent exigent circumstances, the government must obtain a search warrant from a neutral judicial officer who independently verifies that there is a substantial basis to proceed with a governmental intrusion.[146] Presently, however, there is no parallel requirement for independent judicial verification of the minimal *bona fides* of a civil claim before someone's time and money are "seized" through a summons to appear and defend in a civil case in federal court.

The Fourth Amendment applies to civil as well as criminal cases. [147] For much the same reasons that we require a showing of either probable cause or reasonable suspicion in criminal cases, we should not require fellow citizens to come to court and answer charges made against them without verifying that there is a reasonable and credible basis for the government to impose this substantial cost and inconvenience of being involved in a lawsuit. And yet the government arbitrarily imposes that very substantial burden and inconvenience on citizens based on the unverified say-so of a single person without any attempt to corroborate his claim or verify his credibility. The government could not obtain a warrant to search your home, a much lesser intrusion on your privacy than making you a defendant in a lawsuit, based solely on the uncorroborated claims of a single informant who had not been shown to be credible. Rather, except in exigent circumstances, an independent judicial official must verify that the facts provide a substantial basis to credit the informant's story.[148] Yet we do not impose a similar minimal requirement of reasonableness before someone's time and property are seized by the government via an order of summons to appear in a civil case.

The Supreme Court has never ruled on the constitutionality of this aspect of Civil Rule 4,[149] and there are no court of appeals cases on point. The case that comes closest is *Williams v. Chai-Hsu Lu*.[150] There, in the context of a §1983 damage action against state process servers, the Eighth Circuit announced the *ipse dixit* that "[a] court's mere acquisition of jurisdiction over a person in a civil case by service of process is not a seizure under the fourth amendment."[151] But that pronouncement was not accompanied by any analysis, nor was the argument made or ruled upon that the state has an obligation to conduct a preliminary inquiry into the *bona fides* of a civil claim before summoning a person sued to answer. Moreover, to the extent that the court offers any analysis at all, it is one-sided and invalid. The issue is not that the "mere acquisition of jurisdiction over a person" in a metaphysical sense constitutes an unreasonable seizure; the New York practice of giving private notice that suit is about to be brought is part of a state-sanction process for acquiring jurisdiction over a person, but it does not involve a governmental order to appear. On the contrary, one is free if he or she so chooses to ignore the case and rely on whether the plaintiff can prove a sufficient prima facie case to obtain a default judgment. Under the state practice, one is not ordered by the government to appear and defend, but rather merely given notice of the right to do so.

The constitutional "seizure" results from the federal government's additional actions in imposing an official requirement to come to court and to expend resources (either in time or money, and usually both) to answer—upon pain of substantial official financial sanctions—without any attempt to verify that there is a reasonable basis for doing so. An official document signed and sealed by the court tells you that you "must" answer and that if you fail to do so, default judgment "will" be entered for the amount sought in the complaint. That

is not a polite invitation, nor merely a notice of actions being taken against you by another private party. Rather, it is an unmistakable command from the state, backed by a threat of official sanctions if you disobey. [152]

Lower court cases such as *Williams v. Chai-Hsu Lu*, *supra,* holding that a civil summons is not a "seizure" in the constitutional sense also ignore the established body of Fourth Amendment law defining "seizures." The conventional legal test for whether a Fourth Amendment "seizure" has occurred is whether, under all the circumstances, a reasonable person would conclude that someone has been deprived of his freedom by the state, or alternatively is free to go on about his business as he chooses.[153] For example, a roadblock designed to halt a car chase has been held to constitute a "seizure," even though the fleeing suspect was not physically placed under arrest.[154] It is the state's intentional restriction of a person's freedom of movement, and not the particular means chosen by the state to accomplish the restriction, that defines a "seizure" in the constitutional sense.[155] As Justice Scalia explained for a unanimous Supreme Court in the 1989 roadblock case, *Brower v. County of Inyo*, a command by an officer of the state that is intended to restrict someone's freedom of movement with which they comply *is* a "seizure" in the constitutional sense:

> It is clear, in other words, that a Fourth Amendment seizure does not occur whenever there is a governmentally caused termination of an individual's freedom of movement (the innocent passerby), nor even whenever there is a governmentally caused and governmentally desired termination of an individual's freedom of movement (the fleeing felon), but only when there is a governmental termination of freedom of movement *through means intentionally applied*. . . . This analysis is reflected by our decision in *Hester v. United States*, 265 U.S. 57, 44 S.Ct. 445, 68 L.Ed. 898 (1924), where an armed revenue agent had pursued the defendant and his accomplice after seeing them obtain containers thought to be filled with "moonshine whisky." During their flight they dropped the containers, which the agent recovered. The defendant sought to suppress testimony concerning the containers' contents as the product of an unlawful seizure. Justice Holmes, speaking for a unanimous Court, concluded: "The defendant's own acts, and those of his associates, disclosed the jug, the jar and the bottle and there was no seizure in the sense of the law when the officers examined the contents of each after they had been abandoned." Id., at 58, 44 S.Ct., at 446. Thus, even though the incriminating containers were unquestionably taken into possession as a result (in the broad sense) of action by the police, the Court held that no seizure had taken place. *It would have been quite different, of course, if the revenue agent had shouted, "Stop and give us those bottles, in the name of the law!" and the defendant and his accomplice had complied. Then the taking of possession would have been not merely the result of government action but the result of the very means (the show of authority) that the government selected, and a Fourth Amendment seizure would have occurred.*[156]

The official summons in a civil case is the direct written equivalent of the Supreme Court's hypothetical revenue agent shouting, "Stop and give us those bottles in the name of the law," which the Supreme Court specifically and unanimously states *is* "a Fourth Amendment seizure." The subsequent cases also stand for the proposition that a command by the authorities is enough to constitute a "seizure" in the constitutional sense if it is followed by compliance even though no physical force is used.[157] "An arrest requires either physical force . . . *or, where that is absent, submission to the assertion of authority*."[158]

The summons in a civil case certainly meets these criteria for "submission to [official] authority."

> Well into the sixteenth century, . . . the writ of *capias ad respondenum* . . . directed the sheriff to arrest defendants and bring them before the court. Today service of process substitutes for bodily seizure, but behind that innocent-looking piece of paper titled "Summons" stands the full coercive power of the State.[159]

No reasonable person reading the standard form summons reproduced above could conclude that the person receiving it was free to go on about his business. The official-looking form bearing an official seal explicitly informs the recipient that untoward legal consequences will be visited upon him or her by the state if he or she does not do exactly as commanded—"default judgment will be entered against you for the relief demanded in the complaint," which is generally a tidy sum designated by the person suing, again without any review for reasonableness by the state. For example, in one case that made the headlines recently, a D.C. administrative law judge sued his local cleaners for $67 million for allegedly losing his pants.[160] It is indefensible for the state to issue an official threat to one of its citizens that it will impose $67 million in financial penalties if he or she fails to show up in court to answer a lawsuit over a lost pair of pants without any attempt to confirm that the sanctions threatened are reasonably proportional to the questions at issue.[161]

The most thoughtful exploration[162] in modern jurisprudence of whether a summons constitutes a constitutionally-protected "seizure" under the Fourth Amendment is Justice Ginsburg's 1994 concurring opinion in *Albright v. Oliver*.[163] There, after an extensive review of the common law precedents and history, Justice Ginsburg squarely concluded that a person "is equally bound to appear and is hence 'seized' for trial, when the state employs the less strong-arm means of a summons in lieu of arrest."[164] That happened to be a summons in a criminal case, but there is no reason why a summons to appear in a civil case would be any less a "seizure" in the constitutional sense than a summons to appear in a criminal case.

It should be noted, however, that Criminal Rules 4(a) and 9(a), unlike their civil counterpart, have long required a preliminary determination of reasonableness before the state issues a summons requiring someone to appear and defend against criminal charges even though no physical arrest is involved.[165] Similarly, no adverse consequences can be visited on an individual for ignoring an IRS summons until a court determines that it is reasonable and enforces it.[166] And the courts will not enforce an administrative subpoena unless it is determined by a neutral magistrate that it is "reasonable" to require a response.[167] In some circumstances, it has even been held that reasonableness requires shifting the costs of

compliance to the inquiring agency.[168] But there is no parallel requirement that the courts must assess the reasonableness of a civil claim before they compel the person sued to report to court to respond. Nor is there presently a requirement or practice to make someone whole after the fact, even if the claim is speculative or turns out to be unfounded.

Civil Rule 4 not only unreasonably "seizes" the person of the defendant by requiring him or her to come to court to defend, either personally or through an attorney, without any prior determination by the state that is reasonable to compel him or her to do so, but it also at least arguably "seizes" the defendant's property[169] by requiring him or her to expend defense costs without any prior attempt by the state to determine that the financial imposition is justified. However, the deprivation of property is probably more properly analyzed under the Due Process Clause, as discussed in the next section.

C. Rule 4(b) Unconstitutionally Deprives Persons Sued of Property Without Due Process of Law.

The Fifth Amendment provides two separate protections against economic impositions by the federal government: the Takings Clause and the Deprivations Clause.[170] The Deprivations Clause, which Rule 4 violates, is broader than the Takings Clause [171] (which Rule 4 generally does not violate[172]), and their purposes are different. The Takings Clause applies if, but only if, property is confiscated by the government for public use. On the other hand, the Deprivations Clause provides that the protections of procedural and substantive due process must apply before anyone may be "deprived" of use or control of their property by the government, whether or not it is taken for public use by the state. The core purpose behind the Deprivations Clause is to insure that a "legitimate governmental purpose" justifies an imposition on citizens causing them trouble and expense.[173] Rule 4 is deficient in that the government makes no attempt whatsoever to verify that there is a legitimate reason to order someone to answer in court before doing depriving them of property by requiring them to expend resources to answer charges in court.

The key element that triggers the Deprivations Clause is that someone is denied possession or use of money or another recognized form of property[174] by the state. Thus, the Fifth Amendment guarantee against deprivations of property without due process of law has repeatedly been held to apply to situations in which the state imposes costs or requires payments to third parties: for example, a unilateral EPA order requiring a company to spend money to clean up a Superfund site unquestionably constitutes a "deprivation" of property, although four circuits have now held that the procedure does not violate due process because it is reasonable and provides for a pre-deprivation judicial hearing.[175] Similarly, by requiring someone who is sued to expend resources to answer charges in court, the state is clearly imposing costs and thereby "depriving" the person sued of property so as to trigger due process protections. This is true whether they hire counsel, or merely pay for the transportation costs and paper to represent themselves *pro se* (although of course the magnitude is greater when counsel is employed). The costs imposed by litigation are not trivial. According to the Federal Judicial Center, the average cost of a case in 2009 was $15,000,[176] although, unsurprisingly, the costs varied in proportion to a number of variables.[177]

Deprivations of property are not necessarily illegal; they merely must comply with due process, which means that they must be substantively reasonable and accompanied by procedures appropriate to the circumstances. What is unusual about current Rule 4, however, is that the government forswears any inquiry into the reasonableness of its actions before it imposes substantial economic costs on the putative defendants. This unthinking imposition of economic costs on the persons sued without providing reasonably available procedures to assess the reasonableness of the economic harm imposed by the state violates the Deprivations Clause.

Rule 4 sticks out like a sore thumb because it provides no pre-deprivation process whatsoever and rarely is a person who is wrongly sued reimbursed retroactively for the expenses incurred. Rule 4 also arguably offends the equal protection component of the Due Process Clause by automatically taking the word of one group of citizens as the basis for imposing burdens on another group of citizens.[178]

What process is "due" is of course dependent upon the circumstances.[179] At the time that the writ of summons developed in the 13th century, when few people could read or write, much less communicate by email and telephone, commanding someone to appear before the King personally in order to answer charges may have been the most efficient way to determine whether there was a reasonable basis for the claims.[180] But that is no longer the case today, and due process requires a system that is tailored to what is reasonably available.

In a long line of cases beginning in *Sniadach v. Family Finance Corp. of Bay View*,[181] and extended in *Fuentes v. Shevin*,[182] the Supreme Court has held that the Due Process Clause constrains the use of other long-established common law writs and remedies so that not even a temporary deprivation of property by the state is allowed without a prior inquiry appropriate to the circumstances.[183] The 1975 due process decision in *North Georgia Finishing, Inc. v. Di-Chem*, *Inc.*[184] is particularly interesting for our purposes. There, in dicta, the Court suggested that the combination of a "detailed affidavit," a determination of facial validity by a "neutral magistrate," and a bond to pay costs for property wrongfully seized *pendente lite,* could be sufficient to satisfy due process.[185]

The suggestion in *Di-Chem* that a detailed affidavit, review by a neutral judicial officer, and a bond or other procedure to compensate the victim for wrongful deprivations would be sufficient to comply with due process is also consistent with the decision in *Mitchell v. W. T. Grant Co.*[186] That case upheld a Louisiana statute permitting a secured creditor with a pre-existing lien to sequester property pre-judgment. The *Mitchell* Court emphasized the lien-holder's pre-existing interest in preventing dissipation of the previously-encumbered property, but also the requirement of a detailed affidavit from which a judge could determine a clear entitlement to the writ, plus the availability of an immediate post-deprivation hearing with the option for damages.[187]

Civil Rule 4, however, provides none of these three constitutionally-required elements that have been applied to constrain potential abuse of other common law writs (a detailed

affidavit verifying the claim, a neutral judicial evaluation before imposing the burden, and a process for compensating the victim if the deprivation turns out to be invalid). And, unlike *Mitchell v. W. T. Grant*, the plaintiff in an ordinary civil case has no pre-existing lien whatsoever on the defendant's assets. Nor does the theoretical possibility of a suit after the fact for abuse of process or malicious prosecution remedy the defect. These suits require an additional showing of an improper purpose and malice or subjective intent. Merely showing that the suit was objectively unfounded and unreasonable is insufficient.[188] Unlike the temporary deprivations of property by common law writs found unconstitutional in the *Fuentes v. Shevin* line of cases, the deprivation of property worked by the writ of summons is almost always permanent and irreparable because under the American Rule, costs are not assessed against losing parties in litigation. As a result, the state has a particularly strong obligation to provide pre-deprivation procedures.

This line of due process cases from the 1970's was reiterated and clarified in 1991 in *Connecticut v. Doehr*,[189] in which a unanimous U.S. Supreme Court struck down a Connecticut statute that had authorized pre-judgment attachment of real estate as security for a pending civil suit based on an *ex parte* judicial determination of probable cause. The Connecticut pre-judgment attachment procedure imposed a much lesser burden than Civil Rule 4 in that pre-judgment attachment typically imposed no actual financial costs on the defendant. Instead, it merely consisted of entering a *lis pendens* on the land records, thereby notifying other creditors of the pending unrelated claim and establishing the priority of the potential judgment creditor in the case under suit.[190] Nonetheless, a unanimous U.S. Supreme Court declared this procedure unconstitutional because, without prior notice and hearing, or exigent circumstances and a requirement to post a bond to make the owner whole afterwards, the state deprived someone of private property without due process.[191]

For our purposes it is particularly relevant that in *Doehr*, Connecticut tried unsuccessfully to defend its statute by analogy to the Federal Rules of Civil Procedure, arguing "that the statute requires something akin to the plaintiff stating a claim with sufficient facts to survive a motion to dismiss."[192] The U.S. Supreme Court unanimously rejected Connecticut's argument that the plaintiff's unverified say-so in enough detail to survive a motion to dismiss was sufficient to justify even the temporary deprivation of control of real property resulting from a pre-judgment attachment. The *Doehr* Court applied the modern due process framework that had developed since *Snaidach* and its progeny for balancing competing private and public interests against the risk of error under *Mathews v. Eldridge*.[193] The *Doehr* Court explained:

> [T]he statute presents too great a risk of erroneous deprivation under any of these interpretations. . . . *Permitting a court to authorize attachment merely because the plaintiff believes the defendant is liable, or because the plaintiff can make out a facially valid complaint, would permit the deprivation of the defendant's property when the claim would fail to convince a jury, when it rested on factual allegations that were sufficient to state a cause of action but which the defendant would dispute* . . . . The potential for unwarranted attachment in these situations is self-evident and too great to satisfy the requirements of due process absent any countervailing consideration.
>
> . . . *It is self-evident that the judge could make no realistic assessment concerning the likelihood of an action's success based upon these one-sided, self-serving, and conclusory submissions*.[194]

Applying this same analysis to the much more substantial deprivation of property worked by Civil Rule 4—the costs of defense imposed on every person sued, "merely because the plaintiff believes the defendant is liable"—should lead to exactly the same result. Moreover, *Doehr* stands for the proposition that more is required than "one-sided, self-serving, and conclusory submissions," such as those in a typical complaint.

Significantly, this line of due process cases was decided a generation *after* Civil Rule 4 was written, and as far as I have been able to determine, the provisions of Rule 4 have never been seriously reconsidered in light of them. It is not apparent why a requirement to spend money to answer charges in a civil case based on the unverified say-so of a would-be creditor should be any different than the pre-judgment attachment of real property based on the unverified say-so of a would-be creditor that was struck down as unconstitutional in *Connecticut v. Doehr, supra*. Connecticut's pre-judgment attachment statute contained substantially more protection against arbitrariness [195] than are currently provided by Civil Rule 4.

It is also interesting that four Justices in *Doehr* went on to opine that when exigent circumstances do not permit a hearing, a bond to reimburse a person wrongfully deprived of his property might be constitutionally required. [196] This strongly suggests that so-called "cost shifting"[197] may be constitutionally required in situations where courts allow plaintiffs to conduct "fishing expedition" discovery to determine whether they have a valid cause of action, but the plaintiff is unsuccessful in doing so.[198] The other five Justices did not disagree; they simply felt that it was unnecessary to address that issue in the case before them.

For the same reason that the Supreme Court has held that other common law writs and remedies such as replevin and garnishment must be disciplined by the Due Process Clause, so too the writ of summons should be issued only after the state verifies that a deprivation of the proposed defendant's property is justified by the plausibility of the plaintiffs' claims.

D. Rule 4(b) Unconstitutionally Delegates Governmental Power to Private Parties.

The decision to order someone to come to court to answer charges is undeniably an exercise of state power, as pointed out by Philip Howard above.[199] Rule 4, however, makes the issuance of a federal civil summons a ministerial act by the court clerk.[200] It thereby delegates an important exercise of state power to private individuals in violation of the constitutional provision that the judicial power is vested in the courts. Worse yet, there are no standards that private individuals must satisfy in order to exercise this fundamental attribute of state power (beyond properly filling out the form of summons, which is a

patently insufficient check on this delegation of state power). This violates the fundamental constitutional principle that government power may not be delegated to private individuals without appropriate standards to guide its exercise.[201] Far less serious exercises of governmental power than issuing a court order to participate in a lawsuit have been held to violate the principle against delegating government power to private individuals. For example, statute statutes that require the consent of adjoining property owners to a change in zoning classification have been held unconstitutional because they delegate governmental powers to private individuals.[202]

The issue of standardless delegation of governmental power to private individuals is particularly objectionable because the private actors[203] exercising this power, plaintiff's lawyers, have a financial stake in the outcome. If a judge made these same decisions of whom to order to court, but had a financial interest in nuisance settlements to avoid litigation costs, we would instantly recognize a violation of due process.[204] But we allow plaintiffs' lawyers with contingent fee arrangements who will share in the proceeds of any nuisance settlements to require court orders to be issued to any person they choose without any control by the court to insure that the order to appear and defend has a reasonable basis in law and in fact.

This problem of a delegation of state power to those with a financial interest in the outcome is particularly intense when plaintiffs' lawyers are empowered by the state to bring cases that do not currently have a valid basis in law or in fact. The rules properly allow them to bring such cases in the hope that they will later be able to develop a reasonable basis for the claim either through facts unearthed in discovery,[205] or "by a non-frivolous argument for extending, modifying, or reversing existing law or for establishing new law."[206] Some of these speculative cases are reasonable in terms of the benefits they confer on society and probably should be allowed.[207]

It does not follow automatically, however, that the person sued should subsidize the investigation into whether a wrong has been committed. In such "reasonable but speculative" cases, it should be routine for the plaintiff's lawyer to pay the costs that his or her speculation in litigation futures imposes on the persons sued.[208] Normally in a market economy those who make the decision to invest in an economic opportunity are required to pay the costs of the social resources consumed by their endeavor. This is thought to create a self-policing system in which those who are in the best position to determine whether an opportunity is worth pursuing can balance both costs and benefits of the activity in which they choose to engage. The litigation business is unusual, however, in that a plaintiff's lawyer may externalize a substantial portion of the costs of the economic venture that he or she initiates onto the defendant, but the attorney and his client obtain all of the benefits if the venture is successful. In other contexts, this incentive structure, in which one economic actor gets the profits but another assumes the risks, has been criticized by economists for creating runaway speculation.[209]

The present system, however, unconstitutionally delegates all of these decisions to the plaintiff's lawyer without any standards, supervision, or review by the state and merely with the toothless threat of sanctions under Rule 11 if the case turns out to be unreasonable. This is another, more subtle version of the problem of standardless delegations of government power to private individuals discussed above. The policy judgment that plaintiffs should sometimes be allowed to bring cases that are not well-founded in existing law and/or in the facts currently in the plaintiffs' possession does not mean that decision should be delegated to private individuals who have a financial interest in the outcome.[210] But because this fundamentally *judicial* decision to allow a case to go forward despite the absence of sufficient law or evidence to support it has been delegated to private parties to be made *sub silentio*, we currently seem to have no problem with allowing plaintiffs' lawyers with a personal financial stake in the outcome routinely to summon and impose costs on defendants against whom they currently lack sufficient evidence, thereby creating settlement value that inures to the personal benefit of the plaintiff's lawyer. Because this occurs "out of sight, out of mind," judges have no idea how common it is for defendants to be extorted by power delegated by the state into making payments in cases in which they are not legitimately involved.[211]

The best that can be said for these "something may turn up" or "fishing expedition" cases is that they may be filed in good faith, but speculatively, by private parties with a financial stake in the outcome. A more sinister explanation is that experienced plaintiffs' lawyers know that the many people that they are suing may pay nuisance value. They should not be condemned for responding rationally to the lucrative economic opportunities that the ethical and procedural rules currently permit. Traditionally called "strike suits,"[212] such cases are filed not because of their probability of success on the merits but because of the settlement value that they create by imposing defense costs on those who are sued. One can debate the frequency with which such cases occur, and the size of the dead-weight loss that they impose on the economy, but one cannot logically deny that they exist. In a famous article in 1979, Landes and Posner showed formally that even cases with little or no prospect of success do create settlement value in proportion to the costs of litigation. [213] Empirical data are not very good on how large the dead-weight loss to the economy is from such cases. One empirical study of employment discrimination cases concluded that it makes economic sense for an employer to pay at least $4,000 per claim regardless of merit simply to avoid costs of defense.[214]

A strike suit is an "arbitrage" pure and simple: economic value is manufactured not by creating anything socially useful, but simply by doing a transaction over and over where there is a discontinuity between its payoffs and its expected costs.[215] The discontinuity between expected costs and benefits is in turn a function of the endemic judicial reluctance to "shift" costs of consuming resources in litigation from where they fall to those who cause them.[216] Judges should not confuse costs with penalties. There is nothing punitive about requiring an economic actor to pay for resources that are consumed in an activity that they undertake to make a profit.[217] On the contrary, the philosophy behind a market economy is that resources will be used most efficiently if those who decide to consume them pay the marginal costs of production.[218] For the same reasons that electricity will be wasted and over-consumed if government

requires it to be supplied at a price below the marginal cost to make it, litigation will be over-supplied, wasting societal resources, if those who initiate litigation pay only a small fraction of its cost.

The root of the judicial reluctance to impose the costs of litigation on those who are in the best position to determine whether the expenditure of resources is justified is in turn embedded in Rule 4 and the perverse incentives that it creates: judges are required by law and custom to presume that every case filed in court is valid until shown otherwise, and the "showing otherwise" is expensive.

Although this constitutional defect in Rule 4 is perhaps the most clear-cut, it is not desirable to fix Rule 4 by developing more constraining standards for when private parties may exercise the state power to summon. That was the function that the "forms of action" performed until they were abolished by the Field Code in New York in 1848, and at the federal level by the Federal Rules of Civil Procedure in 1938. By delimiting acceptable categories for suit, the state historically constrained the basis on which one party could hale another into court. It is not desirable to bring back the rigidity of the "forms of action." However, without the forms of action to constrain private discretion regarding the basis for suit, the state must now make a PSPD, a preliminary inquiry into whether the plaintiff's claims are sufficiently plausible on both legal and factual grounds that the state may reasonably require the person sued to answer them or routinely award costs afterward.

Courts are already required by statute to do this for civil claims brought *in forma pauperis*. The federal *in forma pauperis* statute provides:

> [T]he court shall dismiss the case at any time if the court determines that—
>
> . . .
>
> (B) the action or appeal—
>
> (i) is frivolous or malicious;
>
> (ii) fails to state a claim on which relief may be granted; or
>
> (iii) seeks monetary relief against a defendant who is immune from such relief.[219]

In a 1989 decision, *Neitzke v. Williams*,[220] a unanimous Supreme Court, speaking through Justice Thurgood Marshall, explained the rationale for differing treatment between *in forma pauperis* cases and those brought by paying customers as follows:

> Congress recognized, however, that a litigant whose filing fees and court costs are assumed by the public, unlike a paying litigant, lacks an economic incentive to refrain from filing frivolous, malicious, or repetitive lawsuits. To prevent such abusive or captious litigation, 1915(d) [now (e)] authorizes federal courts to dismiss a claim filed *in forma pauperis* ". . . . if satisfied that the action is frivolous or malicious." Dismissals on these grounds are often made *sua sponte* prior to the issuance of process, so as to spare prospective defendants the inconvenience and expense of answering such complaints. See *Franklin v. Murphy*, 745 F.2d 1221, 1226 (CA9 1984).[221]

That was, however, before *Iqbal* and *Twombly*. The *Neitzke v. Williams* Court cited with approval *Conley v. Gibson*'s[222] very liberal pleading standard that no actionable set of facts could be proven under the allegations.[223] This standard was later specifically disavowed in *Twombly*.[224] The main concern of the Court in *Neitzke v. Williams* seems to have been to make sure that poor people were given just as much leeway as rich ones to file cases even if they ultimately proved unfounded.[225] This laudable goal of equality between rich and poor litigants was achieved by harmonizing in the wrong direction. Paying customers should be subject to the same *sua sponte* review for frivolousness before service of process as are their fellow citizens who are indigent already are.

The *Neitzke v. Williams* Court noted the issue whether *sua sponte* dismissals under Rule 12(b)(6), as opposed under the *in forma pauperis* statute, are permissible in a footnote, but did not answer it.[226] The Court did state in dicta, however, that

> [a] patently insubstantial complaint may be dismissed, for example, for want of subject-matter jurisdiction under Federal Rule of Civil Procedure 12(b)(1). See, e. g., *Hagans v. Lavine*, 415 U.S. 528, 536 -537 (1974) (federal courts lack power to entertain claims that are "so attenuated and unsubstantial as to be absolutely devoid of merit'") (citation omitted); *Bell v. Hood*, 327 U.S. 678, 682 -683 (1946).[227]

The question left open by the Supreme Court in *Neitzke v. Williams* regarding the authority of a federal court to dismiss *sua sponte* before service of process in an ordinary case under Rule 12(b)(6), as opposed to under 28 U.S.C. §1915(e) in an *in forma pauperis* case, was answered in the affirmative by the D.C. Circuit in *Baker v. Director, United States Parole Commission*.[228] That *per curiam* decision is of particular interest because the panel included then-Circuit Judges Ruth Bader Ginsburg and Clarence Thomas, arguably the most liberal and most conservative Justices of the current Supreme Court. They both joined Judge Lawrence Silberman in a *per curiam* opinion holding that a *sua sponte* dismissal prior to service of process was proper under Rule 12(b)(6), even in a case not brought under the *in forma pauperis* statute, "where the plaintiff has not advanced a shred of a valid claim."[229] Other circuits hold to the contrary,[230] however, and there is a clear circuit spit that will eventually have to be resolved by the Supreme Court. A Supreme Court case addressing that circuit split might be a good occasion to create the Pre-Service Plausibility Determination process advocated by the Article.

Even if the power asserted by the D.C. Circuit in *Baker* to dismiss an occasional case before service *sua sponte* were to be recognized more generally, that would not obviate the need for a change to the language of Rule 4 as proposed below.[231] The principal drafter of the Federal Rules of Civil Procedure, Charles Clark, sagely pointed out long ago that the rules should not only grant judicial power, but especially when they aspire to change judicial behavior, they must also *explain* how and why that power is to be used.[232] In this instance, the practice that the clerk's office issues the summons automatically without any

preliminary determination by the court that it is reasonable to require the person sued to answer is so deeply embedded in our current practice that a change in rule language is required.

III. The Government Must Verify the Plausibility of Civil Claims Before It Orders Persons to Answer Them.

Perhaps the anomalies described above would be tolerable if they were unavoidable, but there is a simple solution, which is routinely followed in many other areas of our law: the PSPD. ***Before summoning someone to spend a substantial amount of time and money defending a lawsuit, a court official should make an inquiry appropriate under the circumstances to verify that there is a plausible basis for the claim that is sufficient in law and fact for it to be reasonable for the state to require the defendant to answer.*** This does not mean that plaintiffs must show that they are going to win their lawsuit. It simply means that the government has an obligation to satisfy itself that there is a sufficiently reasonable basis for the suit so that the state is not complicit in fraud or extortion, or is not itself acting arbitrarily by ordering the defendant to appear and defend. This minimal threshold requirement is not satisfied merely because someone fills in their name and address and the name of the person that they want to sue on a government form.

Rule 4(b) currently reads as follows:

> (b) **Issuance.** On or after filing the complaint, the plaintiff may present a summons to the clerk for signature and seal. If the summons is properly completed, the clerk must sign, seal, and issue it to the plaintiff for service on the defendant. A summons—or a copy of a summons that is addressed to multiple defendants—must be issued for each defendant to be served.

For the reasons described above, it should be amended to read as follows:

> (b) **Issuance.** On or after filing the complaint, the plaintiff may present a summons to the clerk for signature and seal. If the summons is properly completed, **and a magistrate judge or district judge determines from review of the complaint and other appropriate inquiries that it is reasonable to summon one or more of the proposed defendants to answer,** the clerk must sign, seal, and issue it to the plaintiff for service on th**at** defendant. A summons—or a copy of a summons that is addressed to multiple defendants—must be issued for each defendant to be served.

The concept of minimal governmental inquiry before imposing a substantial burden on a citizen is not unknown to our law. In fact, we honor that principle in every other area of law that I can think of—except when summoning someone to defend a civil lawsuit under Rule 4 and its state equivalents.

The system of civil procedure creates a series of "hurdles" of increasing height that are tailored to the appropriateness of moving to the next stage:

> (1) At the Rule 4 stage, the proper question is a very modest one: whether the case appears to be sufficiently plausible that it is reasonable for the state to require the defendant to appear and respond to the complaint;
>
> (2) At the Rule 12 stage, the proper question is a different one: whether the plaintiff has stated a legally-cognizable claim such that it is reasonable to subject the defendant to the costs and intrusion of discovery.
>
> (3) At the Rule 56 stage, the proper question is whether a sufficient dispute of material fact exists after discovery that the case should be heard by the trier of fact.

In *Iqbal* and *Twombly* the Supreme Court correctly perceived the problem of imposing costs on those sued without verifying that there is sufficient merit to the claim to justify doing so, but it located the solution in the wrong place, using the wrong mechanism. The proper function of the complaint in the modern procedural system is to state the plaintiff's legal theories with sufficient particularity that their legal sufficiency can be tested via a motion to dismiss. It is impossible in any system to maximize two or more variables simultaneously.[233] *Ceteris paribus* procedural devices work better when they are not asked to perform multiple, inconsistent functions. While there should be a modest hurdle before the defendant is haled into court by the state, it does not necessarily follow that we should return to detailed fact-pleading in the complaint. There are many well-known deficiencies in a system that requires that the plaintiff be in possession of all the facts necessary to take a case to trial as a pre-condition to bringing a claim.

Rather than re-invent fact pleading, with all of its well-known drawbacks and inefficiencies, we should adapt new procedural devices as part of Rule 4. These procedures should be properly adapted to the purpose of determining whether it is reasonable for the state to summon the persons identified by the plaintiff and put them to the burden and expense of defending a particular claim. That would consist of a two-pronged inquiry: (1) whether a claim is sufficiently plausible based on the available facts and existing law that it is reasonable for the state to compel the persons that the plaintiff wishes to sue to incur the costs and inconvenience of appearing in court; or if not, (2) whether the plaintiff is sufficiently likely to develop the necessary facts or law at a later date.[234] Some "speculative but reasonable" cases should be brought for their broader social utility even though the available facts and/or law do not support the claim. But it does not follow that (1) the power to bring claims that are not currently justified by the available facts or the law should be delegated to private self-interested individuals without any standards or review by the state; and it also does not follow (2) that the costs of the resources consumed in a speculative effort to develop facts or law should be subsidized by the persons sued regardless of how the economic venture ultimately turns out.

A judicial official such as a magistrate judge should engage in a preliminary examination of a lawsuit before summoning the defendant to respond in order to determine that the lawsuit is plausible enough that it is reasonable for the state to put the defendant(s) to the time and expense of responding. In many instances, this could be done simply by reviewing the complaint, particularly if it pleads facts with sufficient specificity and is verified under oath or attaches key items of evidence, such as the contract or promissory note upon which suit is based. Moreover, complaints could identify key pieces of evidence that

the plaintiff does not presently have in its possession but hopes to obtain through discovery.

The incentives created by advance knowledge that the complaint must satisfy a standard of minimal plausibility and reasonableness would do more than any word-smithing of Rule 8 to ensure that complainants plead cases with reasonable specificity. The practice of preliminary judicial review of complaints before service in Germany reportedly has exactly that effect: those drafting complaints want to put enough in them to convince the judicial official reviewing them before service that there is a valid basis for suit so that they will summon the defendant without further ado:

> The expectation of preliminary review helps deter frivolous complaints. Yet that review should not deter many meritorious complaints, since plaintiffs do not plead at their peril. Should the judge have concerns about whether the procedural prerequisites are met, or about whether the complaint sufficiently substantiates the factual allegations, the judge is to direct the plaintiff to clarify the point before dismissing the case.[235]

In situations in which the complaint itself does not contain enough information to verify that it is reasonable for the government to put the defendant to the time and trouble to answer a lawsuit, the reviewing magistrate should telephone or invite in the plaintiff's lawyer for an informal oral conference and ask appropriate questions, in much the same way that judges and magistrates already do before issuing search warrants. This oral conference would be similar to the first status conference that is typically held today, in which the judge finds out from the parties what the case is about.

The conference, if needed because not enough information is provided in support of the complaint, could ordinarily involve only the plaintiff's lawyer to avoid imposing unnecessary burdens on the prospective defendants before the state has verified that there is a reasonable basis to do so. In appropriate instances, the reviewing magistrate could in his or her discretion also invite[236] in the prospective defendant(s) or consult with them by telephone to learn their side of the story.[237]

For example, a reviewing magistrate tasked with determining the *bona fides* of a claim before service could often determine quickly and inexpensively by consulting the defendant that many of the putative defendants either did not exist or did not manufacture the products in question. Over time, reviewing magistrates would develop experience that many defendants are often wrongly included in certain kinds of cases, and they would start asking this question of plaintiffs' lawyers. To forestall the inquiry, plaintiffs might start determining who is involved before they file their cases, and reciting the same in the complaints before they file them.

If a reviewing judge or magistrate decides to hold a conference rather than sign off on the complaint, the preliminary complaint review and verification conference should be on the record before a court reporter. A transcript should be made and, in accordance with the usual final judgment rule, an appeal would be available if the reviewing judge refuses to authorize service of the complaint, but not if the judge decides to proceed with service.

Plaintiffs should be encouraged by gentle questioning about missing evidence to identify in their complaints any crucial "missing link" evidence that they anticipate obtaining through discovery. For example, a plaintiff might state: "Despite having interviewed all of the decedent's known co-workers, I do not currently have product identification evidence for 8 of the 10 manufacturers named, but I hope to obtain it through discovery of their records showing that they sold their products to the employers where he worked." The court can then assess whether it is sufficiently likely that the crucial evidence will turn up that it is reasonable to go forward. The threshold for reasonableness would be lower if plaintiffs' lawyers routinely paid for the costs of inquiries to try to find missing evidence.[238]

As we routinely do in criminal cases, or when courts are asked to enforce administrative subpoenas,[239] in habeas corpus cases, or as many other procedural systems also do in civil cases, it is possible for the state to conduct a modest preliminary inquiry into the *bona fides* of cases *before* the state summons the defendant to appear and begin spending resources. I argue that minimal preliminary inquiry by the state is constitutionally *required*, but regardless of whether that is the case, the constitutional values at stake show that as a matter of policy, the rules should require a magistrate judge to conduct a preliminary inquiry into the likely merit of a claim before those sued are required to answer it. This should be done at the Rule 4 stage, before the federal government orders the persons sued to appear in court and compels them to begin expending their resources to answer the claim.

To those who would object that a PSPD is impractical, I would remind them that (1) we already do it in many civil cases and (2) we did it routinely between 1789 and 1938.

Endnotes

1 Samuel Taylor Coleridge, Biographia Literaria 13(Ernest Rhys ed., 1906), *available at* http://www.archive.org/details/biographialitera027747mbp.

2 Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007) [hereinafter Twombly] (requiring plaintiffs to plead sufficient facts to show their claims are "plausible").

3 Fed. R. Civ. P. 4(b)("If the summons is properly completed, the clerk must sign, seal, and issue it to the plaintiff for service on the defendant."), and cases cited *infra* note 60.

4 The proper word is "hale," not "haul," although they have a common root in Middle English. *See* Definition of Hale, http://www.yourdictionary.com/hale (last visited Dec. 13, 2011).

5 "As a litigant, I should dread a lawsuit above all else, other than sickness and death." Judge Learned Hand, *The Deficiencies of Trials to Reach the Heart of the Matter*, 3 Lectures on Legal Topics 87, 105 (1926).

6 Charles Warren & Louis D. Brandeis, *The Right To Privacy*, 4 Harv. L. Rev. 193 (1890).

7 Fed. R. Crim. Pro. 9(a) ("The court must issue . . . at the government's request, a summons . . . if one or more affidavits accompanying the information establish probable cause to believe that an offense has been committed and that the defendant committed it.").

8 28 U.S.C. §1915(e)(2)(B).

9 For a summary of a typical anti-SLAPP statute, see *Oasis West Realty v. Goldman*, S181781 (Cal. May 16, 2011), *available at* http://scocal.

stanford.edu/opinion/oasis-west-realty-v-goldman-33970 ("Section 425.16, subdivision (b)(1), provides: 'A cause of action against a person arising from any act of that person in furtherance of the person's right of petition or free speech under the United States or the California Constitution in connection with a public issue shall be subject to a special motion to strike, unless the court determines that the plaintiff has established that there is a probability that the plaintiff will prevail on the claim. . . . Only a cause of action that satisfies *both* prongs of the anti-SLAPP statute—i.e., that arises from protected speech or petitioning *and* lacks even minimal merit—is a SLAPP, subject to being stricken under the statute.").

10 United States v. Morton Salt, 338 U.S. 632 (1950) (court will enforce administrative subpoena if, but only if, "reasonable").

11 For a typical rule, see, e.g., Local Rule 72-3.2, United States District Court, Central District of California, *available at* http://www.cacd.uscourts.gov/CACD/LocRules.nsf/a224d2a6f8771599882567cc005e9d79/4ef5256bde5a03708825768e005825e3?OpenDocument L.R. 72-3.2 ("**Summary Dismissal of Habeas Corpus Petition.** The Magistrate Judge promptly shall examine a petition for writ of habeas corpus, and if it plainly appears from the face of the petition and any exhibits annexed to it that the petitioner is not entitled to relief, the Magistrate Judge may prepare a proposed order for summary dismissal and submit it and a proposed judgment to the District Judge."). The district court may enter an order for the summary dismissal of a habeas petition "[i]f it plainly appears from the face of the petition and any exhibits annexed to it that the petitioner is not entitled to relief in the district court. . . ." *Id.* The district court may enter an order for the summary dismissal of a habeas petition "[i]f it plainly appears from the face of the petition and any exhibits annexed to it that the petitioner is not entitled to relief in the district court. . . ." Rule 4, Rules Governing Section 2254 Cases (West 1977). Summary dismissal is appropriate if the allegations in the petition are "vague [or] conclusory" or "palpably incredible" or "patently frivolous or false." Blackledge v. Allison, 431 U.S. 63, 75-76 (1977).

12 See *infra* pp. 115-117.

13 See *infra* pp. 125-128.

14 *See* Harold J. Krent, The Private Performing the Public: Delimiting Delegations to Private Parties (Nov. 24, 2010) *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1714497 ("Although the propriety of the private exercise of public powers rarely has been litigated, the Supreme Court struggled with that question during the New Deal when Congress created a number of innovative governance structures combining public and private entities in an effort to end the Great Depression. Indeed, on several occasions, such as in *Carter v. Carter Coal* and in *A.L.A. Schechter Poultry Corp. v. United States*, the Court invalidated delegations in part because of the role accorded private parties." (footnotes and citations omitted)).

For more general discussions of the issues raised by delegations of government functions to private actors, see Paul R. Verkuil, Outsourcing Sovereignty (2007); Gillian E. Metzger, *Privatization as Delegation*, 103 Colum. L. Rev. 1367 (2003); Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. Rev. 543 (2000).

15 See *infra* pp. 117-118.

16 Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007).

17 Elsewhere I have argued that regulating by incentives is more efficient than by judicial command and control. E. Donald Elliott, *Managerial Judging and the Evolution of Procedure*, 53 U. Chi. L. Rev. 306 (1986); E. Donald Elliott, *Toward Incentive-Based Procedure: Three Approaches for Regulating Scientific Evidence*, 69 B.U. L. Rev. 487 (1989). While I adhere to my view from a quarter century ago that incentive-based procedure would be a theoretical first-best solution, I have detected no willingness on the part of either the judiciary or the Congress to abandon the so-called "American Rule" and move to a loser-pays system; instead, our current legal culture regulates litigation behavior, if at all, by judicial command-and-control (aka "managerial judging" or judicial discretion). For example, I argued (very persuasively I thought) in the second article cited above that the courts could better regulate scientific evidence through a system of economic incentives rather than preliminary screening by judges. Nonetheless, in *Daubert v. Merrell Dow Pharmaceuticals, Inc.,* 509 U.S. 579 (1993), the Supreme Court imposed a system of preliminary screening by judges. The proposals in this paper should be regarded as a second-best approach, like *Daubert,* that moves in the right direction by improving the incentive structure of the litigation market in a way that is not only politically plausible because it is more consistent with our legal culture, but is also arguably constitutionally required.

18 *In re* Sinking Fund Cases, 99 U.S. 700, 718-19 (1878) ("The United States cannot any more than a State interfere with private rights, except for legitimate governmental purposes. They are . . . prohibited from depriving persons or corporations of property without due process of law.").

19 Fed. R. Crim. Pro. 9(a) ("The court must issue . . . at the government's request, a summons . . . if one or more affidavits accompanying the information establish probable cause to believe that an offense has been committed and that the defendant committed it.").

20 The metaphor is Margaret Mead's: "[I]f a fish were to become an anthropologist, the last thing it would discover would be water." *Quoted in* Doing the Ethnography of Schooling: Educational Anthropology in Action 24 (G.D. Spindle ed., 1982).

21 A notable exception is Professor Carrington's thoughtful article on Rule 4. Paul D. Carrington, *Symposium: The Fiftieth Anniversary of The Federal Rules of Civil Procedure: Article: Continuing Work on The Civil Rules: The Summons*, 63 Notre Dame L. Rev. 733 (1988). However, Professor Carrington also focuses on the mechanics and consequences of service of process, and does not address the considerations raised in this article.

22 *See, e.g.*, Samuel Issacharoff, Civil Procedure 17 (Foundation Press, 2d ed. 2009). Following an introductory chapter on general concepts of due process, even Professor Issacharoff, one of the most thoughtful of modern proceduralists, jumps right into the general rules of pleading and Rule 12(b) motions to dismiss, without even mentioning Rule 4. This is the conventional approach, but it overlooks the important incentive structure that is already established by Rule 4. The person being sued—now arbitrarily reclassified by the state as a "defendant"—must expend resources to convince the court that the charges against him are baseless.

23 *See, e.g.*, Barbara Allen Babcock, Toni M. Massaro & Norman W. Spaulding, Civil procedure: Cases and Problems 35-55 (4th ed. 2009) [hereinafter Babcock Civil Procedure Casebook].

24 Bell Atlantic Corp. v. Twombly, 550 U.S. 544 (2007).

25 Ashcroft v. Iqbal, 129 S. Ct. 1937 (2009).

26 See i*nfra* p. 112 and note 36.

27 In theory, Rule 11 may provide such a remedy, but in practice it is rarely invoked, in large part because it is focused on "sanctions" (punishment) rather than "costs" (reimbursement). *See* Robert Cooter, *Prices and Sanctions*, 1984 Colum. L. Rev. 1523 ("Officials should create prices to compel decisionmakers to take into account the external costs of their acts, whereas officials should impose sanctions to deter people from doing what is wrong.").

28 *Id.* at 557-58 (citations omitted).

29 SOLIDFX, LLC v. Jeppesen Sanderson, Inc., 11-CV-01468-WJM-BNB, 2011 WL 4018207 (D. Colo. Sept. 8, 2011) (*Twombly* "does not erect an automatic, blanket prohibition on any and all discovery before an antitrust plaintiff's complaint survives a motion to dismiss.").

30 Fed. R. Civ. P. 12(i).

31 341 F. Supp. 2d 449 (D.N.J. 2004).

32 "Church & Dwight Co., the makers of Arm & Hammer, was forced to retain Morgan, Lewis & Bockius to file multiple briefs in the federal court at not inconsiderable expense to rid itself of this nuisance suit." Posting of Ted Frank to Overlawyered, http://overlawyered.com/2006/12/because-we-all-love-wacky-pro-se-suits-ward-v-arm-hammer/ (Dec. 18, 2006).

33 See *infra* p. 117 regarding the discretion that the clerk's office exercised prior to the current version of Rule 4 not to issue a summons if a long time had passed since the events.

34 See *infra* pp. 115 regarding the "ministerial" and non-discretionary nature of current Rule 4(b).

35 "'Rent seeking' is one of the most important insights in the last fifty years of economics . . . . The idea is simple but powerful. People are said to seek rents when they try to obtain benefits for themselves through the political arena."

David R. Henderson, *Rent Seeking*, *in* The Concise Encyclopedia of

Economics (2d ed. 2008), *available at* http://www.econlib.org/library/Enc/RentSeeking.html.

36 *See* Neitzke v. Williams, 490 U.S. 319, 326-27 (1989) ("Rule 12(b)(6) authorizes a court to dismiss a claim *on the basis of a dispositive issue of law* . . . .") (emphasis supplied).

37 Professor Bone's otherwise excellent analysis of the economics of civil procedure suffers by assuming that detailed fact pleading will screen out baseless cases. Robert G. Bone, Civil Procedure: The Economics of Civil Procedure 125-157 (2003); *id.* at 126 ("One advantage of using detailed pleading requirements to screen frivolous suits is that pleading operates as a gatekeeper."); *see also* Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 Colum. L. Rev. 433 (1986).

It has never been demonstrated that detailed pleading requirements actually "screen" cases to any significant degree. It is equally plausible that clever lawyers will file most of the same cases anyway, merely pleading them in more detail. This is borne out by a study by the Federal Judicial Center of motions to dismiss after *Iqbal* and *Twombly*, which found more motions to dismiss were made but leave to amend was generally granted, resulting merely in more detailed pleadings. Joe S. Cecil, George W. Cort, Margaret S. Williams & Jared J. Bataillon, Fed. Judicial Ctr., Motions to Dismiss for Failure to State a Claim After Iqbal: Report to the Judicial Conference Advisory Committee on Civil Rules (2011); *see also* Patricia W. Hatamyar, *The Tao of Pleading: Do* Twombly *and* Iqbal *Matter Empirically?*, 59 Am. U. L. Rev. 553, 556 (2010) (study of reported cases indicates that pleading stage dismissals have increased after *Twombly* and *Iqbal*, from forty-six to fifty-six percent of cases in which the motion to dismiss is made). However, such short-term effects of rules changes tend to equilibrate as lawyers adapt their strategies to the new rules. *See* Linda Cohen & Matthew Spitzer, *Solving the Chevron Puzzle*, 57 Law & Contemp. Probs. 67 (1994); *see also* Ian Ayres, *Playing Games with the Law*, 42 Stan. L. Rev. 1291 (1990) (arguing the dynamic game theory models are inherently superior to static micro-economic models for understanding law because players can adjust their strategies as rules change).

38 See *infra* pp. 128-129 for a description of the preliminary processes that should be incorporated into Rule 4 to screen claims before they are served, the "Pre-Service Plausibility Determination." They differ from fact-pleading in that, in appropriate circumstances, the court should ask questions, rather than being bound by the assumption that the factual allegations of the complaint are true even if implausible and unsubstantiated. In an early article after *Twombly*, Richard Epstein recognized that the "plausibility" standard inherently involved factual inquiry and thereby elided the traditional distinctions between motions to dismiss and motions for summary judgment. Richard A. Epstein, Bell Atlantic v. Twombly*: How Motions to Dismiss Become (Disguised) Summary Judgments*, 25 Wash. U. J.L. Pol'y 61 (2007). Epstein has subsequently proposed a system in which judges would re-evaluate on an ongoing basis throughout the course of a case whether the case is strong enough to go to the next stage. Richard A. Epstein, *Of Pleading and Discovery: Reflections on* Twombly *and* Iqbal *with Special Reference to Antitrust*, 2011 U. Ill. L. Rev. 187 (2011). I regard my suggestion for a Pre-Service Plausibility Determination before service of process as in the same spirit.

39 *See, e.g.*, Robert L. Carter, *The Federal Rules of Civil Procedure as a Vindicator of Civil Rights,* 137 U. Pa. L. Rev. 2179, 2179 (1989), *citing* Judith Resnik, *Failing Faith: Adjudicatory Procedure in Decline,* 53 U. Chi. L. Rev. 494, 497 & n.11, 501 & n.29 (1986); *see also* Robert L. Carter, *Civil Procedure as a Vindicator of Civil Rights: The Relevance of* Conley v. Gibson *in the Era of Plausibility Pleading*, 52 Howard L.J. 17 (2008-2009).

40 "In almost all other countries, except Japan and China, the winning party, whether plaintiff or defendant recovers at least a substantial portion of litigation costs." ALI/Unidroit Principles of Civil Procedure 6 & sources cited *id.* n.8 (2006).

41 For a definition, see i*nfra* p. 126.

42 See *infra* pp. 117-122 discussing the drafting history of Rule 4.

43 See *infra* pp. 116-117, discussing the 1789 Judiciary Act and the 1792 Process Act, An act for regulating processes in the Courts of the United States, and providing compensation for the officers of the said Courts, and for jurors and witnesses. Ch. 36, §1, 1 Stat. 275 (May 8, 1792) ("*Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled*, That all writ processes issuing from the supreme or a circuit court, shall bear test of the chief justice of the supreme court (or if that office shall be vacant) of the associate justice next in precedence; and all writs and processes issuing from a district court, shall bear test of the judge of such court (or if that office shall be vacant) of the clerk thereof, which said writs and processes shall be under the seal of the court from whence they issue, and signed by the clerk thereof."). This followed verbatim in relevant part a statute enacted on a temporary basis by the First Congress. An Act to regulate Processes in the Courts of the United States. Ch.21, §1, 1 Stat. 93 (Sept. 29, 1789); *see* Middleton Paper Co. v. Rock River Paper Co., 19 Fed. 252, 253-254 (Cir. W.D. Wisc. 1884)("The summons, notice, writ or whatever it may be called, by virtue of which a defendant is required to come in court and answer, litigate his rights and submit to the personal judgment of the court is certainly process within the meaning of the law of congress . . . is to be issued by the clerk of this court, under the seal of the court and tested in the name of the Chief Justice of the United States. . . . *It is no doubt the policy of the law to keep process under the immediate supervision and control of the court*." (emphasis supplied)).

44 Daubert v. Merrell Dow Pharms., Inc.*,* 509 U.S. 579 (1993).

45 From 1983 to 1993, the Federal Rules of Civil Procedure experimented with mandatory imposition of sanctions under Rule 11. Most commentators agree that this experiment with mandatory financial sanctions for frivolous cases and motions was a disaster, William W. Schwarzer, *Rule 11 Revisited*, 101 Harv. L. Rev. 1013 (1988), and therefore anything like it is unlikely to be tried again soon. There is, however, an important conceptual difference between costs for consuming resources and sanctions as punishment. *See* Cooter, *supra* note 27. The distinction is often overlooked by judges who tend to equate the two.

46 "Sometimes there are reasons to sue even when one cannot win. . . . The first attorney to challenge *Plessy v. Ferguson* was certainly bringing a frivolous action. But his efforts and the efforts of others eventually led to *Brown v. Board of Education*." Eastway Constr. Corp. v. City of New York, 637 F. Supp 558, 575 (E.D.N.Y. 1986). The author agrees with the first sentence, but not the second. Claims are not frivolous merely because they advance an "argument for extending, modifying, or reversing existing law or for establishing new law." Fed. R. Civ. P. 11(b)(2). But the rhetoric is symptomatic of a concern about over-deterring claims worth hearing that is prevalent in our legal culture. It does not follow, however, that the decision of what arguments for changes to existing law should take up the time of others should be delegated without judicial supervision to plaintiffs' lawyers with a financial stake in the outcome. See *infra* pp. 125-126.

47 Fed. R. Civ. P. 26(b)(2)(C)(iii):

> (C) *When Required.* On motion or on its own, the court must limit the frequency or extent of discovery otherwise allowed by these rules or by local rule if it determines that:
>
> . . . .
>
> (iii) the burden or expense of the proposed discovery outweighs its likely benefit, considering the needs of the case, the amount in controversy, the parties' resources, the importance of the issues at stake in the action, and the importance of the discovery in resolving the issues.

48 *See, e.g.*, United States v. Morton Salt, 338 U.S. 632 (1950) (court will enforce administrative subpoena if, but only if, "reasonable").

49 There are undoubtedly rejoinders to many of the constitutional arguments that I propose, but I will leave them to others. This is not only because the length of this Article already strains the patience of law review editors, but because my primary purpose to to locate *Twombly* into the context of history and values of constitutional dimension, and also to suggest that the problem of distorted incentives is better solved by Pre-Service Plausibility Determinations by the judiciary than by enhanced pleading requirements and motions to dismiss.

50 Oliver Wendell Holmes Jr., *The Path of the Law*, 10 Harv. L. Rev. 457, 461 (1897) (By "law" I mean "[t]he prophecies of what the courts will do in fact and nothing more pretentious . . . .").

51 For example, a distinguished proceduralist, Professor Arthur Miller, has recently published a long and impassioned defense of keeping the courts

open to all comers no matter how unreasonable and unsupportable their charges may be, based primarily on the history of the federal rules. Arthur R. Miller, *From* Conley *to* Twombly *to* Iqbal*: A Double Play on the Federal Rules of Civil Procedure*, 60 Duke L.J. 1 (2010). His historical focus overlooks the fundamental countervailing constitutional principles of privacy (about which he has been passionate in other contexts) and that the state may not arbitrarily favor one group of citizens over another without a reasonable basis for doing so. *Cf.* Oliver Wendell Holmes Jr., *The Path of the Law*, 10 Harv. L. Rev. 457, 469 (1897) ("It is revolting to have no better reason for a rule of law than that so it was laid down in the time of Henry IV.") Moreover, according to Professor Miller, history apparently began in 1938 with the adoption of the Federal Rules of Civil Procedure. In fact, however, on this point the Federal Rules were a sharp and unwise departure from our long tradition that persons sued were also entitled to reasonable protection against an unwarranted invasion of their privacy. See *infra* pp. 115-117.

52 See *infra* pp. 124-125.

53 *See, e.g.*, Connecticut v. Doehr*,* 501 U.S. 1 (1991) (prejudgment attachment of real property without prior notice and hearing, exigent circumstances or requirement to post a bond violates 14th Amendment due process), and other cases discussed *infra* at pp. 124-125.

54 *See, e.g.*, Babcock Civil Procedure Casebook, *supra* note 23, at 475 ("[T]he Rules transferred power in the form of access to information from corporate defendants to individual plaintiffs."); Paul Brodeur, Outrageous Misconduct: The Asbestos Industry On Trial (1985).

55 Thomas O.McGarity & Wendy E. Wagner, Bending Science: How Special Interests Corrupt Public Health Research (2008).

56 "But framing the problem in terms of assisting individual plaintiffs in their suits against corporate defendants is unsatisfactory. Discovery concededly may work to the disadvantage as well as to the advantage of individual plaintiffs. Discovery, in other words, is not a one-way proposition. It is available in all types of cases at the behest of any party, individual or corporate, plaintiff or defendant." Hickman v. Taylor, 329 U.S. 495, 507 (1947).

57 See *infra* p. 129.

58 From an incentive-based perspective, one of the problems with the filing fee is that there is no marginal cost for adding additional defendants (beyond the minimal cost for service). Therefore, we should not be surprised that some defendants are named who have little or nothing to do with the matter. It costs the plaintiff's lawyer merely postage to name additional defendants, but many of them can be expected to pay nuisance value to settle.

59 Fed. R. Civ. P. 4(b).

60 Bauers v. Heisel, 361 F.2d 581, 595 n.3 (3d Cir. 1966), *cert. denied*, 396 U.S. 1021 (1967); Charles A. Wright & Arthur R. Miller, 4A Federal Practice & Procedure: Civil § 1084 (3d ed. 2010) ("The current rule . . . makes it clear that the only formal requirement for the issuance of a summons is the filing of a valid complaint with the court."). The use of the word "valid" before the word "complaint" in the Wright & Miller treatise might be read to suggest that there is discretion under the existing rule for the clerk's office to decline to issue a summons if it determines that the complaint is palpably deficient. While that is the result for which this Article argues, both the language of the present rule and the case law construing it would make it difficult to accomplish this result without a rule change. *But cf.* Mitchell v. Beaubouef, 581 F.2d 412 (5th Cir. 1978), *reh'g denied,* 586 F.2d 842 (5th Cir. 1978), *cert. denied*, 441 U.S. 966 (1979) (If a prisoner's *pro se* complaint is deemed legally sufficient under the liberal standard appropriate to *pro se* prisoner litigation, the process must be served on the defendant.).

61 See *infra* for a discussion of 28 U.S.C. §1915(e), under which cases brought *in forma pauperis* are singled out for *sua sponte* review before service of process. *See also* Neitzke v. Williams*,* 490 U.S. 319 (1989), discussed *infra* p. 127.

62 The sign and seal are essential. Ayres v. Jacobs & Crumplar, P.A., 99 F.3d 565, (3d Cir. 1996)("Requiring the Clerk to sign and issue the summons assures the defendant that the process is valid . . . . [A] summons not issued and signed by the Clerk with the seal of the court affixed thereto fails to confer personal jurisdiction over a defendant even if properly served."); *accord* 2 James W. Moore, Moore's Federal Practice 4.05 (2d ed. 1996) ("Under Rule 4(b) only the clerk may issue the summons . . . . [A] summons issued by the plaintiff's attorney is a nullity."); 4A Charles Alan Wright & Arthur R. Miller, Federal Practice and Procedure § 1084 (2d ed. 1987).

63 Philip K. Howard, *There Is No Right to Sue*, Wall St. J., July 31, 2002, at A14, *available at* http://commongood.org/society-reading-cgpubs-opeds-13.html; *see also* Philip K. Howard, *Making Civil Justice Sane: Judges Should Stop Unreasonable Lawsuits Before They Start*, City J., Spring 2006, at 64, *available at* http://www.city-journal.org/printable.php?id=1989 ("Juries in a criminal case are our protection against abuses of state power. But a private lawsuit, we seem to have forgotten, is a use of state power against another private citizen. Filing a lawsuit is just like indicting someone—it's just an indictment for money. Without the protection of a disinterested prosecutor and a grand jury, the defendant needs the protection of the judge to decide whether the claim has legal merit, leaving the jury to decide disputed facts."); *see also* William A. Alderson, A Practical Treatise Upon the Law of Judicial Writs and Process in Civil and Criminal Cases: The Sufficiency, Validity, Amendment and Alteration of Process; Its Execution and Return, and The Powers and Liabilities of Officers Thereunder §3 at 9 (New York, Baker, Voorhis & Company 1895)("[S]ummonsing the party to answer to a complaint and enforcing a judgment were originally private acts, which were not authorized but only permitted by the state. Progress in the modes of judicial procedure resulted in rendering such acts those of the state, and the issuing and execution of all writs and process are now the exercise of state powers. This fact is important to remember in determining the validity of writs and process and the acts of the officer performed in executing them. Courts have too often been forgetful of this fact, which should constitute the premise of the argument, and have pronounced upon such matters as though the state had no concern therein.").

64 It is a commonplace in cultural anthropology that a "culture" consists of those things that the people in it do not see because they take for granted. Lucila L. Salcedo, Social Issues 12 (1999) ("Culture also affects all the things that we take for granted and what we question."); Edgar Schein, Corporate Culture Survival Guide 119-120 (1999)("shared mental models that members of an organization hold and take for granted"); Geert Hosstede, *National Cultures and Corporate Cultures*, *in* Communication Between Cultures 51 (Larry A. Samovar, Richard E. Porter & Edwin R. McDaniel eds., 1984) ("Culture is the collective programming of the mind which distinguishes the members of one category of people from another.").

65 William Feilden Craies, *Summons*, *in* Encyclopedia Britannica (1911) (Defining a summons as "(1) a command by a superior authority to attend at a given time or place or to do some public duty; (2) a document containing such command, and not infrequently also expressing the consequences entailed by neglect to obey."), *available at* http://en.wikisource.org/wiki/1911_Encyclop%C3%A6dia_Britannica/Summons; *see also* Leas & McVitty v. Merriman, 132 F. 510 (Cir. W.D. Va. 1904) (distinguishing between a federal summons which orders the defendant to appear, and summons in some states that merely provide notice of a claim but are not orders from the court to appear).

66 See *infra* pp. 118-120.

67 That the state must define reasonable and predictable grounds upon which someone may be held to account is arguably the core meaning of clause 39 of Magna Carte (1215), upon which our concept of "due process of law" is based: "No free man shall be taken or imprisoned, or dispossessed, or outlawed, or banished, or in any way destroyed, . . . except by the legal judgment of his peers or by the law of the land."

68 An 1890 decision by Queens Bench, *Reg. v. Colonel Byrde and Others, Justices, and the Pontypool Gas Company*, 63 Law Times Rep. N.S. 645 (QB Div., 1890), reiterates this long-standing understanding. ("The justices may . . . in the exercise of their discretion refuse to issue a summons, even though there is evidence before them of an alleged indictable misdemeanour, if they consider that the issue of the summons would be vexatious or improper" (citing Reg. v. Ingham, 14 Q. B. 396 (1849).)). The *Pontypool Gas Co.* case involved forfeiture of statutory penalties for failure to construct a reservoir in a timely manner, and thus it is debatable whether it is properly classified as a criminal or civil case. However, since the same writs were used to summon the prospective defendant, this distinction seems not to have mattered.

69 See *infra* pp. 115-116, discussing the practice in Colonial courts as described by Judge Betts in 1838.

70 Samuel R. Betts, A Summary of Practice in Instance, Revenue and Prize Causes, in the Admiralty Courts of the United States for the

Southern District of New York; and also on Appeal to the Supreme Court: Together with the Rules of the District Court 23-24 (New-York, Halsted and Voorhies, 1838) ("In this district process emanates from the court correspondent to the libellant's case. The actual practice for a half century was to read the libel in open court, and thereupon pray and receive directions for the appropriate process. This direct agency of the court has been discontinued since the revolution, but the principle upon which the usage was founded, yet enters into and influences the practice.").

71 *Id.* (emphasis supplied).

72 See also *Middleton Paper Co. v. Rock River Paper Co.,* 19 Fed. 252, 254 (Cir. W.D. Wisc. 1884), an ordinary, non-admiralty civil case between two paper companies involving garnishment of a debt, in which the court stated "It is no doubt the policy of the law to keep process under the immediate supervision *and control of the court*." (emphasis supplied).

73 Conkling served as a federal district judge in the northern district of New York from 1825 to 1852. Alfred Conkling, http://en.wikipedia.org/wiki/Alfred_Conkling (last visited Dec. 14, 2011).

74 2 Alfred Conkling, The Admiralty Jurisdiction, Law and Practice Courts Of The United States: With An Appendix, Containing The New Rules of Admiralty Practice Prescribed by the Supreme Court of The United States, Those of the Circuit and District of The United States for The Northern District of New York, and Numerous Practical Forms of Process, Pleadings, Stipulations, Etc., Comprising The Entire Progress of a Suit in Admiralty Accompanied by Explanatory Notes 70-72 (Albany, W.C. Little & Co. 1857).

75 The Judiciary Act of 1789, September 24, 1789. 1 Stat. 73. Chap. XX §14. ("*And be it further enacted*, That all the before-mentioned courts of the United States, shall have power to issue writs of *scire facias*, *habeas corpus*, and all other writs not specially provided for by statute, which may be necessary for the exercise of their respective jurisdictions, and agreeable to the principles and usages of law.").

76 Thomas Sergeant, Constitutional Law: Being a View of the Practice and Jurisdiction of the Courts of the United States, and of Constitutional Points Decided 143 (Philadelphia, Abraham Small, 2d ed. 1830) (citing United States v. Burr, Appendix, 2d part, 186 (opinion of Marshall, C.J.).

77 An act for regulating processes in the Courts of the United States, and providing compensation for the officers of the said Courts, and for jurors and witnesses. Ch. 36, §2, 1 Stat. 275 (May 8, 1792).

78 *See* Dwight v. Merritt, 4 Fed 614 (Cir. N.Y. 1880) (holding that a summons issued by a lawyer pursuant to state practice is invalid to compel someone to answer in federal court).

79 An act for regulating processes in the Courts of the United States, and providing compensation for the officers of the said Courts, and for jurors and witnesses. Ch. 36, §1, 1 Stat. 275 (May 8, 1792). This followed verbatim in relevant part a statute enacted on a temporary basis by the First Congress in 1789 a few months after ratification of the Constitution. An Act to regulate Processes in the Courts of the United States. Ch.21, §1, 1 Stat. 93 (Sept. 29, 1789).

80 "Teste, *practice*. The teste of a writ is the concluding clause, commencing with the word witness, &c." John Bouvier's Law Dictionary 562 (2d. ed., 1843).

81 William A. Alderson, A Practical Treatise Upon the Law of Judicial Writs and Process in Civil and Criminal Cases: The Sufficiency, Validity, Amendment and Alteration of Process; Its Execution and Return, and the Powers and Liabilities of Officers Thereunder §39 at 70 (New York, Baker, Voorhis & Company 1895) ("Once the seal was everything, the signature nothing. In modern times the signature is regarded at least as of much importance as the seal. . . . *The signature is independent evidence of the authorized delegation of the power of state in judicial proceedings*."(emphasis supplied and footnote omitted)); *see also* Ins. Co. v. Hallock, 73 U.S. 556 (1887) (unanimously invalidating title to land purchased at sheriff's sale because writ under which sale had been conducted had not been properly sealed: "Without the seal it is void. We cannot distinguish it from any other writ or process in this particular.").

82 *See* Middleton Paper Co. v. Rock River Paper Co., 19 Fed. 252, 253-54 (Cir. W.D. Wisc. 1884) ("The summons, notice, writ or whatever it may be called, by virtue of which a defendant is required to come in court and answer, litigate his rights and submit to the personal judgment of the court is certainly process within the meaning of the law of congress and the rule of the court, which is to be issued by the clerk of this court, under the seal of the court and tested in the name of the Chief Justice of the United States. . . . *It is no doubt the policy of the law to keep process under the immediate supervision and control of the court*."(emphasis supplied)).

83 *Id.*

84 *See* William Stewart Simkins, A Federal Suit at Law 22 (1912) ("Form of process for the commencement of suits is controlled by the conformity act, sec. 914, U. S. Rev. Stat., U. S. Comp. Stat. 1901, p. 684, except as to the official signature, seal and test, which by sec. 911, U. S. Rev. Stat., U. S. Comp. Stat. 1901, p. 083, is required to all writs and processes issuing from the courts of the United States. Gillum v. Stewart, 112 Fed. 32; Chamberlain v. Mensing, 47 Fed. 436; Shepard v. Adams, 168 U. S. 624, 42 L. ed. 604, 18 Sup. Ct. Rep. 214. Congress having thus legislated as to the process, it must be followed (Dwight v. Merritt, 18 Blatchf. 305, 4 Fed. 614), though the State law permitted an attorney to issue the summons (Middleton Paper Co. v. Rock River Paper Co. 19 Fed. 252; AEtna Ins. Co. v. Hallock (AEtna Ins. Co. v. Doe) 6 Wall. 556, 18 L. ed. 948).").

85 Congress created the system of magistrates (now called "magistrate judges") in the Federal Magistrates Act of 1968, Pub. L. 90-578, 82 Stat. 1115 (1968), a generation after the federal rules. However, a system of "United States Commissioners" had existed since 1793 to try petty offenses, issue search and arrest warrants, set bail, and the like. Commissioners were limited by background and experience to criminal cases, and I can find no evidence that the possibility of using commissioners or magistrate judges rather than assistant clerks to screen cases was considered by the drafters of Rule 4. Today, "[a]s a practical matter, the sub-judiciary has become indispensable. Federal judges likely could not manage their caseloads effectively without delegating some tasks to magistrates and special maters." Babcock Civil Procedure Casebook, *supra* note 23, at 683.

86 William A. Alderson, *supra* note 63, at 24-88; *see also* Current Decisions, *Process—Amendment—Void Summons Not Amendable*, 32 Yale L. J. 297 (1923).

87 Edson R. Sunderland, *Process*, *in* 32 Cyclopedia of Law and Procedure 412, 426 (William Nash & Howard Pervear Nash eds., 1909) (citing Steves v. Carson, 40 P. 569 (1895); Reese v. Kirby, 68 Ga. 825).

88 See *infra* pp. 120 & 121 (describing how Rule 4 was deemed to supersede the 1792 statute).

89 *See supra* note 32.

90 *Supra* p. 112.

91 *In re* Kinney, 202 F. 137, 138 (1st Cir. 1913).

92 For more information about the underlying dispute, see *Kinney v. Plymouth Rock Squab Co.*, 214 F. 766 (1st Cir. 1914). It appears that Kinney was a frequent litigant.

93 *In re* Kinney, 202 F. at 138.

94 *Id.*

95 135 F. 336 (3d Cir. 1905).

96 135 F.2d at 140 ("The questions sought to be presented in this case relate to the interpretation to be given to a law of the state, and the complaint is that this law is being misinterpreted and misapplied, to the injury of the plaintiff in his rights of property. In all such cases, where there is not the requisite diverse citizenship and amount in controversy to give the court jurisdiction, the remedy for the injuries complained of is in the state courts. As, then, the Circuit Court had no jurisdiction of the proposed action against the state judges, it follows that the use plaintiff, Kinney, sustained no legal injury whatever by the clerk's noncompliance with his *praecipe*, and failure to file his papers." (internal quotation omitted)).

97 *See* Charles Clark, *Fundamental Changes Effected by the New Federal Rules I*, 15 Tenn. L. Rev. 551, 563-64 (1939).

98 *Id.* at 564.

99 *Id.*

100 *Id.*

101 *Id.* The full context of Clark's description of what occurred is as follows:

> Now, I want to run over a few of the more striking things in the early part of the Rules, and, of course, there is not time to go into great detail. The first general matter is the way suit is commenced. We had much discussion about this matter. The New Yorkers wanted their system, which is simple service of summons on the opposing side with an exchange of pleadings between the parties, and with the Court not in the case until some action is asked of it. Under the New York system, therefore, a case can go forward very far before the Court or any of its officers, even the clerk, may know it exists. It is a very simple system, and I think it works quite satisfactorily.
>
> To many lawyers this system seemed undignified and over-simple. It came to be dubbed the "hip-pocket rule," because one lawyer said, "Why, that is just a case where the lawyer carries around the case in his hip-pocket," since the lawyer would have the pleadings and they would not be filed in the Court until some action was requested of the Court. That was one of the alternative plans we suggested in the preliminary draft. Due to the objections of lawyers, who were long on dignity, however, we adopted the original procedure in the Federal Courts, to-wit: that an action is started by filing the complaint with the clerk, and the court's process is issued by the clerk and served by a marshal. That is provided here at the beginning, Rule 3 and 4. You start a suit by taking your complaint to the clerk, and the clerk issues the summons and the summons and complaint are served by a marshal. There is a provision in Rule 4(c) that special appointments in place of the marshal shall be made freely when substantial savings in trial fees will result, but the procedure indicated is the one that was the more familiar system throughout the country.

Charles Clark, *Fundamental Changes Effected by the New Federal Rules I*, 15 Tenn. L.Rev. 551, 563-64 (1939).

102 "In the preliminary draft we presented two alternative methods of beginning an action. One provided that to begin a suit it is necessary to file a complaint with the clerk of court, have summons issued under the seal of the court and delivered to the marshal for service, and that all other pleadings and papers must be filed as well as served. The other method proposed was that permitted in many code states, which allows the lawyers to prepare the summons and complaint in their own offices and serve them without filing, and allows all papers to be withheld from the files until the point is reached at which some judicial action is asked for. All those members of the Advisory Committee who had practiced under the latter system favored it. Those members who had not practiced under this system were either opposed to it or doubtful. The reaction from the profession has been overwhelmingly in favor of the first system, which requires the complaint to be filed when the action is commenced. The Advisory Committee in its last draft has, therefore, adopted this system. Those of us who have practiced under the other yielded reluctantly. We know that the more informal system is more convenient, saves time and results in a saving of expense in cases which are settled or dismissed without judicial action, and we know from experience that the prediction of the opponents of this system of abuses and dire consequences that will flow from it are not borne out by actual experience in those states where this system is used. Nevertheless, it is after all largely a matter of speed and convenience, and as the bar of the country seems to prefer the more formal system, the Advisory Committee have recommended it to the Court." Outline of Address by William D. Mitchell on Proposed Federal Rules Civil Procedure Before the Judicial Section of the American Bar Association (Sept. 1937) (from the Clark Papers at Yale University Library, copy on file with Florida Law Review). It is not entirely clear whether Mitchell actually delivered the remarks verbatim or whether Clark or someone else merely prepared the outline of talking points for him as background for his speech, but for our purposes, it hardly matters as either way the Mitchell outline shows the drafters' contemporary understanding.

103 Box 101 in Clark Papers at Yale University Library.

104 Clark, *supra* note 101, at 564.

105 132 F. 510, 513 (Cir. W.D. Va. 1904) (emphasis supplied); *see also* Shepard v. Adams, 168 U.S. 618, 624 (1898) ("The state Code of Colorado provides that civil actions shall be commenced by the issuing of a summons or the filing of a complaint; that the summons may be issued by the clerk of the court or by the plaintiff's attorney; it may be signed by the plaintiff's attorney; it may be served by a private person not a party to the suit. All writs and process issuing from a federal court must be under the seal of the court, and signed by the clerk, and bear *teste* of the judge of the court from which they issue. Rev. St. § 911. The process and writs must be served by the marshal or by his regularly appointed deputies. Id. §§ 787, 788.").

106 Supreme Court of the State of New York, Summons, https://iapps.courts.state.ny.us/fbem/forms/summons.pdf (last visited Dec. 14, 2011).

107 New York State Unified Court System, New York City Civil Court, Starting a Case, Issuance of the Summons, http://www.nycourts.gov/courts/nyc/civil/starting.shtml#issuance (last visited Dec. 14, 2011).

108 Babcock Civil Procedure Casebook, *supra* note 23, at 104.

109 See, e.g., N.Y. CVP. LAW § 3215(a), which provides upon failure to appear, the clerk shall enter default judgment for a sum certain but only "upon submission of the requisite proof."

110 It is true that the state typically tells the plaintiff in advance what the plaintiff is going to need to do in order to acquire jurisdiction and get a default. But that advance notice to the plaintiff of what is going to be required later to show that the prospective defendant was given fair notice does not raise any of the federal constitutional issues raised by the present federal practice of officially ordering the defendant to court.

111 4 Fed 614 (Cir. N.Y. 1880).

112 *Id.* at 614-15.

113 Admittedly, not all state systems for summons still follow the New York model. Under the pernicious influence of Federal Rule 4, some may since have adopted the federal model in which the state delegates to the plaintiff's attorney the power of the state to issue an official order compelling the person sued to appear without any government quality control. If so, they too are unconstitutional. One would have to assess them individually, just as state pre-judgment attachment systems had to be assessed and amended individually in the wake of *Sniadach*. But it has long been held that state systems for summoning persons to answer in court are subject to constitutional due process restrictions. Pennoyer v. Neff, 95 U.S. 714 (1878); Mullane v. Cent. Hanover Bank & Trust Co., 339 U.S. 306 (1950).

114 One of the most influential of the drafters, Edson Sunderland of Michigan, had even written about the power of courts to refuse to serve summons in patently frivolous cases. Edson R. Sunderland, *Process*, *in* 32 Cyclopedia of Law and Procedure 412, 426 (William Nash and Howard Pervear Nash eds., 1909) ("[I]t is within the discretion of the court to allow or refuse the issuance of summons after a long delay.").

115 Charles E. Clark, *Fundamental Changes Effected by the New Federal Rules I,* 15 Tenn L. Rev. 551 (1939).

116 *Id.* at 551; *see also* Steven N. Subrin, *Charles E. Clark and His Procedural Outlook: The Disciplined Champion of Undisciplined Rules, in* Judge Charles Edward Clark 115, 115-52 (Peninah Petruck ed., 1991).

117 *Id.* at 564.

118 Armistead M. Dobie, *The Federal Rules of Civil Procedure*, 25 Va. L. Rev. 261, 264 (1939) (emphasis supplied).

119 Former Equity Rule 12 of 1912 provided, "Whenever a bill is filed, and not before, the clerk shall issue the process of subpoena thereon, *as of course*, upon the application of the plaintiff, which shall contain the names of the parties and be returnable into the clerk's office twenty days from the issuing thereof. At the bottom of the subpoena shall be placed a memorandum, that the defendant is required to file his answer or other defense in the clerk's office on or before the twentieth day after service, excluding the day thereof; otherwise the bill may be taken *pro confesso*." (emphasis supplied). *Reprinted in* Charles C. Montgomery, Montgomery's Manual Of Federal Procedure §911 at 428 (1914), *available at* http://www.archive.org/stream/manualoffederalp00montiala/manualoffederalp00montiala_djvu.txt.

120 1937 Advisory Committee Note to Rule 3, http://www.gpo.gov/fdsys/pkg/USCODE-2009-title28/html/USCODE-2009-title28-app-rulesofci-other-dup1.htm.

121 Montgomery, *supra* note 119, §913 at 429.

122 Samuel Taylor Coleridge, Biographia Literaria 13 (1817).

123 *See also* Farzana v. Ind. Dep't of Educ., 473 F.3d 703 (7th Cir. 2007) (Easterbrook, J.) ("By refusing to accept complaints (or notices of appeal) for filing, clerks may prevent litigants from satisfying time limits. To prevent this—to ensure that judges rather than administrative staff decide whether a document is adequate—Fed. R. Civ. P. 5(e) was amended in 1993 to require clerks to accept documents tendered for filing. The last sentence of this rule provides: 'The clerk shall not refuse to accept for filing any paper presented for that purpose solely because it is not presented in proper form as required by these rules or any local rules or practices.'").

124 Pueblo de Taos v. Archuleta, 64 F.2d 807, (10th Cir. 1933).

125 485 F.2d 558 (7th Cir. 1973).

126 *Id.* at 559.

127 *Id.* at 560.

128 Dear v. Rathje*,* 391 F. Supp. 1 (N.D. Ill. 1975).

129 *Id.* at 10 n.2.

130 *Id.* at 10.

131 *Id.*

132 Dear v. Rathje, 532 F.2d 756 (7th Cir. Mar. 16, 1976) (table).

133 One Cassandra who saw clearly the potential for the new rules to increase strike suits was Francis M. Finch, an Associate Justice of the New York Court of Appeals. According to the Mitchell outline found in the Clark papers at Yale, *supra* note 102, Judge Finch objected that the new rules would greatly increase the potential for strike suits, but his perceptive remarks were interpreted as merely relating to discovery, and were dismissed as relevant only to "admittedly bad" conditions in New York City where many lawyers were considered to have "low ethical standards" but not to other parts of the country where lawyers were deemed more ethical and less susceptible to economic incentives:

> At the same meeting [1936 annual meeting of the ABA] Judge Finch of the Court of Appeals of the State of New York made an address deploring the extent to which strike suits and dishonest or blackmailing cases are instituted, and he suggested that the proposed rules would open the way still further for this sort of abuse. His illustrations were taken from conditions in the City of New York. His principal suggestion was that the law should punish the plaintiff who brings a strike suit by requiring him to pay not merely the ordinary costs, but all the expenses of the defendant, including reasonable counsel fees, if the defense is successful. The Advisory Committee believes that any substantial change in the present basis for taxing costs or disbursements is a matter for the Congress and not properly embodied in the proposed rules of practice and procedure. It may be that in large metropolitan areas like New York City where the conditions are admittedly bad and many dishonest actions are brought in the courts, the rules relating to discovery and examination before trial offer opportunities to lawyers of low ethical standards. As applied to the country as a whole, we think the rules relating to these subjects are in line with modern enlightened thought on the subject and will not be subjected to abuse. Uniform rules of practice and procedure must be drawn to meet conditions generally throughout the country and not special conditions in a few areas. Our suggestion is that in places like New York City the remedy is an improvement in the machinery for disbarring or disciplining lawyers guilty of misconduct.

Outline of Address by William D. Mitchell on Proposed Federal Rules Civil Procedure Before the Judicial Section of the American Bar Association (Sept. 1937), *supra* note 102.

134 Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 557-58 (2007) (citations omitted).

135 Historical Note to 1998 Revision of 28 U.S.C. §1691, at 5 ("Based on title 28, U.S.C., 1940 ed., § 721 (R.S. § 911; Mar. 3, 1911, ch. 231, § 291, 36 Stat. 1167). *Provisions as to* teste *of process issuing from the district courts were omitted as superseded by Rule 4 (b) of the Federal Rules of Civil Procedure*.") (emphasis supplied), *available at* www.gpo.gov/fdsys/pkg/USCODE.../USCODE-1998-title28-partV.pdf.

136 28 U.S.C. §1691.

137 *Id.* ("**Seal and teste of process.** All writs and process issuing from a court of the United States shall be under the seal of the court and signed by the clerk thereof.")

138 28 U.S.C. § 2072(b) ("Such rules shall not abridge, enlarge or modify any substantive right. All laws in conflict with such rules shall be of no further force or effect after such rules have taken effect.").

139 462 U.S. 919 (1983).

140 H.R. Rep. No. 889, at 28 (1988).

141 *See generally* Peter G. McCabe, *Renewal of the Federal Rulemaking Process*, 44 Am. U. L. Rev. 1655, 1663 (1995).

142 Jackson v. Stinnett, 102 F.3d 132, --- note 3 (5th Cir. 1996) (Garza, J.) ("Another good reason not to read the abrogation clause to nullify provisions of the PLRA is that such a reading approaches a violation of the Presentment Clause and the nondelegation doctrine. The abrogation clause of the Rules Enabling Act purports to give the Supreme Court the legislative power to repeal any federal law governing practice and procedure in the courts. Under the Rules Enabling Act, the Court need only report such changes to Congress in the form of a rule, which would acquire the force of law without Congress ever casting a single vote. To say the least, such a power would strain the Constitution's limits on the exercise of the legislative power. U.S. Const. art. I, § 7, cl. 2; INS v. Chadha, 462 U.S. 919, 950-51, 103 S.Ct. 2764, 2784, 77 L.Ed.2d 317 (1983); A.L.A. Schechter Poultry Corp. v. United States, 295 U.S. 495, 529, 55 S.Ct. 837, 843, 79 L.Ed. 1570 (1935); see also Note, supra, 98 Harv.L.Rev. at 836-37. To avoid such a drastic result, we will not construe the abrogation clause to dictate that Rule 24(a) invalidates Congress's subsequent amendments of i.f.p. procedure.").

143 524 U.S. 417 (1998).

144 Clinton v. City of New York, 524 U.S. at 446 n.40 ("The Government argues that the Rules Enabling Act, 28 U.S.C. § 2072(b), permits this Court to 'repeal' prior laws without violating Article I, §7. Section 2072(b) provides that this Court may promulgate rules of procedure for the lower federal courts and that '[a]ll laws in conflict with such rules shall be of no further force or effect after such rules have taken effect.' See *Sibbach* v. *Wilson & Co.,* 312 U.S. 1, 10 (1941) (stating that the procedural rules that this Court promulgates, "if they are within the authority granted by Congress, repeal" a prior inconsistent procedural statute); see also *Henderson* v. *United States,* 517 U.S. 654, 664 (1996) (citing §2072(b)). In enacting §2072(b), however, Congress expressly provided that laws inconsistent with the procedural rules promulgated by this Court would automatically be repealed upon the enactment of new rules in order to create a uniform system of rules for Article III courts. As in the tariff statutes, Congress itself made the decision to repeal prior rules upon the occurrence of a particular event—here, the promulgation of procedural rules by this Court."). A similar argument that "Congress itself made the decision to repeal prior [administrative decisions] upon the occurrence of a particular event" was made unsuccessfully in *Chadha*. *See* E. Donald Elliott, INS v. Chadha*: The Administrative Constitution, the Constitution and the Legislative Veto*, 1983 Sup. Ct. Rev. 125, 135-137 (1984).

145 U.S. Const. amend. IV, cl. 1.

146 A functional, albeit adventurous, argument might be advanced that a summons under Civil Rule 4(b) to appear and defend should be considered a "warrant" for purposes of the additional protections of the Warrant Clause of the Fourth Amendment. A strong argument has been made on historical grounds that the higher standard requiring an advance judicial determination of probable cause (as opposed to mere reasonableness) was imposed under the Warrant Clause because officers acting under the protection of a warrant were not responsible for their actions at common law. Akhil Reed Amar*, The Fourth Amendment, Boston, and the Writs of Assistance* , 30 Suffolk U. L. Rev. 53 (1996); *see also* William J. Stuntz, *Warrant Clause*, *in* The Heritage Guide to the Constitution 326 (Edwin Meese III, David F. Forte & Matthew Spaulding eds., 2005)("When the Fourth Amendment was written, the sole remedy for an illegal search or seizure was a lawsuit for money damages. Government officials used warrants as a defense against such lawsuits."). As a practical matter, the well-known judicial reluctance to "shift costs" and/or impose costs in even the most abusive situations means that someone obtaining a civil summons under Rule 4(b) is, as a practical matter, immune from ever having to answer in damages for the costs that they impose on others, much like an officer serving a search warrant at common law.

147 For an early article observing that the Fourth Amendment applies to

civil as well as criminal cases, see Louis J. DeReuil, *Applicability of the Fourth Amendment in Civil Cases*, 1963 Duke L. J. 472 (1963). Unfortunately, however, DeReuil, who was serving at the time as an Internal Revenue Service attorney, largely limited his observations to summonses in tax cases, but he clearly maintains that the Fourth Amendment applies to orders of summons in civil cases.

148 United States v. Harris, 403 U.S. 573 (1971) (warrant may issue if there is a "substantial basis for crediting" the informant).

149 The Supreme Court last considered Rule 4 in *Hanna v. Plumber*, 380 U.S. 460 (1965) (interpreting the *Erie* doctrine to allow Rule 4 to govern service or process even when this would lead to a different outcome than the state rule).

150 335 F.3d 807 (8th Cir. 2003).

151 *Id.* at 809.

152 Larry L. Teply, Ralph U. Whitten & Denis F. Mclaughlin, Cases, Text, and Problems On Civil Procedure 32 (2d ed., 2002) ("A summons is a paper that notifies the defendant that the actions has been commenced. *It also commands the defendant to appear and defend the action by a certain date or the court will enter a judgment (a default judgment) against the defendant for the remedy demanded by the plaintiff*." (emphasis supplied)).

153 United States v. Mendenhall, 446 U.S. 544, 554 (1980) (Stewart, J., majority op.) ("[A] person has been 'seized' within the meaning of the Fourth Amendment only if, in view of all the circumstances surrounding the incident, a reasonable person would have believed that he was not free to leave.").

154 Brower v. County of Inyo, 489 U.S. 593 (1989).

155 *Id.* at 596-97.

156 *Id.* at 596-98 (first emphasis original; second emphasis supplied).

157 United States v. Johnson, 620 F.3d 685, 691 (6th Cir. 2010) (When the police yelled "stop" and the defendant obeyed, the defendant was seized.); United States v. Salazar, 609 F.3d 1059, 1066-67 (10th Cir. 2010) (An individual was seized when he got out of a truck after a command from an officer within a patrol car with flashing lights.); United States v. Jones, 562 F.3d 768, 775 (6th Cir. 2009) (A defendant was seized when he "complied with [the officer's] order to stop."); *see also* Brendlin v. California, 551 U.S. 249, 262 (2007) (A fleeing man is not seized until he is physically overpowered, but one sitting in a chair may submit to authority by not getting up to run away.).

158 California v. Hodari D, 499 U.S. 621, 626 (1991) (emphasis supplied).

159 Babcock Civil Procedure Casebook, *supra* note 23, at 104.

160 Pearson v. Chung, 961 A.2d 1067, 1070 (D.C. 2008).

161 By contrast, the magistrate courts of the Republic of South Africa specifically authorize the clerk to decline to issue a summons if an excessive amount is claimed for attorney's costs or court fees. Torquil M. Paterson, Eckard's Principles of Civil Procedure in the Magistrate's Courts 94 (5th ed., 2005).

162 Admittedly, there are lower court cases that come out the other way, but most of them merely announce the result that being required to come to court, without more, is not a constitutional "seizure" without the type of deeper historical and functional analysis made by Justice Ginsburg. *See, e.g.*, Britton v. Maloney, 196 F.3d 24, 30 (1st Cir. 1999) ("Absent any evidence that [plaintiff] was arrested, detained, restricted in his travel, or otherwise subject to a deprivation of his liberty before the charges against him were dismissed, the fact that he was given a date to appear in court is insufficient to establish a seizure within the meaning of the Fourth Amendment.").

163 510 U.S. 266, 279 (1994) (Ginsburg, J., concurring).

164 *Id.*

165 United States v. Gobey, 12 F.3d 964, 967 (10th Cir. 1993) (Under the federal rules [of criminal procedure] . . . . a summons cannot be issued in the first instance without probable cause, the decision of a neutral magistrate, and the requisite particularity."); *accord* United States v. Greenberg, 320 F.2d 467, 472 (9th Cir. 1963); United States v. Hondras, 176 F. Supp. 2d 855, 858 (E.D. Wis. 2001).

> The language of Rule 9(a) indicates that the issuance of a summons in a criminal case requires probable cause when it says that at the request of the government, the court must issue a summons "if one or more affidavits accompanying the information establish probable cause." *U.S. v. Herndon*, 546 F. Supp. 2d 854, 857-58 (E.D. Cal. 2008).

Charles Alan Wright & Andrew D. Leipolda, Federal Rules of Criminal Procedure § 51 A (2010).

166 Schulz v. I.R.S., 395 F.3d 463, 465 (2d Cir. 2005), *as clarified on reh'g*, 413 F.3d 297 (2d Cir. 2005) ("[A]bsent an effort to seek enforcement through a federal court, IRS summonses apply no force to taxpayers, and no consequence whatever can befall a taxpayer who refuses, ignores, or otherwise does not comply with an IRS summons until that summons is backed by a federal court order.").

167 United States v. Morton Salt, 338 U.S. 632 (1950) (Court will enforce administrative subpoena if, but only if, "reasonable.").

168 SEC v. Arthur Young & Co., 584 F.2d 1018, 1034 (D.C. Cir. 1978), *cert. denied,* 439 U.S. 1071 (1979) ("[T]he power to exact reimbursement as the price of enforcement is soundly exercised only when the financial burden of compliance exceeds that which the party ought reasonably be made to shoulder.").

169 A "seizure" of property occurs when "there is some meaningful interference [by the state] with an individual's possessory interests in that property." Soldal v. Cook County, Ill., 506 U.S. 56, 61 (1992).

170 "No person shall . . . be deprived of . . . property, without due process of law; nor shall private property be taken for public use, without just compensation." U.S. Const. amend. V.

171 *See* Jennifer B. Arlin, *Of Property Rights and the Fifth Amendment: FIRREA's Cross-Guarantee Reexamined*, 33 Wm. & Mary L. Rev. 293 (1991), *available at* http://scholarship.law.wm.edu/wmlr/vol33/ iss1/13 ("A taking is distinct from a deprivation in several ways. First, when the government 'takes' property, it takes it for public use and is required to pay just compensation. . . . This clause is stricter than the first clause of the Fifth Amendment, the Deprivations Clause, because it can apply only to private property being taken for public use. The Takings Clause requires no process; its only requirement is that the former property owner be reimbursed 'justly' for the value of the property." (citations omitted)).

172 The narrow exception in which the Takings Clause may also be implicated is the special circumstance discussed hereafter in which defendants are required to subsidize investigations in the "reasonable but speculative" category of cases discussed hereafter, *infra* p. 126.

173 *In re* Sinking Fund Cases, 99 U.S. 700, 718-19 (1878) ("The United States cannot any more than a State interfere with private rights, except for legitimate governmental purposes. They are . . . prohibited from depriving persons or corporations of property without due process of law.").

174 In addition, although not relevant here, some cases state that "property" is defined more broadly for purposes of the Deprivation Clause than for the Takings Clause. Corn v. City of Lauderdale Lakes, 95 F.3d 1066, 1075 (11th Cir. 1996) ("'Property' as used in the Just Compensation Clause is defined much more narrowly than in the due process clauses.").

175 GE v. Jackson, 610 F.3d 110 (D.C. Cir. 2010), *cert. denied*, 79 U.S.L.W. 3685 (U.S. June 6, 2011) ("The parties agree that the costs of compliance and the monetary fines and damages associated with noncompliance qualify as protected property interests."); Solid State Circuits, Inc. v. EPA, 812 F.2d 383, 391-92 (8th Cir. 1987); Wagner Seed Co. v. Daggett, 800 F.2d 310, 316 (2d Cir. 1986); Employers Ins. of Wausau v. Browner, 52 F.3d 656, 664 (7th Cir. 1995); *see also* Sackett v. EPA, 622 F.3d 1139 (9th Cir, 2010), *cert. granted*, 80 U.S.L.W. 3003 (U.S. June 28, 2011) (No. 10-1062).

176 Binyamin Appelbaum, *Investors Put Money on Lawsuits to Get Payouts*, N.Y. Times, Nov. 14, 2010, *available at* http://www.nytimes.com/2010/11/15/ business/15lawsuit.html.

177 Emery G. Lee III & Thomas E. Willging, Fed. Judicial Ctr., Litigation Costs in Civil Cases: Multivariate Analysis - Report to the Judicial Conference Advisory Committee on Civil Rules (2010); *see also* William H. J. Hubbard, Civil Justice Reform Group, Preliminary Report on the Preservation Costs Survey of Major Companies (2011) (Costs of litigation in some cases are significantly higher as costs exhibit a "long tail," meaning that a small fraction of cases account for most of the expenses associated with individual cases.).

178 *See, e.g.*, Hampton v. Mow Sun Wong, 426 U.S. 88, 100 (1976) ("The federal sovereign, like the States, must govern impartially. . . ."); *see also* Buckley v. Valeo, 424 U.S. 1, 93 (1976); Weinberger v. Wiesenfeld, 420 U.S. 636, 638 n.2 (1975) ("Equal protection analysis in the Fifth Amendment area is the same as that under the Fourteenth Amendment.").

179 *See, e.g.*, Mullane v. Cent. Hanover Bank & Trust Co., 339 U.S. 306 (1950).

180 Andrew H. Hershey, *Justice and Bureaucracy: The English Royal Writ and "1258,"* 113 Eng. Hist. Rev. 829, 838 (1998) (describing difficulties and expense of travel to court to complain or answer). At a later date, an alternative procedure called the *querela* developed in which someone could present their claims orally to four knights in a local country court, rather than travel to where the King and his Chancery clerks were present. *Id.* at 844-850.

181 395 U.S. 337 (1969) (Statute permitting prejudgment garnishment of wages without notice and prior hearing violates due process).

182 407 U.S. 67 (1972) (State replevin provisions that permitted vendors to have goods to be seized through an *ex parte* application to a court clerk and posting of a bond violates due process.).

183 It is well-settled that even preliminary and temporary deprivations of property require process appropriate to the circumstances. Niki Kuckesd, *Civil Due Process, Criminal Due Process*, 25 Yale L. & Pol'y Rev. 1, 12 (2006) ("Due process requires hearing procedures with respect to temporary or preliminary deprivations, as well as for those that are final and permanent.").

184 419 U.S. (1975) (invalidating an *ex parte* garnishment statute that failed to provide for notice and prior hearing or to require a bond, a detailed affidavit setting out the claim, the determination of a neutral magistrate, or a prompt post-deprivation hearing).

185 *Id.* at 606-608.

186 416 U.S. 600 (1974).

187 *Id.* at 615-618.

188 A cause of action for abuse of process generally requires proof of an ulterior motive or improper purpose. *See* Nat'l Ass'n of Prof'l Baseball Leagues, Inc. v. Very Minor Leagues, Inc., 223 F.3d 1143, 1152 (10th Cir. 2000) ("The elements of abuse of process are (1) the improper use of the court's process (2) *primarily* for an ulterior purpose (3) with resulting damage to the plaintiff asserting the misuse."); Vittands v. Sudduth, 730 N.E.2d 325, 332 (Mass. App. Ct. 2000) ("The essential elements of the tort of abuse of process are (1) process was used; (2) for an ulterior or illegitimate purpose; (3) resulting in damage.") (internal quotations omitted); *see also* 1 Am. Jur. 2d Abuse of Process § 6 ("ulterior motive or purpose generally required in an abuse of process action"); *id.* ("[M]ere ill will or spite toward the adverse party in a proceeding does not constitute an ulterior or improper motive, where the process is used only for the purpose for which it was designed and intended.").

In addition, in most jurisdictions, a suit for malicious prosecution requires not only proof of an improper purpose, but also subjective knowledge by the person filing suit that there were no reasonable grounds for suing. *See* 52 Am. Jur. 2d Malicious Prosecution § 1 ("A malicious prosecution may be briefly defined as one that is begun in malice and without probable cause to believe it can succeed, and that finally ends in failure."). Merely filing a lawsuit that the state would have determined on preliminary review to be insufficiently well-founded to require the person sued to answer would not necessarily be actionable either as an abuse of process nor a malicious prosecution. *See* Campbell v. City of San Antonio, 43 F.3d 973, 979 (5th Cir. 1995) (Malicious prosecution claims require a showing of malice, ill will, or improper purpose.). Nor do prevailing rules and practices for assessing costs require the person suing to reimburse those who were sued for their costs merely because the claims under suit turn out to be unfounded. *See generally* Charles A. Wright & Arthur R. Miller, 10 Federal Practice & Procedure: Civil § 2668 (3d ed., 2010).

189 501 U.S. 1 (1991).

190 "We agree with the Court of Appeals that the property interests that attachment affects are significant. For a property owner like Doehr, attachment ordinarily clouds title; impairs the ability to sell or otherwise alienate the property; taints any credit rating; reduces the chance of obtaining a home equity loan or additional mortgage; and can even place an existing mortgage in technical default where there is an insecurity clause. . . . [T]he State correctly points out that these effects do not amount to a complete, physical, or permanent deprivation of real property . . . ." *Doehr*, 501 U.S. at 11; *see also Doehr*, 501 U.S. at 26, 27 (Rehnquist, C.J., and Blackmun, J., concurring) ("In the present case, on the other hand, [unlike prior precedents] Connecticut's prejudgment attachment on real property statute, which secures an incipient lien for the plaintiff, does not deprive the defendant of the use or possession of the property.").

191 *Doehr*, 501 U.S.

192 *Id.* at 13.

193 424 U.S. 319, 335 (1976).

194 *Doehr*, 501 U.S. at 13-14 (both emphases supplied).

195 *Id.* at 14-15 ("Connecticut points out that the statute also [in addition to an *ex parte* judicial determination of probable cause] provides an 'expeditiou [s]' postattachment adversary hearing; notice for such a hearing; judicial review of an adverse decision; and a double damages action if the original suit is commenced without probable cause."(citations and footnotes omitted)).

196 *Id.* at 18-23.

197 That term should be grating to anyone graduating from any law school that teaches law and economics after about 1980, as Nobel Prize-winning economist Ronald Coase showed in a famous article long ago that costs do not naturally "belong" to either plaintiffs or defendants. Ronald Coase, *The Problem of Social Cost*, 3 J.L. & Econ. 1 (1960).

198 *See* Martin H. Redish & Colleen McNamara, *Back to the Future: Discovery Cost Allocation and Modern Procedural Theory*, 79 Geo. Wash. L. Rev. 773 (2011).

199 See *supra* text at note 64. *See also Doehr*, 501 U.S. at 9 (stating question as "what process must be afforded by a state statute enabling an individual to enlist the aid of the State to deprive another of his or her property by means of the prejudgment attachment or similar procedure").

200 See *supra* p. 115.

201 Carter v. Carter Coal Co., 298 U.S. 238, 310 (1936) (Fifth Amendment due process limits authority of federal government to delegate to other coal producers the power to fix wages and hours.); *see also* President George H.W. Bush, Statement on Signing the Negotiated Rulemaking Act of 1990 (Nov. 29, 1990), *available at* http://www.presidency.ucsb.edu/ws/index.php?pid=19115.

202 Seattle Title Trust Co. v. Roberge, 278 U.S. 116 (1928) (zoning variance only by consent of adjacent owners unconstitutional); Eubank v. City of Richmond, 226 U.S. 137 (1912) (setting of property line by adjacent owners unconstitutional).

203 Even if lawyers admitted to practice before a court are considered "officers of the court," they still have a financial interest in the decisions that they make. And note also that the power to require the clerk to issue a court order of summons is not limited to officers of the court, but may be exercised by any person, whether or not admitted to practice before the court.

204 Tumey v. Ohio, 273 U.S. 510 (1927) (Due process is violated if judge has a personal, direct and substantial financial interest in the outcome.); *see also* Caperton v. A. T. Massey Coal Co., Inc., 129 S. Ct. 2252 (2009) (Due process is violated by judge who received large campaign contributions from litigant.).

205 Fed. R. Civ. P. 11(b)(2) ("[T]he factual contentions have evidentiary support or, if specifically so identified, will likely have evidentiary support after a reasonable opportunity for further investigation or discovery.").

206 *Id.*

207 *See* Robert G. Bone, Twombly, *Pleading Rules, and the Regulation of Court Access*, 94 Iowa L. Rev. 873, 932-35 (2009).

208 *Cf.* Redish & McNamara, *supra* note 198 ("liken[ing] the discovery process to a quasi-contract, and argu[ing] that it is morally untenable to allow the requesting party to retain the benefit of its opponent's labor without, at the very least, reimbursing the costs of discovery incurred by the producing party").

209 George A. Akerloff & Paul M. Romer, *Looting: The Economic Underworld of Bankruptcy for Profit*, 2 Brookings Papers of Economic Activity 1 (1993).

210 *Compare* Young v. United States *ex rel.* Vuitton et Fils S.A., 481 U.S.

787, 802-09 (1987) (Counsel for a party that is the beneficiary of a court order may not be appointed to undertake criminal contempt prosecutions for alleged violations of that order.).

211 I am indebted to my sometime co-teacher Chief Justice Randy T. Shepard of the Indiana Supreme Court for pointing out to me that judges rarely see the costs that unfounded suits impose on others.

212 "A strike suit is a non-meritorious action brought to blackmail management into a settlement so that management can avoid the costly process of continued litigation, particularly the costs of discovery." Merritt B. Fox, *Required Disclosure and Corporate Governance*, 62 Law & Contemp. Probs. 113, 199 (Summer 1999).

213 William Landes & Richard Posner, *Adjudication as a Private Good*, 8 J. Legal Stud. 235 (1979) (describing relationship between size of "stakes" in litigation and settlement). It is an implication of Landes and Posner's famous formula that even a case with an expected value of zero on both sides will create settlement value in the form of a joint asset, the litigation costs that can be avoided by settling. Why defendants would be willing to pay settlement value rather than litigate to discourage future strike suits is a more complicated puzzle in game theory, but it too has been solved. *See* Bone, *supra* note 37, at 64.

214 David Sherwyn, Samuel Estreicher & Michael Heise, *Assessing the Case for Employment Arbitration: A New Path for Empirical Research*, 57 Stan. L. Rev. 1557, 1579 (2005) ("Because it costs employers (1) between $4000 and $10,000 to defend an EEOC charge, (2) at least $75,000 to take a case to summary judgment, and at least $125,000 and possibly about $500,000 to defend a case at trial, it almost always makes good business sense to settle a case for $4000."). Costs will vary, however, by geographic area of the country and type of case.

215 "[A]n arbitrage is a transaction that involves no negative cash flow at any probabilistic or temporal state and a positive cash flow in at least one state; in simple terms, a risk-free profit." Arbitrage, http://en.wikipedia.org/wiki/Arbitrage (last visited Dec. 14, 2011).

216 Some of the reasons that judges are reluctant to second-guess decisions by litigants are catalogued in Elliott, *Managerial Judging*, *supra* note 17.

217 *See* Redish & McNamara, *supra* note 198.

218 *See* Thomas K. McCraw, Prophets of Regulation: Charles Francis Adams; Louis D. Brandeis; James M. Landis; Alfred E. Kahn 224 (1984) ("Sound regulatory policy [Cornell economist and Carter Administration official Alfred Kahn] never tired of explaining, requires that buyers pay marginal cost of all the goods and services they receive. If five units of an item cost $40 to produce and six units costs $60, then the marginal cost of the item is not $8 or $10 but $20. If the sixth unit is priced at $10 (that is, at average cost), consumers will purchase too many units—often not just one too many, but several—and since consumers only have a certain amount of money to spend, they will be able to buy too few units of other items, relative to what they would do under allocative efficiency. When good and service are not priced according to marginal costs, therefore, consumers will automatically bring about a misallocation of society's resources. In order to prevent this unhappy result, Kahn believed, the prices of all goods and services should be set 'at the margin'—that is, they should be pegged to the cost of producing one more unit at a particular time.").

219 28 U.S.C. §1915(e)(2)(B).

220 490 U.S. 319 (1989).

221 Neitzke v. Williams, 490 U.S. 319, 326 (1989).

222 Conley v. Gibson, 355 U.S. 41, 47 (1957).

223 *Williams*, 490 U.S. at 326.

224 Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 561-63 (2007) (concluding that the *Conley* standard of "no set of facts can be proved" under the allegations of the complaint is "best forgotten").

225 *Williams*, 490 U.S. at 329-30 ("Under Rule 12(b)(6), a plaintiff with an arguable claim is ordinarily accorded notice of a pending motion to dismiss for failure to state a claim and an opportunity to amend the complaint before the motion is ruled upon. These procedures alert him to the legal theory underlying the defendant's challenge, and enable him meaningfully to respond by opposing the motion to dismiss on legal grounds or by clarifying his factual allegations so as to conform with the requirements of a valid legal cause of action. This adversarial process also crystallizes the pertinent issues and facilitates appellate review of a trial court dismissal by creating a more complete record of the case. . . . By contrast, the sua sponte dismissals permitted by, and frequently employed under, 1915(d), necessary though they may sometimes be to shield defendants from vexatious lawsuits, involve no such procedural protections. To conflate the standards of frivolousness and failure to state a claim, as petitioners urge, would thus deny indigent plaintiffs the practical protections against unwarranted dismissal generally accorded paying plaintiffs under the Federal Rules. A complaint like that filed by Williams under the Eighth Amendment, whose only defect was its failure to state a claim, will in all likelihood be dismissed *sua sponte*, whereas an identical complaint filed by a paying plaintiff will in all likelihood receive the considerable benefits of the adversary proceedings contemplated by the Federal Rules. Given Congress' goal of putting indigent plaintiffs on a similar footing with paying plaintiffs, petitioners' interpretation cannot reasonably be sustained. According opportunities for responsive pleadings to indigent litigants commensurate to the opportunities accorded similarly situated paying plaintiffs is all the more important because indigent plaintiffs so often proceed pro se and therefore may be less capable of formulating legally competent initial pleadings."). Congress evidently disagreed, however, as it subsequently amended 28 U.S.C. §1915 to add "failure to state a claim on which relief can be granted" as a ground for *sua sponte* dismissal in *in forma pauperis* cases.

226 490 U.S. at 318 n.8 ("We have no occasion to pass judgment, however, on the permissible scope, if any, of sua sponte dismissals under Rule 12(b)(6).").

227 490 U.S. at 318 n.6.

228 916 F.2d 725 (D.C. Cir. 1990).

229 916 F.2d at 726; *accord* Omar v. Sea-Land Service, Inc., 813 F.2d 986, 991 (9th Cir. 1987).

230 Frankos v. LaVallee, 535 F.2d 1346 (2d Cir. 1976); Perez v. Ortiz, 849 F.2d 793, 797-98 (2d Cir. 1988); Morrison v. Tomano, 755 F.2d 515, 516-17 (6th Cir. 1985); Jefferson Fourteenth Assocs. v. Wometco de P.R., Inc., 695 F.2d 524, 526-27 (11th Cir. 1983); *cf.* Literature, Inc. v. Quinn, 482 F.2d 372, 374 (1st Cir. 1973) (failure to give plaintiff prior notice "might well justify reversal," but reversed on other grounds).

231 See *infra* p. 128.

232 Charles Clark, *Special Problems in Drafting and Interpreting Procedural Codes and Rules*, 3 Vand. L. Rev. 493, 501 (1950) ("RULES AS EXPLAINING AS WELL AS GRANTING POWER . . . [W]ithout a tradition for the exercise of discretion, a general grant of power is likely to accomplish little. Habitually courts act according to precept and custom. If left to their own devices, without any precise guide beyond a general authorization, they will stick to what they have known in the past.").

233 John von Neumann & Oskar Morgenstern, Theory of Games and Economic Behavior 11 (1947), *quoted in* Garrett Hardin, *The Tragedy of the Commons*, 162 Science 1243 (1968).

234 *See* Fed. R. Civ. Pro. 11(b)(3) ("By presenting to the court a pleading, written motion, or other paper—whether by signing, filing, submitting, or later advocating it—an attorney or unrepresented party certifies that to the best of the person's knowledge, information, and belief, formed after an inquiry reasonable under the circumstances: (3) the factual contentions have evidentiary support *or, if specifically so identified, will likely have evidentiary support after a reasonable opportunity for further investigation or discovery . . . .*" (emphasis supplied)).

235 *Id.*

236 An optional invitation that provides an optional *opportunity* to be heard is different in a constitutional sense than an order that one must appear. See *supra* at pp. 118-120 (describing state systems of summons without a mandatory court order to appear).

237 *Cf.* Connecticut v. Doehr, 501 U.S. 1 (1991) (condemning relying solely on "one-sided, self-serving, and conclusory submissions").

238 *See generally* Redish & McNamara, *supra* note 198.

239 United States v. Morton Salt, 338 U.S. 632 (1950) (Court will enforce administrative subpoena if, but only if, "reasonable.").

A Modest Proposal for Human Limitations on Cyberdiscovery

*By Richard M. Esenberg**

Note from the Editor:

In December 2010, the Federalist Society heard from a number of federal judges and civil procedure experts about amendments to the Federal Rules of Civil Procedure, including the process that would be undertaken to amend the rules and some proposed amendments that might be offered. Based on the comments and perspectives received, the Federalist Society determined that it could add value to the broader discussion over amending the rules by asking experts to flag issues or perceived problems with the rules as they currently exist, and to identify the range of solutions that are being offered to address these problems. This back-and-forth culminated in four papers, one of which follows. A version of these papers will appear in the Florida Law Review, and they are published here with permission.

I. Introduction

I show my Civil Procedure students a video on electronically stored information ("ESI") created by Jason Baron and Ralph Losey.[1] The video, set to the type of pulsating electronic music normally heard prior to kickoff, sets forth a series of factoids about ESI. There will soon be more bytes of ESI than stars in the universe, it would take six million years to read each web page in the known universe, and we are awash in billions and trillions of e-mails, tweets, text messages and Google searches.[2] The video refers to studies showing that most of this information is never produced—and often not even thought of—in the discovery process.[3] It points out that the most common forms of retrieval, such as Boolean key word searches, find a relatively small percentage of "relevant" documents.[4] For Baron and Losey, the "near future" is that litigants cannot "afford the whole truth," but they suggest (with, I hope and suspect, tongues in cheek) that the "far future" is discovery conducted by artificial intelligence agents. The answer to the challenges of E-discovery, in other words, is the creation of E-lawyers.

The video is an engaging and well done representative of an emerging genre in the litigation literature which I prefer to call Electronic Gothic. It tends, unintentionally or otherwise, to frighten litigants and lawyers about the irresistible world of litigation holds, search protocols, document retention, preservation of records, recovery of lost materials, data mining, metadata, and iterative multi-phase discovery. The vehicle is often tales about sanctions for the loss or destruction of information that a party did not know it had, was (at least subjectively) unaware that is was obligated to keep, or had inadvertently deleted.

Law firms have formed E-discovery groups, and lawyers have fashioned careers as "E-discovery attorneys." One such lawyer admonishes law students to embrace their "inner geek," saying that, "if you did not go to law school to work with computers and data bases, then you might want to rethink being a litigator . . . ."[5] Another prominent expert on the discovery of ESI pointed out that lawyers tend to be drawn to the profession from a certain acuity in "liberal arts logical analysis"—i.e., the verbal and analytic skills that have traditionally been at the heart of the lawyerly craft. The profession, he suggested, needs to remake itself.

E-discovery is certainly here to stay because, absent a disaster that sends civilization to the stone ages, the digitalization of life is here to stay. The complexity of managing ESI in litigation is almost certain to grow as what we can create and where we can send it grows increasingly robust. While some of these advances may aid in the management of E-discovery, it seems a safe bet, as Losey and Baron suggest, that the location and production of ESI is going to get much harder before it gets appreciably easier. But, I want to suggest, the answer—even in the near term—is not to lament our inability to get at the "whole truth" and dream of robo-lawyers. Whether or not Losey and Baron are right in suggesting that we cannot afford "the whole truth," it is beyond doubt that ESI cannot be treated like paper in discovery.

But it is less obvious that much of the "truth" is really lost. The idea, undergirding much of discovery practice, that any information anywhere that might conceivably be helpful on any issue ought to be available for perusal is a notion that only lawyers could love. Other professions—doctors, design engineers, research scientists—have long had to accept the idea that a certain quantity of information will have to be "enough" and that one must somehow live with the ensuing uncertainty. Lawyers cannot now—and never have been able to—do otherwise. But something in the notion of open discovery, self interest (more discovery is more work), and the human fear of "missing something" [6] seem to have made lawyers peculiarly resistant to this idea.[7]

While the growth of ESI is irresistible, it faces an unmovable limiting principle. However voluminous and dynamic electronic information may become, human beings remain blissfully limited in their capacity to process information. As long as litigation remains concerned with the endeavors of mortals, the percentage of nonduplicative ESI that is in fact relevant to the "whole truth" is likely to remain rather limited. Aided by modern technology, human beings may come to create information that is dynamic and voluminous by increasingly committing all random thoughts to writings digitally retained. But only so much of it can ever be used. Consequently, it is likely not possible that all—or even a substantial part of it—will be relevant to whatever human activity has become the object of litigation and necessary for a fair and just resolution of the underlying controversy.

The development of E-discovery principles and rules have been an effort to balance cost against the value of the information utilizing the traditional tools of judicial management of

** President & General Counsel of the Wisconsin Institute for Law & Liberty and Adjunct Professor of Law at Marquette University Law School*

discovery, i.e., ad hoc and factually intensive balancing. This will continue to be necessary. But I want to suggest another paradigm. As ESI multiplies, organizations will have to find ways to retain and have access to that information which is necessary to conduct business, i.e., to sell and design things, to hire and fire people, and to do all the other things that happen in the real world and become the subject of litigation.

There ought to be, at minimum, a strong presumption that the retention and retrieval policies created to manage this information outside the litigation process are likely to catch almost all the information that is relevant within it.[8] Although this concept has found its way into the 2006 amendments to the Federal Rules and pertinent case law, there is more work to be done.

II. ESI Is Different

A. The Challenges of ESI

As noted above, the digitalization of life has threatened to overwhelm the process of relatively unfettered party directed discovery. The challenges presented by the discovery of electronically stored information may not be entirely "new," but they are "more."[9] The electronic revolution has resulted in a substantial—indeed geometric—increase in matters committed to writing. What may have been communicated by phone or in person or not communicated at all may now be expressed in e-mails, text messages, tweets, etc. Efforts to retrieve information or records of the transmission of these communications that, in the past, were unlikely to have even been created are now memorialized in the records of search engines and the "metadata" of information systems.[10] Human interactions and communications are now increasingly recorded somewhere. As two commentators recently observed:

> Information inflation reflects the fact that civilization has entered a new phase. Human beings are now integrated into reality quite differently than before. They can instantaneously write to millions. They engage in real time writing of instant messages, wikis, blogs and avatars. Consequently, the flux of writing has grown exponentially, with resulting impact on cultural evolution. All this affects litigation. Vast quantities of new writing forms challenge the legal profession to exercise novel skills.[11]

This is the temptation of E-discovery: the notion that "somewhere" in that mass of information "someone" may have written "something" that will be relevant to the issues in litigation.

Not only are more records created, they are far more likely to remain in existence not only "somewhere" but often in multiple places. The storage of electronic information, while expensive, is easier and less expensive than the retention of what have traditionally been much smaller quantities of paper records. These stored records can, moreover, often be searched electronically to identify some subset of at least potentially relevant materials. This, too, creates opportunities to find "something" that might advance a litigant's cause.

But there are other aspects of ESI that confound these opportunities. Electronic data is dynamic. It can be altered—sometimes automatically and unintentionally—through the normal operation of the system that created it. Because there is a cost—both in dollars and system efficiencies—to retaining it, it may be automatically deleted or "overwritten." While its deletion may not be irrevocable, it may make it relatively inaccessible, i.e., it can be recovered only at great cost and effort. An electronic document can, moreover, be repeatedly duplicated and transmitted to numerous recipients. Thus, it can be found in numerous "places"—not all of which are self-evident. The advent of "cloud computing" and applications like Google documents (or the simple fact that home computers may be put to business and professional use) raises the likelihood that certain documents may reside "out" of the responding organization.

We can go on. As the volume of information metastasizes, it overwhelms the capacity of human beings—and traditional electronic search methods—to review. ESI will generally have associated "metadata" that may provide information about when documents were created, altered, and transmitted. Deciphering that data—and even the documents themselves—may require an understanding—or even the use—of the system on which they were created. Because ESI may be automatically deleted or altered, the onset of litigation (or the apprehension of its potential) may require intervention to suspend those processes. Although notions of preserving relevant evidence—or sanctioning parties for spoliation—are not new, implementing these "litigation holds" is complicated and expensive,[12] requiring an understanding of just where diffuse forms of information can be found and predicting what may be relevant to litigation in which the claims and defenses may be nascent, ill-defined and imperfectly understood.

Finally, efforts to locate, preserve, and retrieve ESI are less transparent and straightforward than simply searching paper records. They require the application of expertise and can often result in complicated disputes about what can and cannot be readily obtained and lead to satellite litigation and "discovery about discovery." This substantially increases the cost of discovery management and disputes. It requires software, consultants, and, as noted earlier, attorneys specially versed in the nature of the game.

B. Responding to the Challenges

Of course, these problems have not gone unnoticed and unaddressed. In 2004, a group of prominent jurists, practitioners, and academics announced (and then subsequently revised) the Sedona Principles.[13] These fourteen principles seek to balance the need for discovery of ESI against its cost and unique challenges. They create a duty to preserve information but not one that requires a party to take "every conceivable step" or preserve "'deleted, shadowed fragmented or residual' information absent a showing of special need and relevance." In ordering discovery, courts should balance "cost, burden and need" considering the "nature of the litigation and amount of controversy." The primary (but apparently not exclusive) focus of E-discovery should be on "active data and information as opposed to disaster recovery back-up tapes and other sources that are not reasonably accessible." "Cost shifting" from the "responding" to the "requesting" party can happen on satisfaction of a multi-factor test. One commentator recently extolled the "enduring relevance" of the Principles.

In a now famous series of opinions in a case called *Zubulake v. UBS Warburg LLC*, Judge Shira A. Scheindlin attempted a similar balance in the context of employment litigation involving the preservation and production of a large volume of e-mails. The decisions, now a staple of most Civil Procedure textbooks, largely track the Principles, announcing a set of principles calling for a level of discovery and burden that is "just right." The *Zubulake* series repeated the now well-accepted notion that "[t]he universe of discoverable material has expanded exponentially and 'discovery is not just about uncovering the truth, but also about how much truth parties can afford to disinter.'"[14] The cases recognized a *seven* factor test for shifting the cost of discovery.[15]

But this cost shifting, at least in Judge Scheindlin's view, should not apply to "readily accessible ESI" to which the "normal rules of discovery" should apply.[16] The decisions made clear that a party must implement a "litigation hold" on ESI once it is on notice (i.e., knows or should know) that the information may be relevant to current or future litigation:

> As a general rule, once a party reasonably anticipates litigation, it must suspend its routine document retention or destruction policy and put in place a "litigation hold" to ensure the preservation of relevant documents. A litigation hold does not apply to inaccessible backup tapes, which may continue to be recycled on the schedule set forth in the company's policy. On the other hand, if backup tapes are accessible, then such tapes would likely be subject to the litigation hold.[17]

Under these circumstances:

> First, counsel must issue a "litigation hold" whenever litigation is reasonably anticipated. Counsel has a duty to remind employees that the litigation hold is still in effect. Second, counsel must communicate with "key players" and remind them of the duty to preserve. Further, counsel must become fully familiar with his client's document retention policies. Last, counsel must instruct his clients to produce all relevant data and ensure the evidence is stored in a safe place to avoid intentional or inadvertent destruction of potentially relevant data. Once counsel takes these steps, the client is fully on notice of its discovery obligations.[18]

Zubulake and cases like it are certainly helpful, but drawing lessons from reported decisions is difficult. The cases are fact-intensive, and the pertinent facts are highly technical. In a recent opinion by Judge Scheindlin, subtitled "Zubulake revisited," it takes eighty pages to describe the E-discovery malfeasance of the plaintiff.[19] A recent article describing E-discovery cases since 2006 offers relatively little guidance beyond the common place. Having read summaries of approximately 150 cases, one knows little that is new. That is not a criticism. It's not clear that more guidance is readily found.

In 2006, the Federal Rules of Civil Procedure were amended to address the problems presented by ESI. Pursuant to amended Rule 26(b), ESI need not be produced from sources that the responding party has identified as not reasonably accessible because of undue burden or cost subject to judicial review. Courts may limit discovery if it is unreasonably cumulative or duplicative or can be obtained from another source that is more convenient, less burdensome, or expensive. They may restrict discovery if the seeking party has had ample opportunity to obtain the information or if the burden or expense of proposed discovery outweighs "its likely benefit considering the needs of the case, the amount in controversy, the parties' resources, the importance of the issues at stake in the action and the importance of discovery in resolving the issues." By case law, although not rule, parties are required to take reasonable steps to preserve ESI when they "know or should know" of the potential to litigation for which it may be relevant. Federal Rule of Civil Procedure 37(e) provides that "absent extraordinary circumstances, a court may not impose sanctions when ESI is lost" as a result of the routine, good faith operation of an electronic information system."

C. The Inadequacy of the Response

All of this is eminently reasonable but, it would seem, not particularly effective. The standard for implementing a litigation hold is, for example, an invitation for an argument. It does little to define what must be held or how prescient the holding party must prove to be. Whether or not something is "reasonably accessible" is undefined, as is the "routine, good faith operation" of an information system. Of course, all legal standards are more or less underdetermined, but the vague nature of these standards may be more problematic in the context of discovery which is—and must largely remain—a process largely managed by the parties and one in which judicial intervention is difficult due to the nascent and ill-defined nature of the issues to be tried and the complex and technical nature of ESI.

Judges must assess such claims or evaluate the burden, need, and proportionality of proposed discovery with incomplete knowledge of the claims and defenses. The complexity of evaluating competing claims about the nature of the information sought and the cost of obtaining it may require the equivalent of a small (or not so small) trial—something difficult to do in the context of motion practice. The best way to avoid a premature (and perhaps incorrect) decision is often to err on the side of permitting discovery.

The standard of Rule 26(b)(2)(C)(iii)—the balancing of cost and burden proportionately—may be the principle by which most disputes regarding the scope of discovery are resolved.[20] That rule and its associated principles all contemplate relatively uncabined balancing of multiple factors. Nothing is foreclosed. Little is mandated. No relevant factors are excluded, but no particular result is mandated. While it is difficult to formulate specific legal rules that will do much more under complex circumstances, multi-faceted and ambiguous balancing in response to complicated and expensive questions provide little guidance. Standards that call for things that are "reasonable" and prohibit that which is "undue" are no better than admonitions to do "right." A seven part test for anything permits almost any result.

While some have endorsed—or at least accepted—the notion of judge as manager,[21] it is hard to imagine, given the volume of litigation, that discovery could work as anything other than a process that is largely party managed. There is simply too much litigation and cases and too much information

for judges of special masters to become involved in more than a fraction of cases. Management of the process by the parties works best if there are rules that effectively provide relatively clear direction or both sides have comparable incentives driving them within a realm of "reasonable behavior." In cases in which both parties are more or less equally subject to the costs and burdens of electronic discovery, each side can expect the other to be as aggressive or reasonable as it has been. This form of mutually assured destruction may discipline the parties and temper the discovery "arms race." But, in cases of asymmetrical information, i.e., those in which the bulk of information (particularly ESI) resides with one party, incentives diverge. Where the burden of responding to discovery is largely borne by one side, there are fewer incentives to self discipline.

Even when we do move to judicial management, judges must assess such claims or evaluate the burden, need, and proportionality of proposed discovery with incomplete knowledge of the claims and defenses. Although the rule requires parties to meet and confer and a mantra of the E-discovery industry is to call for "collaborative" discovery, parties famously disagree about the value of their cases and the extent of the burden that they are asking another to assume. However they agree on the principle of proportionality, that agreement is swamped by radically different perceptions of the amount at stake and the likelihood of recovery.

D. The Implications of Inadequacy

If the only implication of this were to increase the costs of discovery, it would be bad enough. But increasing the cost of litigation, particularly in the context of a system with at least some form of notice pleading, changes the dynamics of the litigation process and the calculus surrounding the management of litigation risk. The ability to assert a colorable claim, i.e., one that can survive a motion to dismiss and trigger the process of discovery, is an asset. Because it costs something—often quite a lot—to make such a claim go away—and litigation risk can rarely be dismissed—whatever increases the cost of the process increases the value of that asset. This materially alters the settlement calculus.

III. Another Response

A. A Modest Presumption

The rules ought to be amended to strengthen the presumption, begun with the 2006 amendments, that adherence to retention and retrieval policies adopted outside the context of litigation and consistently applied ought to be the measure of a party's obligation to maintain and produce ESI. The idea, not unrelated to Rule 34's longstanding option to produce records as they are kept in the ordinary course of business, is rooted in the idea that most organizations formulate such policies in good faith and, in fact, probably cannot know in advance whether the retention of information will "hurt" or "help" their litigation prospects. How much ESI to keep, where to keep it, and how to get at it are generally determined by the need to have access to information necessary to do business. Policies are presumably adopted in a way that will permit access to records that one needs to address the design and performance of products, the management of employees, and other aspects of the business that are likely to become the subject of litigation. If that is the case, most all relevant information will remain accessible under such generally applicable and neutrally framed policies.

To be sure, the current Federal Rules *permit* courts to limit E-discovery to documents resident in these systems, and, at least on its face, Fed. R. Civ. P. 26(b) creates a presumption against the discovery of ESI that is not reasonably accessible. But it may be well to make clear that, absent extraordinary circumstances, a party is required to produce only that ESI resident in the active systems maintained by it in the ordinary course of business. What I am suggesting is a bit of a paradigm shift. Perhaps we need be less concerned with whether the discovery of ESI falls beyond a pale of acceptable burden and cost and more concerned with whether the information sought can be found within a set of sources most likely to contain relevant records and can be accessed in a way that a party's normal records management system permits.

An example of such an approach is reflected in an amendment to Rule 26 proposed by certain defense bar organizations in a white paper presented in a recent conference on civil litigation at Duke University Law School, specifying that certain categories of ESI that are not available in the ordinary course of business need not be produced:

> (B) Specific Limitations on Electronically Stored Information.
>
> (i) A party need not provide discovery *of the following categories* of electronically stored information ~~from sources~~, *absent a showing by the receiving party of substantial need and good cause, subject to the proportionality assessment pursuant to Rule 26(b)(2)(C):*
>
> *(a) deleted, slack, fragmented, or other data only accessible by forensics;*
>
> *(b) random access memory (RAM), temp files, or other ephemeral data that are difficult to preserve without disabling the operating system;*
>
> *(c) on-line access data such as temporary internet files, history, cache, cookies, and the like;*
>
> *(d) data in metadata fields that are frequently updated automatically, such as last-opened dates;*
>
> *(e) information whose retrieval cannot be accomplished without substantial additional programming, or without transforming it into another form before search and retrieval can be achieved;*
>
> *(f) backup data that are substantially duplicative of data that are more accessible elsewhere;*
>
> *(g) physically damaged media;*
>
> *(h) legacy data remaining from obsolete systems that is unintelligible on successor systems; or*
>
> *(i) any other data that are not available to the producing party in the ordinary course of business* and that the party identifies as not reasonably accessible because of undue burden or cost *and that* on motion to compel discovery or for a protective order, if any, the party from whom discovery of such information

is sought shows is not reasonably accessible because of undue burden or cost.[22]

The proposed amendment provides additional guidance for both parties and courts and, importantly, roots that guidance in deference to systems established to conduct business. It retains current language requiring that, under certain circumstances, a party seeking to withhold information that might otherwise be discoverable must demonstrate that it is not reasonably accessible due to undue burden and cost. However, it makes clear that certain specified sources of information need not be searched or produced without such a showing, including information whose retrieval would require substantial additional programming or transformation or which cannot be obtained in the ordinary course of business. Although the proposed amendment does not unambiguously establish "active" ESI under a generally applicable retention policy as the entire universe for E-discovery, the recognition that most relevant documents are likely to be found within records retained and accessible under such policies informs its restrictions on the scope of discovery.

This will not obviate the need for litigation holds. The fact of litigation or its reasonable anticipation may affect the need to retain ESI, and parties ought to remain under an obligation to preserve potential ESI once litigation has been commenced or can be reasonably anticipated. An amendment proposed by the white paper delivered at Duke calls for parallel restrictions on the type of ESI that must be preserved, once again providing more particular guidance that reflects a judgment about where potentially relevant information is most likely to be found:

> (2) Specific Limitations on Electronically Stored Information
>
> Absent court order demonstrating that the requesting party has (1) a substantial need for discovery of the electronically stored information requested and (2) preservation is subject to the limitations of Rule 26(h)(1), a party need not preserve the following categories of electronically stored information:
>
> (A) deleted, slack, fragmented, or other data only accessible for forensics;
>
> (B) random access memory (RAM), temp files, or other ephemeral data that are difficult to preserve without disabling the operating system;
>
> (C) on-line access data such as temporary internet files, history, cache, cookies, and the like;
>
> (D) data in metadata fields that are frequently updated automatically, such as last-opened dates;
>
> (E) information whose retrieval cannot be accomplished without substantial additional programming, or without transferring it into another form before search and retrieval can be achieved;
>
> (F) backup data that are substantially duplicative of data that are more accessible elsewhere;
>
> (G) physically damaged media;
>
> (H) legacy data remaining from obsolete systems that is unintelligible on successor systems; or
>
> (I) any other data that are not available to the producing party in the ordinary course of business.

It is certainly possible that the exclusion of these sources of information from preservation and production will eliminate some information that might be relevant to litigation. It is less clear that they will render the results less accurate.

The amendments proposed at Duke also modify Rule 37(e) to make clear that sanctions may not be imposed for the failure to preserve ESI in the absence of a finding of willful conduct. This expansion of the rule's safe harbor provision furthers the emphasis on normally followed retention and retrieval procedures. The difficulty, however, is that sanctions for failure to preserve documents generally contain some presumption that the lost information would have helped the requesting party or hurt whomever has failed to produce it. But, in the absence of some finding of willfulness, that presumption is unwarranted. Although a responding party might certainly be required to restore the cost of recovering lost ESI, further sanctions as a consequence of negligence are problematic at least in the absence of some information about whether lost ESI would have helped or hurt the responding party.

B. Cost Allocation

The amendments proposed by the defense bar do some additional useful things such as limiting the number of document requests and the sources that can be searched.[23] Nevertheless, limitation of the universe of ESI that must be preserved and produced won't resolve all of the special challenges presented by E-discovery. Even active data systems maintained by parties in the ordinary course of business may produce enormous quantities of information. Presumably, parties will create methods of retrieving pertinent information for business purposes that balance the needs of that information with the cost of retrieval. Those systems ought to be treated as presumptively sufficient.

But most regularly maintained data bases are subject to some form of keyword or other electronic search that will, even without duplicates, result in mass quantities of information that will be exceeding expensive—or even stretch human capacities—to review. Perhaps the best solution to this problem is to place the cost of discovery with the requesting party. Internalization of externalized costs is generally thought to lead to greater rather than lesser efficiency. Perhaps the best way to ensure that the cost of discovery is proportional to what is at stake is to ask whether the party seeking it—the one who is presumably in the best position to know—is willing to pay for it.

A full consideration of this idea is beyond the scope of this paper. While this may be thought to burden the ability of less wealthy litigants to pursue a claim, the investment of substantial resources into litigation on behalf of nonwealthy parties thought by counsel to have a meritorious claim is quite common in a variety of contexts and has not materially impeded the pursuit of claims.

Although these costs would presumably be taxable upon resolution of the case on the merits, very few cases are resolved on the merits. To be sure, the fact that the cost of discovery is potentially taxable would affect the settlement calculus and indirectly discipline discovery. But a more direct impact would require these costs to be paid at the time that they are incurred. While this might lead to pretrial satellite litigation over the reasonableness of those costs, this seems more manageable and predictable than the more amorphous standards that currently control. It would involve the rather straightforward question of what undertaking a particular task has or will cost and not an assessment of whether, at some point in the future after underdeveloped issues become clear, it will have been "worth it."

IV. Conclusion

I close with a story from my young days as a lawyer. Rising to begin the introduction of my rebuttal case in a trial to the bench, the judge looked down at me and said, "Now Mr. Esenberg, you do what you need to do. But first ask yourself if anything you are about to do proves anything that hasn't been proven four times already, because I'm ready to rule." I sat down, learning an important lesson of trial advocacy: When to stop.

"When to stop" E-discovery is a difficult question. My modest suggestion is that lawyers take their cue from the ways in which such information is managed in the "real world." The electronic revolution has enabled many wonderful things, but, in litigation and elsewhere, we ought not to allow our desire for the perfect to become the enemy of the good.

Endnotes

1 Jason R. Baron and Ralph E. Losey, *E-discovery: Did You Know?, available at* http://www.youtube.com/watch?v=bWbJWcsPp1M&feature=player_embedded.

2 *Id.*

3 *Id.*

4 *Id.*

5 Alison A. Grounds, *Aspatore*, 2010 WL 3251514, at 14 (Aug. 2010).

6 Because ESI results in more human communication being recorded, it fuels the dream of the "smoking gun," i.e., the idea that, in a fit of candor, ill temper, or frustration, someone will write something that becomes "money" for the requesting party. It is unclear, however, that the ability to get at random thoughts that were previously unrecorded actually results in more accurate litigation outcomes.

7 *See, e.g.*, White Paper: Reshaping the Rules of Civil Procedure for the 21st Century: The Need for Clear, Concise and Meaningful Amendments to Key Rules of Civil Procedure 30 (submitted to the 2010 Conference on Civil Litigation, Duke Law School, May 10-11, 2010 on behalf of Lawyers for Civil Justice, DRI, Federation of Defense & Corporate Counsel, International Association of Defense Counsel) (May 2, 2010) ("Many attorneys believe that zealous advocacy requires extensive discovery.").

8 Some suggest that normal record management systems should be either driven by—or framed with—E-discovery in mind. *See, e.g.*, Steven C. Bennett, *Records Management: The Next Frontier in e-Discovery?*, 41 Tex. Tech. L. Rev. 519 (2009); Alison A. Grounds, *Evolving Technology and Strategies in the Area of e-Discovery*, Aspatore, 2010 WL 3251514, at * 12 (Aug. 19, 2010) ("I am seeing a trend where the makers of electronic records management systems are understanding that there needs to be an e-discovery component in their systems . . . ."); Capitol Records Inc. v. MP3 Tunes, LLC, 261 F.R.D. 44 (S.D.N.Y. 2009) ("The day undoubtedly will come when burden arguments based on a large organization's lack of internal ediscovery software will be received about as well as the contention that a party should be spared from retrieving paper documents because it had filed them sequentially, but in no apparent groupings, in an effort to avoid the added expense of file folders or indices."). My suggestion here is that it ought to be business necessity—and not the needs of litigation—that should drive records management.

9 One commentator notes that "E-discovery expenses of $3,000,000 in just five months are fairly commonplace." Ralph C. Losey, *Lawyers Behaving Badly: Understanding Unprofessional Conduct and Discovery*, 60 Mercer L. Rev. 983, 1000 (2002) (citing Kentucky Speedway, LLC v. Nascar, Inc., 2006 WL 5097354 (E.D. Ky. 2006)).

10 "Metadata" may identify who created the document, the date it was created and when it was opened or edited. Jessica DeBono, *Preventing and Reducing Costs and Burdens Associated with E-discovery, The 2006 Amendments to the Federal Rules of Civil Procedure*, 59 Mercer L. Rev. 963, 968 (2008).

11 George L. Paul & Jason R. Baron, *Information Inflation: Can the Legal System Adapt?* 13 Rich. J. L. & Tech. 10, 67-68 (2007).

12 One commentator describes the process as follows:

> A litigation hold consists of several components that must be implemented in a timely manner. The time element is extremely important when dealing with electronically stored information because such information can be destroyed or modified in the usual course of a company's business and a computer system's routine operations. A litigation hold must be customized to the anticipated litigation, depending on the nature and scope of the claims; however, a number of different procedures and records should be included in most cases. First, notice of the litigation hold should be provided to all relevant employees to preserve information. Second, a plan establishing how relevant electronically stored information will be retrieved and preserved must be created. Third, notice (and records of such notice) directing record custodians to suspend the destruction of relevant information should be maintained. Fourth, a record identifying what evidence has been preserved should be created. Fifth, monitoring procedures to ensure employees are utilizing the litigation hold should be implemented. Sixth, notification (and records of such notification) regarding the termination of the hold when litigation is no longer anticipated should be maintained.

DeBono, *supra* note 10, at 987-88 (footnotes omitted).

13 The Sedona Principles: Best Practices Recommendations & Principles for Addressing Electronic Document Production (2d ed., 2007), available at http://www.thesedonaconference.org; *id.* at Principle 12 (providing that "[u]nless it is material to resolving the dispute, there is no obligation to preserve and produce metadata absent agreement of the parties or order of the court").

14 Zubulake v. UBS Warburg, 217 FRD 309, 311 (S.D.N.Y. 2003) (*Zubulake I*).

15 *Zubulake I*, 217 F.R.D. at 320-24. The factors are: 1. The extent to which the request is specifically tailored to discover relevant information; 2. The availability of such information from other sources; 3. The total cost of production, compared to the amount in controversy; 4. The total cost of production, compared to the resources available to each party; 5. The relative ability of each party to control costs and its incentive to do so; 6. The importance of the issues at stake in the litigation; 7. The relative benefits to the parties obtaining the information. Not much is excluded.

16 Zubulake v. UBS Warburg, 216 F.R.D. 280 (S.D.N.Y. 2003) (*Zubulake III*).

17 Zubulake v. UBS Warburg, 220 F.R.D. 212, 218 (S.D.N.Y. 2003) (*Zubulake IV*). An exception applies if the company can identify employee documents that are stored on backup tapes. If that is the case, the tapes should be preserved if the information contained on those tapes is not otherwise available. *Id.*

18 Zubulake v. UBS Warburg, ___ F.R.D. ___, 433-34 (S.D.N.Y. 2004) (*Zubulake V*).

19 Pension Comm. of Univ. of Montreal Pension Plan v. Bank of Am. Secs.

LLC, 2010 WL 184312 (S.D.N.Y. Jan. 15, 2010).

20 Thomas Y. Allman, *The "Two-Tiered" Approach to E-Discovery: Has Rule 26(b)(2)(B) Fulfilled Its Promise?*, 14 Rich. J. L. & Tech. (2008).

21 *Symposium on Ethics and Professionalism in the Digital Age, Transcript—Morning Session*, 60 Mercer L. Rev. 863, 8967 (2009) (Judge notes that "the role of the judge is in the process of extraordinary transformation because of e-discovery.").

22 White Paper, *supra* note 7, at 25.

23 I am old enough to have been a seasoned litigator when courts began to limit—and rather arbitrarily at that—the number of interrogatories and both the number and length of depositions. How, we wondered, could the search for Truth be continued? It turns out we managed quite well.

Pleading, Discovery, and the Federal Rules: Exploring the Foundations of Modern Procedure

*By Martin H. Redish**

Note from the Editor:

In December 2010, the Federalist Society heard from a number of federal judges and civil procedure experts about amendments to the Federal Rules of Civil Procedure, including the process that would be undertaken to amend the rules and some proposed amendments that might be offered. Based on the comments and perspectives received, the Federalist Society determined that it could add value to the broader discussion over amending the rules by asking experts to flag issues or perceived problems with the rules as they currently exist, and to identify the range of solutions that are being offered to address these problems. This back-and-forth culminated in four papers, one of which follows. A version of these papers will appear in the Florida Law Review, and they are published here with permission.

Introduction

The Federal Rules of Civil Procedure are rapidly approaching their 75th birthday, which will come in the year 2013. 75 years is a long time, and while the Rules have of course been amended significantly at various points over the years, their basic structure remains largely the same as in their original formulation. When first promulgated in 1938, the Rules had an immediate and dramatic impact on civil adjudication by replacing long accepted procedural practices with very different methods of resolving disputes. There can be little question that the new system, spearheaded by the genius of Advisory Committee Reporter Charles Clark,[1] radically altered not only the actual procedures themselves, but also the underlying set of values that had previously rationalized our procedural system. The problem, right from the start, was that there was precious little articulation of either what the new value system was or why it was deemed preferable to the value structure underlying the old system.

To be sure, at the most basic level the stark differences between the two systems must have been obvious to all involved. In place of the draconian requirements of the demanding fact pleading standard, which required a plaintiff to know all of the circumstances surrounding his injury in detail at the time of the pleading, the new Federal Rules demanded considerably less at the pleading stage. The information that was unavailable at the pleading stage could now be gathered through a complex system of court-enforced discovery.[2] But exactly *why* this dramatic change was made was never fully clarified by any of the key actors. Thus, while it was clear that the change was premised in some sense on the notion that "[m]utual knowledge of all the relevant facts gathered by both parties is essential to proper litigation,"[3] the deep structure of the underlying value system was never satisfactorily articulated.

In part, this failure may have been due to the pressures imposed by narrow political considerations. In his scholarly work defending the new procedural system embodied in the Federal Rules, Judge Clark mystifyingly characterized the changes as merely the natural evolution of the preexisting process.[4] Yet that statement could not have been further from the truth. One can reasonably surmise that Judge Clark's characterization of the Rules' intended impact on existing procedural practices was largely a strategic effort to allay fears about the seemingly dramatic nature of the changes being adopted. However, it may also partially have been the result of the traditional failure of scholars to consider procedural issues from a "deep structural" perspective. By "deep structure" I refer to a synthesis of the fundamental social, moral, political and economic values which society seeks to foster in shaping its civil litigation process.[5] As a general matter, procedural scholarship focuses on what can be described as "second order" analysis, which refers to issues surrounding the shaping of specific procedural doctrines. Only rarely, however, have procedural scholars sought to tackle procedural questions as foundational as the intersection between procedure and democratic theory. This characterization is even more applicable to procedural scholarship at the time the Federal Rules were adopted, when legal scholars focused almost exclusively on narrow, even technical, issues of legal doctrine and analysis. Thus, although no one—including both those who agreed and those who disagreed with the changes brought about by the Federal Rules—could doubt the dramatic impact of Clark's revisions on our nation's sociopolitical and economic structure, it appears that absolutely no efforts were ever made at the time to view those changes through the lens of foundational political or economic theory.

This failure is truly unfortunate, since the choices made in shaping the Rules will necessarily impact our socio-economic and political structure, whether we are fully aware of that impact or not.

The rapid approach of the Rules' anniversary provides an appropriate opportunity to begin such a deep structural analysis with the benefit of almost 75 years of experience. The analytical inquiry appears to be timely for at least three additional reasons, as well. First, in two decisions over the last three years, *Bell Atlantic Corp. v. Twombly*[6] and *Ashcroft v. Iqbal*,[7] the Supreme Court caused an enormous stir—among judges, scholars and practitioners—over the proper pleading standard. Critics of these decisions (and there are almost too many to count) have mounted a variety of attacks on the Court's recent statements concerning the level of factual detail required in a complaint filed in federal court.[8] These pleading decisions have been criticized for improperly abandoning the notice-pleading standard embodied in Rule

** Louis and Harriet Ancel professor of Law and Public Policy, Northwestern University School of Law*

8(a) and the Court's famed decision of *Conley v. Gibson*,[9] for reintroducing the pre-Federal Rules "fact pleading" standard, and for improperly preventing plaintiffs from having their "day in court" as a means of vindicating their substantive rights.[10] On the other hand, scholarly defenses of the Court's decisions in *Twombly* and *Iqbal* have been relatively few and far between.[11] It would probably not be an overstatement to suggest that the combination of lower court confusion and intense scholarly controversy caused by two Supreme Court decisions concerning the Federal Rules over so short a time period is unprecedented.

The second reason that reconsideration of the theoretical foundations of our procedural system is timely is the elephant in the room that appears to have driven the Supreme Court's controversial pleading decisions: the Court's lingering concern over the serious burdens caused by the elaborate discovery process that represented the original Federal Rules' most significant innovation. Designed to enable litigants to gather the information necessary to facilitate accurate decision making and the effective vindication of substantive rights,[12] the discovery process has a dark side that seems to have been largely undervalued at the time of the Rules' framing. At least in an important category of litigation—those cases in which significant amounts of discovery are likely to take place—the costs and burdens inherent in the discovery process threaten to give rise both to serious inefficiencies in the adjudicatory process and to a potentially pathological and coercive skewing of the applicable substantive law being enforced.[13] The Court clearly reasoned in its recent pleading decisions that unless the pleading standards effectively perform some form of meaningful gatekeeping function, the harms caused by excessive and burdensome discovery could easily overwhelm the adjudication in much of modern high stakes litigation.[14] Yet even with the pleading standard performing this filtering function, the fact remains that in the cases that are allowed to proceed beyond the pleading stage, the burdens and costs of discovery are likely to continue to be substantial. The problems of excessive discovery, then, remain a significant concern.

The final reason that a reconsideration of the foundations of modern civil procedure is now timely is that both Congress[15] and the Rules Advisory Committee[16] are presently contemplating the possibility of major changes in the Rules. It could be disastrous if either the Committee or Congress were to alter the current adjudicatory structure without first exploring and articulating a coherent perception of the foundational political and socio-economic underpinnings of the procedural system they seek to fashion. The purpose of this Article is to begin that important undertaking.

In this Article, I first articulate my understanding of the basic value structure that is appropriately deemed to underlie our procedural system. In doing so, I seek to fashion the deep structure of modern procedure—what I refer to as "the litigation matrix."[17] In the following section, I consider how modern pleading standards need to be shaped in order to implement that matrix of underlying values most effectively. In so doing, I seek to explain why, despite some unfortunate and largely unnecessary confusion caused by the Court's opinions in *Twombly* and *Iqbal,* the "plausibility" approach the Court attempted to fashion in those decisions actually represents a wise balance of all of the competing and complementary underlying values.[18] In this Article, I will explain why, despite the torrent of criticism to which it has been subjected, the *Twombly-Iqbal* "plausibility" standard represents the fairest and most efficient resolution of the conflicting interests.

In Section III, I turn to an issue inextricably intertwined with the pleading controversy, the troublesome questions surrounding discovery reform. I believe that foundational precepts of economic, moral and political theory dictate a dramatic ex ante change in the structural operation of the discovery process which, if implemented, would undoubtedly reduce the costs and burdens of the process while preserving the bulk of its beneficial functions. That change, simply put, would be to recognize that the costs of discovery are, from the outset, properly attributed to the requesting party, rather than the responding party. Indeed, classic notions of quantum meruit—long recognized as an indisputable moral and legal dictate in the law of contracts—permit no other conclusion.[19] Were this alteration in the nature of the discovery process to be implemented, an immediate economic externality—one that currently plagues all discovery requests—would be removed. As a result, the discovery system would be relieved of most forms of even non-abusive "excessive" discovery requests—discovery that is simply not justified on the basis of a rational cost-benefit analysis.[20] It may also be necessary to consider imposition of direct structural limits on the nature and scope of discovery, though exactly how those limits should be framed will not be free from controversy. In the final section, I consider alternative ways the current Federal Rules could be amended in order to implement these insights.

I should emphasize that in shaping and applying the litigation matrix to the questions of pleading and discovery, I in no way intend to imply that either the factors to be included in that matrix or the manner in which they interact is free from debate or controversy. Nor do I intend to imply that even were we able to develop a consensus as to the abstract normative elements to be included within the matrix, determining how that matrix should apply to individual situations would always be free from controversy. The goal of this Article, rather, is merely to shift the nature of the ongoing debate about the nature and scope of the Federal Rules of Civil Procedure to an inquiry into the moral, economic and political factors that are properly deemed to provide the theoretical foundations of modern procedure.

I. Exploring the Deep Structure of Modern Procedure: Shaping the Litigation Matrix

There exists no officially recognized list of values which our procedural system is appropriately deemed to foster or achieve. Approximately a decade ago, however, I suggested what I considered to be a consensus grouping of broad normative goals that, when synthesized appropriately, make up the normative deep structure of modern procedural theory. "Some of these goals," I noted at the time, "are affirmative, goals the procedural system should accomplish. Others are negative, goals that attempt to limit the dangers to which the procedural system may give rise."[21] At that time, I included

six elements within my foundational litigation matrix: "(1) decision making accuracy; (2) adjudicatory efficiency; (3) political legitimacy; (4) maintenance of the substantive-procedural balance; (5) predictability; and (6) fundamental fairness."[22] Some 10 years later, I see no reason to depart from this broad framework as my normative theoretical anchor for a critical analysis of the modern procedural system.

On one level, it could be argued that decision making accuracy appears so obvious a consideration that it hardly requires explanation. It is nevertheless helpful to articulate the foundational rationales for inclusion of this factor. They are both systemic and individualist. From the perspective of the individual litigant (either plaintiff or defendant) whose rights have been substantively distorted because of inaccuracy in the fact-finding in her individual suit, the procedural system loses most or all of its political legitimacy. From a systemic perspective, absent decision making accuracy in the individual case it is impossible to ensure implementation and vindication of the substantive legal framework enacted by the democratic society. As modes of implementing its chosen framework, governing authorities often vest substantive rights in private individuals.[23] Without the accurate finding of facts in the individual suit, it is highly likely that the substantive law will either be over-enforced (in cases in which the facts are found incorrectly against the defendant) or under-enforced (in cases in which the facts are found incorrectly against the plaintiff). Either way, the underlying substantive law has been subverted by the procedural system.

It is also necessary, however, to recognize the existence of important competing socioeconomic goals. In certain contexts, the unbending pursuit of factual accuracy will give rise to prohibitive costs which, in turn, will lead to an impact far more harmful than beneficial on the rest of society. These costs may be conceptualized as either "internal" or "external."[24] The former category includes the burdens to which the truth-finding process gives rise directly—i.e., the costs and burdens suffered by the litigants as a result of the devices and structures adopted to achieve fact-finding accuracy. In shaping its procedural due process standards, the Supreme Court has long recognized the need for some utilitarian balance that needs to be struck between the search for truth and the financial and structural burdens required to attain the truth.[25] Even where the burdens are not so great as to reach the level of a constitutional violation, however, purely as a matter of social policy those structuring the procedural system must take this concern into account in shaping the devices designed to lead to the finding of truth. At some point, it will simply be too costly for all concerned to take every conceivable step toward truth finding.

In contrast to these "internal" costs, "external" costs include losses incurred by society beyond the scope of the individualized adjudicatory process. In certain situations, use of a particular procedure designed to attain factual accuracy will impose costs that extend beyond the four walls of the courtroom, thereby undermining substantive interests which society has chosen to protect. Evidentiary testimonial privileges provide a perfect illustration of situations in which the system has made a choice to value external interests over truth-finding.

These external costs link the goal of utilitarian limits on the truth-finding process to the fourth foundational goal of a procedural system: maintenance of the substantive-procedural balance. This factor is premised on the recognition that procedural rules will often have an inescapable collateral impact on interests that exist well beyond the walls of the courthouse—in other words, on the substantive concerns and goals society has sought to foster and implement. It would be a serious mistake to ignore these collateral impacts, because they will occur whether or not we acknowledge their existence. How a society shapes its procedural system will inevitably risk over- or under-enforcing its substantive law. Moreover, such "back door" procedural alteration of governing substantive law is especially invidious in a democratic society because it constitutes a change in the DNA of applicable substantive law, not through the transparent democratic processes of legislative modification (with all of the attendant controls of democratic representation and accountability) but rather through furtive or indirect means. Thus, though it may not always be possible to achieve the proper balance, the goal in fashioning procedural rules must be, to the extent reasonably possible, neither to over- or under-enforce substantive law as a result.

While the substantive-procedural interaction focuses primarily on the shaping of citizens' "primary conduct"[26] (i.e., their non-litigation behavior), the foundational goal of predictability focuses on the concern that litigants receive clear and consistent messages as to how they should prepare for litigation. To avoid this uncertainty, it is necessary, to the extent feasible, to establish governing rules that provide litigants with a clear understanding of what is expected from them. It is important to comprehend, however, that it would be counterproductive to adopt, in the name of predictability, narrow, mechanistic rules of procedure. The most one can reasonably expect is for generally framed rules to provide broad guidance that will be fleshed out through the inductive process of case-by-case development.

While the final element in the litigation matrix, fundamental fairness, is essential to the foundational analysis, it is simultaneously fraught with danger. On the one hand, any system that fails to include at least some concern for dictates of fundamental fairness in the procedural system's treatment of the litigants ignores important elements of the social contract between government and citizen that is necessarily implicit in a democratic society. On the other hand, the concept of fundamental fairness is, at least at its outer limits, so vague and malleable that it can far too easily be invoked conclusorily for manipulative or hidden purposes. Thus, to the extent the foundational moral calculus underlying modern procedure includes considerations of fundamental fairness the inquiry must be confined to situations where one is able to articulate specific and logically defensible inferences from explicit normative premises.

II. Pleading and the Litigation Matrix

A. Pleading and the Risk of the Wrong Guess

With the promulgation of the Federal Rules of Civil Procedure in 1938, the federal judicial system dramatically altered the prevailing theory of pleading, and in doing

so significantly modified the nature of the relationship between procedure and the substantive law it is created to implement. It has been thought by many, however, that the Supreme Court in its 2007 decision in *Bell Atlantic Corp. v. Twombly*[27] substantially reinterpreted and restructured the pleading requirements that had been included in the original Federal Rules in ways that dangerously undermined the core philosophical precepts underlying those Rules.[28] The Court followed its decision in *Twombly* two years later in *Ashcroft v. Iqbal*,[29] and once again many considered the decision to be inconsistent with the original Rules.

There is little doubt that a procedural system's chosen pleading standard can have a significant impact on the implementation of underlying substantive law. At one extreme, pleading standards that require a plaintiff to supply detailed facts about defendants' illegal behavior at a point in the process at which it would be difficult for the plaintiff to know that information could result in serious under-enforcement of substantive rights and proscriptions; legitimate suits would be filtered out at an early stage of the process. At the other extreme, overly lax pleading standards that enable a plaintiff to get past the pleading stage asserting nothing more than vague and unsupported legal conclusions could invite so-called "strike suits," frivolous claims brought solely to coerce defendants into making unjustified settlements to avoid the burdens and costs of the discovery process.

In choosing a generally-applicable pleading standard, it is difficult to walk this procedural tightrope. Whichever pleading standard is ultimately adopted, there will always exist a serious risk that in a significant percentage of cases the result would either be over- or under-deterrence of substantively proscribed behavior. Either result would upset the delicate balance between substance and procedure that is central to the smooth functioning of a constitutional democracy. The question then becomes, on which side of the equation are we willing to risk being wrong? We have seen such a form of weighing in other legal contexts. For example, the criminal system has made the categorical ex ante judgment that we would prefer to let a guilty person go free rather than send an innocent person to prison.[30] In the pleading context, the task is to fashion a workable standard under which the risks are allocated in a manner that optimizes the symbiotic interaction between procedure and the substantive law it is designed to enforce. I refer to this effort as a search for the party on whom to impose "the risk of the wrong guess." In the pleading context, where the court of course lacks perfect knowledge of the facts, the question at the time of the motion to dismiss is to determine whether it is likely more fair and efficient to risk dismissing a deserving plaintiff on the one hand or imposing the burdens of the pre-trial process on a defendant who would ultimately prevail on the merits.

On the civil side, whether one chooses a pleading system that risks pushing deserving plaintiffs out of court prematurely or instead selects a system that risks over-deterrence of defendant behavior (as well as the resulting internal and external economic inefficiencies) depends on certain foundational substantive assumptions about economic and political theory. If one begins with a strong presumption in favor of the value of wealth redistribution and an overriding concern that laws regulating corporate or governmental behavior be enforced, then one is likely to choose a pleading system that demands less of plaintiffs, thereby placing a risk of over-enforcement on defendants. If, on the other hand, one were to begin with an overriding substantive concern about the costs and harms of over-deterrence and the possible waste of litigation resources and believe that courts should not transfer wealth absent a strong and clear reason to do so, then we are far more likely to adopt a more demanding pleading standard. Such a standard would place the risk of deciding incorrectly more on the plaintiffs who are seeking to enforce the law.[31]

Throughout its history, the nation has made very different choices about which party should bear the risk of the wrong guess at the pleading stage. In the following section, I explore these alternatives and the shifts from one to another presumption at different points in the nation's history. In so doing, I will explore the inherent intersection between the pleading standard and the enforcement of controlling substantive law.

B. The Evolution of Pleading in the Federal Courts

1. The Shift from Fact Pleading

Prior to the adoption of the Federal Rules in 1938, the generally accepted pleading standard was "code pleading," named because of its origins in the reform statutory codes of the nineteenth century, particularly the Field Code in New York, which had been designed to replace the common law writ system. It was adopted in an effort to democratize the litigation system by making it more understandable and therefore more accessible to the common person.[32] Instead of focusing on the conceptual niceties of legal pigeonholing that had characterized common law pleading, the codes shunned the pleading of legal conclusions in favor of an intensive emphasis on the need for detailed facts.[33] Demurrers to the face of complaints on grounds of a lack of factual specificity were commonplace, and as a result the pleading stage played a significant role in the litigation process. Not surprisingly, this focus on factually detailed allegations often made it difficult for plaintiffs to proceed past the pleading stage, since at the outset of the case they often lacked access to key information concerning defendant's specific behavior, that was not readily available to them or under the control of the defendant.

Under the intellectual leadership of Charles Clark, the Federal Rules dramatically altered the prevailing pleading dynamic.[34] Instead of demanding facts that stated a cause of action, the Rules now demanded only that the pleadings provide "a short and plain statement of the claim showing that the pleader is entitled to relief . . . ."[35] Under this system, the motion to dismiss was to play a far smaller role than had the demurrer in code pleading jurisdictions.[36] Instead, the plaintiff was to have access to an array of elaborate discovery devices,[37] enforceable by the court,[38] to enable him to acquire the information needed to pursue the case to trial. The only exceptions to this substantially softened pleading standard were cases of fraud and mistake, which, pursuant to Rule 9(b), remained subject to fact pleading requirements.[39]

As opponents of the Rules were quick to point out, the obvious dangers in this system were the invitation to meritless suits brought solely for purposes of seeking coercive settlements or engaging in fishing expeditions. Elaborate discovery devices often require substantial investments of time, effort and money on the part of litigants. Once the motion to dismiss is effectively eliminated as a filter, there is nothing to stop plaintiffs from initiating the process and quickly obtaining access to potentially burdensome and expensive discovery.[40] But whatever the legitimacy of the concern was at the time of the Rules' adoption, with the development of modern products liability law and class action procedure, in at least a certain category of complex cases the problem of discovery abuse has evolved into a real danger.[41] While the Rules drafters over the years have undertaken a number of significant and often controversial measures to reduce the frequency of such abuse,[42] in *Twombly* Justice Souter pointed out that their success had been, to say the least, less than consistent.[43]

2. Understanding the Pleading Standard of the Federal Rules

The task facing both the drafters of the Rules and the courts asked to interpret and enforce them is to devise a method that, to the extent feasible at the outset of a litigation, imposes the risk of the wrong guess on the party most likely (as best we can predict at the pleading stage) to be arguing the factually incorrect position. In this way, we will reduce the costs of over- or under-deterrence as best we can. Thus, where a complaint alleges non-conclusory facts which, if true, make the court believe that the complaint "plausibly" alleges a valid claim—i.e., there is a reasonable likelihood that a legally cognizable wrong has been committed—it is appropriate to permit the complaint to proceed, even though the court or jury may ultimately determine that no wrong was actually committed. On the other hand, where no reasonable basis exists, on the face of the complaint's factual allegations, to plausibly suspect that a legal wrong has been committed, the risk of under-deterrence of the substantive law must be placed on the plaintiff. To be sure, the difference between these two will not always amount to the difference between night and day. There will no doubt be many close cases. But that difficulty rarely disqualifies a legal standard, nor should it here.

It is important to note that use of this standard should not be considered either a doctrinal innovation or a departure from the drafters' intent underlying Rule 8(a) when it was originally adopted in 1938. The so-called notice pleading system, when properly construed, should not—nor, I believe, was it ever intended to—serve as an "Open, Sesame" to plaintiffs seeking to engage in the equivalent of legalized blackmail or to conduct fishing expeditions through the wasteful and inefficient use of the discovery process. Indeed, anyone who would reject this "plausibility" standard[44] as overly restrictive and under-protective of a plaintiff's substantive and procedural rights should be required to articulate the elements of the less demanding standard with which they would replace it. The only conceivably less restrictive alternative is a standard that would permit a plaintiff merely to allege, in the most vague and conclusory manner, that a defendant had committed a violation of law. While presumably the plaintiff would need to assert violation of a specific right, that requirement hardly provides either the defendant or the system with meaningful protection against waste or abuse (both internal and external) due to the delay and burdens of what turns out to have been wasted discovery. It is simply too easy for a plaintiff to camouflage a total absence of any real basis for suit under a conclusory allegation of law violation. The realistic alternative to a standard grounded in an assessment of a complaint's plausibility, then, is not this substantially less demanding version of notice pleading (what can be appropriately described as "notice pleading minus"); use of such a standard would amount to the imposition of no standard at all and an invitation to procedural chaos. The only even arguably viable alternative to an approach grounded in reasonable suspicion is therefore the even *more* demanding fact pleading standard of the pre-Federal Rules days—a standard the Rules' drafters wisely rejected in all but the narrowest category of exceptions.[45]

It is important to understand that this plausibility standard (which can properly be viewed as a "notice pleading plus" standard) significantly differs from the considerably more demanding fact pleading standard employed in both the pre-Federal Rules codes and currently in Rule 9(b) for allegations of fraud or mistake.[46] This can be conclusively demonstrated by hypothetically applying both standards to the important post-Federal Rules pleading decision, *Conley v. Gibson*.[47] There the Supreme Court overturned a dismissal of the complaint in a suit by African-American union members who accused their union of conspiracy with their employer to engage in racial discrimination, in violation of applicable federal labor laws. Though the complaint included no specific or direct factual allegations describing the nature of the alleged discriminatory conspiracy, it did allege that the railroad for which they had worked abolished 45 jobs held by African-Americans and secretly filled all those jobs with whites. It further alleged that despite repeated pleas, "the Union, acting according to plan, did nothing to protect them against these discriminatory discharges and refused to give them protection comparable to that given white employees."[48] If the allegation that the plaintiffs' union made no efforts on plaintiffs' behalf despite the fact that they all had been replaced by white workers was not in and of itself sufficient to make a reasonable observer suspect of defendant's behavior, the complaint also alleged a history of past discriminatory acts on the part of the union.[49]

Who could reasonably dispute that the *Conley* plaintiffs had alleged far more than enough to make a reasonable observer conclude that unlawful behavior on the part of the defendants had been plausibly alleged? To be sure, it may turn out that proof at trial of the truth of the complaint's non-conclusory factual allegations, standing alone, would not have amounted to evidence sufficient to reach a jury. But that is not the question that the plausibility standard should be deemed to ask at the pleading stage. Rather, plausibility demands only that the complaint's non-conclusory factual allegations make a reasonable observer believe that the defendant likely violated plaintiff's rights and that discovery might well reveal confirming evidence of that fact. But if the adoption of the Federal Rules' revised pleading standard altered the pre-

existing fact pleading standard in any meaningful way, surely the complaint in *Conley* must be found to have alleged enough to allow plaintiffs to invoke the Federal Rules' discovery devices in search of the evidence they would need at trial.

In striking contrast to the plausibility standard, fact pleading requires the allegation of substantial factual detail in describing defendant's unlawful behavior: who did what, with or to whom, and when they did it.[50] In a fact pleading system the plaintiff is expected to know, prior to filing suit, exactly what happened. For example, under a fact pleading regime plaintiffs would not be allowed simply to allege, in a conclusory manner, that their union had conspired to discriminate against them, as was basically true of the complaint in *Conley*.[51] Rather, under a fact pleading regime the plaintiffs would have had to allege specifically at what point the union had conspired and with whom, and elaborate on the detailed nature of the conspiracy—something most plaintiffs who had been the victims of a conspiracy would be unable to do without access to discovery, even if they had suffered its consequences.

The plausibility standard, in contrast, does not demand that the plaintiffs possess knowledge of facts which they could not reasonably be expected to know at the litigation's outset. Rather, it demands merely that the description of the facts plaintiffs do know—i.e., the events that plaintiff knows to have taken place—give rise to the plausible claim that what took place resulted from unlawful behavior.[52] Thus, under the plausibility standard, in certain situations the plaintiff may still be permitted to plead in terms of legal conclusions, something that is foreign to a fact pleading system. For example, under a plausibility standard a plaintiff may plead using such legally-conclusory terms as "conspiracy," or "negligence," without explaining in detail exactly how the defendants' behavior qualifies for such descriptions, as long as the plaintiff's description of what consequences he suffered or of the manner in which the surrounding situation has been altered by defendant's actions is reasonably suggestive of unlawful behavior. Plaintiffs will be permitted to rely on conclusory allegations where it appears doubtful that the situation described factually in the complaint would have taken place absent some departure from the legally-required norm. Thus, while the *Conley* complaint survives under a plausibility standard, it fails the far more demanding fact pleading standard.

Properly understood, the plausibility standard asks merely whether the allegations contained in the complaint describe a situation that on its face gives rise to a finding of sufficient suspicion of unlawful behavior by defendant to justify taking the case to the discovery stage. The inquiry can be thought of in terms of "risk-reward": the more suspicious the circumstances alleged, the more likely it is that the risks associated with incurring the costs of discovery will be justified, because the more likely it is that use of the discovery process will bear fruit. The plausibility standard, then, is simply a matter of playing the odds as best they can be assessed with the limited knowledge the court possesses at the point at which a complaint is filed. Thus, the inquiry a court is to make under an approach grounded in an inquiry into the plausibility of the complaint's allegations of unlawful behavior differs significantly in its expectations of what the plaintiff must provide at the pleading stage from its expectations of what the plaintiff must be able to prove at trial.

A standard grounded in an effort to ascertain plausibility at the pleading stage is fully justified by the socio-political values that make up the underlying litigation matrix.[53] The approach strikes an appropriate balance between competing interests. Any standard less demanding would be far too lax in allowing plaintiffs with questionable claims to proceed to discovery, with all of its accompanying inefficiencies and undue burdens. Similarly, a more factually-demanding standard would, in most cases, risk skewing the substantive-procedural balance in the opposite direction.

It is certainly true that under a plausibility standard erroneous dismissal of a certain number of meritorious suits will occur. Judges are human and therefore fallible; at this early stage of the litigation, with an absence of complete information, even educated guesses still remain, at some level, guesses. Thus there will always exist the risk that pleading requirements will, in an individual case, under-enforce the underlying substantive law. However, the Rules' adoption of the plausibility standard represents the logical outgrowth of the common sense conclusion that we should be willing to risk a certain degree of under-enforcement. Incurring such a risk is necessary to avoid the burdens and inefficiencies that would be caused by the significantly greater amount of over-enforcement that would flow from a less demanding pleading standard.

Plausibility is thus far more consistent with a "notice pleading" standard than it is with a fact pleading standard. However, it is appropriately distinguished from a standard that demands nothing more from a plaintiff than a wholly unsupported, conclusory allegation of a legal wrong. Both could, I suppose, be described as "notice pleading," but plausibility is properly labeled "notice pleading plus," while the absurdly lax standard is appropriately described as "notice pleading minus."

C. Explaining the Supreme Court's Recent Pleading Decisions

1. *Twombly*

Twombly was the first decision to expressly articulate Rule 8(a)'s pleading standard in terms of plausibility. However, as we shall see, the standard is consistent with the holdings of all major pleading precedents.

The case involved an allegation of a conspiracy in violation of Section I of the Sherman Act. The Supreme Court has long made clear that in order to violate the Sherman Act's prohibition of "contracts, combinations or conspiracies in restraint of trade," the defendants must have actually conspired; mere "conscious parallelism," whereby the defendants intentionally act in a parallel manner absent any communication among them, is not actionable.[54] Of course, a pattern of parallel behavior is certainly consistent with the existence of an actual conspiracy, so the question arises whether an allegation of consciously parallel behavior combined with a conclusory assertion of conspiracy suffices to satisfy Rule 8(a). In *Twombly*, the Court considered whether a complaint that conclusorily alleged the existence of an actual

controversy on the basis of defendant's parallel anti-competitive behavior can defeat a motion to dismiss. In holding that an antitrust complaint failed to satisfy the requirements of Rule 8(a), the Court reasoned that "[w]hile a complaint attacked by a . . . motion to dismiss does not need detailed factual allegations . . . a plaintiff's obligation to provide the 'grounds' of his 'entitle[ment] to relief' requires more than labels and conclusions, and a formulaic recitation of the elements of a cause of action will not do. . . . Factual allegations must be enough to raise a right to relief above the speculative level . . . ."[55] To satisfy the notice pleading standard imposed by Rule 8(a) (the "plus" version of that standard, it should be noted), the Court concluded, a claim under section 1 of the Sherman Act must provide "enough factual matter (taken as true) to suggest that an agreement was made." Justice Souter, speaking for the Court, emphasized that "[a]sking for plausible grounds to infer an agreement does not impose a probability requirement at the pleading stage; it simply calls for enough fact to raise a reasonable expectation that discovery will reveal evidence of illegal agreement."[56] The Court drew a distinction between allegations "plausibly suggesting" unlawful behavior on the one hand, and those "merely consistent with" such behavior on the other. The former satisfy pleading requirements; the latter do not.[57]

Applying its plausibility standard to the facts alleged in Twombly's complaint, the Court found that "without some further factual enhancement [beyond the mere assertion of parallel conduct by defendants], [the complaint] stops short of the line between possibility and plausibility of 'entitle[ment] to relief.'"[58] This was because "nothing contained in the complaint invests either the action or inaction [on the part of the defendants] alleged with a plausible suggestion of conspiracy."[59]

2. *Iqbal*

Plaintiff in *Iqbal*, a Muslim and a citizen of Pakistan, was arrested on criminal charges by federal officials after the attacks of September 11, 2001. He alleged that he had been arrested and abused while in custody as part of a sweeping policy established by defendants Ashcroft and Mueller—at the time, respectively Attorney General and Director of the Federal Bureau of Investigation—to detain Muslims such as plaintiff in highly-restrictive conditions, for no reason other than their religion.[60]

In dismissing the complaint, the Court applied *Twombly*'s "plausibility" standard. "A claim has facial plausibility," wrote Justice Kennedy on behalf of the majority, "when the plaintiff pleads factual content that allows the court to draw the reasonable inference that the defendant is liable for the misconduct alleged."[61] The Court added: "The plausibility standard . . . asks for more than a sheer possibility that a defendant has acted unlawfully. . . . Where a complaint pleads facts that are 'merely consistent with' a defendant's liability, it stops short of the line between possibility and plausibility of 'entitlement to relief.'"[62] "Threadbare recitals of the elements of a cause of action, supported by mere conclusory statements," Justice Kennedy noted, "do not suffice."[63]

Applying these dictates to plaintiff's complaint, the Court found the allegations wanting. His claims, Justice Kennedy concluded, are "bare assertions" that "amount to nothing more than a 'formulaic recitation of the elements' of a constitutional discrimination claim."[64] Plaintiff's allegations provided no basis, Justice Kennedy reasoned, on which to surmise that either Ashcroft or Mueller had been a part of any scheme against Muslim men on the basis of nothing more than their religion.

D. *Explaining* Twombly *and* Iqbal

As previously noted, the large majority of scholarly commentary on both of these decisions has been mercilessly critical.[65] The view of many commentators is that in both *Twombly* and *Iqbal,* the Court abandoned the salutary goals of the notice pleading system adopted in the original Federal Rules in 1938. Despite all the critical commentary surrounding the Court's opinions, however, a closer look reveals that in *Twombly* and *Iqbal* the Court in reality did nothing more than impose the pleading standard that should be deemed to have been in force since the original adoption of Rule 8(a) in 1938. To be sure, its use of the label, "plausibility," was new. The substance of the standard, however, was not. The key advance in these decisions was that while the governing standard had always been plagued by ambiguity as to exactly how lenient its demands of factual detail actually were, after *Twombly* and *Iqbal* all uncertainty was removed.

The manner in which all of the Court's pleading decisions may be reconciled is by understanding Rule 8(a)'s pleading standard as the imposition of a requirement consistent with the "risk-of-the-wrong guess" analysis described previously.[66] Pursuant to that analysis, the complaint must allege facts sufficient to justify the imposition on defendant of the risk of a mistaken decision on its motion to dismiss.

To satisfy this standard, the allegations must amount to more than simply the unsupported and conclusory assertion of law violation. But that does not mean that a plaintiff's reliance on the pleading of a legal conclusion, in and of itself, will automatically lead to a complaint's dismissal. The issue is far more complicated than such an all-or-nothing approach would suggest. In certain situations, a complaint's allegations do not necessarily have to include claims of specific facts concerning the commission of unlawful acts on the part of the defendant. Rather, in a manner conceptually analogous to the evidentiary doctrine of *res ipsa loquitur* at trial, it is conceivable that a description of nothing more than the circumstances, as plaintiff knows them to be at the time of the filing of the complaint, could permit an objective observer to reasonably suspect that unlawful behavior might have occurred. The observer could reach this conclusion by reasoning that the results described in the complaint are unlikely to have occurred absent unlawful behavior, and that discovery could provide evidentiary support for the complaint's allegations.

While the Court's opinions in *Twombly* and *Iqbal* may at some levels be susceptible to confusing and inconsistent misinterpretation, when properly understood those decisions should actually reduce, rather than increase, doctrinal

confusion. After *Twombly* and *Iqbal,* complaints lacking specific detail should be deemed sufficient to allow the pleader to proceed to discovery when and only when they allege nonconclusory facts which render the allegation of legal wrongdoing "plausible." Under this standard, where factual gaps exist in plaintiff's allegations, the complaint will be deemed valid when and only when (1) the very allegation of the resulting harm to plaintiff and its surrounding circumstances gives rise to reasonable suspicion of unlawful behavior on the part of one of the participants in the relevant events, and (2) it is reasonable to believe that use of discovery devices will allow plaintiff to fill in sufficient evidentiary detail to get past a summary judgment motion and to proceed to trial.

F. Reconciling the Prior Pleading Decisions

The *Twombly* Court did not need to heighten the existing pleading standard from "notice pleading" (at least the "plus" version of that test)[67] to plausibility, because since its inception the standard of Rule 8(a) had generally been construed to demand that something approaching a reasonable plausibility standard be satisfied. Close examination of the leading pleading decisions since the inception of the Federal Rules reveals that existing doctrine is consistent with—if not inexorably dictated by—the "suspect circumstances" or "plausibility" version of notice pleading.

"Plausibility," then, is simply a new description of an established approach. As already demonstrated, the poster child for notice pleading, *Conley v. Gibson,* quite clearly qualifies under a plausibility standard.[68] The second-most-famous notice pleading decision of the period, authored by Judge Charles Clark, is the Second Circuit's decision in *Dioguardi v. Durning*.[69] There, an immigrant alleged in a self-drafted complaint that two cases of his "tonics" being shipped through customs had mysteriously disappeared.[70] While he provided nothing in the way of supporting detail, there was no reason to expect that he could supply it without having access to discovery. Judge Clark, invoking the revised pleading philosophy of the Federal Rules, rejected a motion to dismiss.[71] The complaint in *Dioguardi* clearly satisfied the plausibility standard as it has been explained in this Article. At the very least, one could reasonably conclude that the situation warranted further investigation through resort to the Federal Rules' discovery processes. In short, there existed enough suspicion to shift the risk of the wrong guess to defendants.

The most recent major decision in which the Supreme Court applied the precepts of notice pleading prior to *Twombly* was *Swierkiewicz v. Sorema, N.A.*,[72] a decision that was expressly reaffirmed in *Twombly*. The plaintiff, a 53-year-old native of Hungary, sued his former employer, a reinsurance company headquartered in New York and principally owned and controlled by a French parent corporation, for discrimination on the basis of national origin pursuant to Title VII of the 1964 Civil Rights Act[73] and on the basis of age pursuant to the Age Discrimination in Employment Act of 1967.[74] He had served as the company's chief underwriting officer, until being replaced by a much younger individual with only one year of underwriting experience at the time he was promoted. In contrast, plaintiff at the time had 26 years of experience. The district court dismissed the complaint because plaintiff "ha[d] not adequately alleged a prima facie case, in that he ha[d] not adequately alleged circumstances that support an inference of discrimination."[75]

In rejecting the defendant's motion to dismiss, the Supreme Court made clear that the complaint was not to be judged by the strict fact pleading standard of Rule 9(b), which is textually reserved for allegations of fraud or mistake.[76] A complaint controlled by Rule 8(a) need not include facts establishing a prima facie case of discrimination, the Court held. Unhelpfully, in its explanation the Court did little more than repeat the language of Rule 8(a) by stating that the complaint "must contain only 'a short and plain statement of the claim showing that the pleader is entitled to relief.' "[77] The Court pointed to the Federal Rules' system of "simplified notice pleading" that "relies on liberal discovery rules and summary judgment motions to define disputed facts and issues and to dispose of unmeritorious claims."[78] Because the plaintiff's complaint "gives [defendant] fair notice of the basis" for his claims, the motion to dismiss should be denied.[79]

Focusing solely on this language, it would seem arguable that *Swierkiewicz* is in conflict with both *Twombly* and *Iqbal*, despite the *Twombly* Court's insistence that *Swierkiewicz* is reconcilable with its plausibility standard.[80] After all, as vague as its allegations may have been, it is true that the complaint in *Twombly* gave the defendant "fair notice" of the type of conspiracy that plaintiffs were alleging.[81] But if one examines closely the situation in *Swierkeiwicz,* one can see that the facts pleaded in the complaint actually do satisfy a plausibility standard. The complaint alleged that (1) the plaintiff was of an age where age discrimination was a reasonable possibility, and (2) the plaintiff was far more qualified to serve in his position than the younger individual who replaced him. These allegations give rise to more than the mere possibility that age discrimination had occurred. At the very least, they give rise to a suspicion of unlawful conduct sufficient to allow plaintiff to get to the next stage of the process, discovery, to ascertain whether there was fire behind the smoke.

Whether the complaint's allegations, if supported by evidence at trial, would have been sufficient to resist a summary judgment motion is open to question. Under the standard of proof for trial established in *McDonnell Douglas Corp. v. Green*,[82] an employment discrimination suit at trial must evidentiarily establish a prima facie case, meaning that plaintiff must present evidence that supports an inference of discrimination."[83] If, at trial, a plaintiff presented evidence that did nothing more than establish that he qualified for protection against discrimination and that he was considerably more qualified for the position than his replacement, a jury could quite probably infer discrimination on the part of defendant.[84] In any event, at the very least *Swierkiewicz* appears to stand for the proposition that all a plaintiff must do is satisfy suspect circumstances, rather than allege the specific elements of a prima facie case. Thus, to the extent *Twombly* is ambiguous on the point, it is reasonable to choose to construe it, in accordance with *Swierkiewicz* (which the *Twombly* Court deemed to still be good law), as satisfying the requirements of the plausibility standard.

In the post-*Twombly* decision of *Erickson v. Pardus*, the Court once again appeared mysteriously to return to the rhetoric of the lax version of notice pleading (what I have called "notice pleading minus"). In his civil rights complaint, a prisoner alleged that he had been removed from a year-long treatment plan for hepatitis C, and that as a result his illness worsened.[85] It is true that the Court in a per curiam opinion mysteriously spoke in terms strongly reminiscent of the deferential notice pleading minus standard, under which a plaintiff is allowed to plead legal conclusions as long as he provides notice to the defendant of what it is alleged he had done unlawfully, and did so without any effort to distinguish *Twombly*.[86] Yet the two decisions are easily reconcilable on the basis of the *Twombly* Court's plausibility standard. We have already seen that in *Twombly* the Court found that in a complaint alleging an unlawful conspiracy in restraint of trade although an allegation of parallelism was *consistent with* illegal conspiracy, it was equally consistent with purely lawful behavior on defendants' part.[87] In contrast, in *Erickson,* according to the complaint, the plaintiff, who suffered from hepatitis C, had been removed from a one-year treatment plan by prison officials, and his illness worsened. It is, of course, possible that his illness would have worsened anyway had the treatment ultimately failed. But at the pleading stage, at least, it is reasonable to infer that if prison doctors had chosen the plaintiff on this extended treatment plan in the first place, there was at least a legitimate chance that it would have beneficial effect on plaintiff's medical condition. Otherwise, we would have to assume that prison officials had chosen to waste taxpayers' money and medical resources in a wholly worthless effort.

Whether a causal relationship in fact existed between removal from treatment and worsening of the illness would be an issue to be resolved at a later point in the proceeding (either on summary judgment or at trial). But surely it would be difficult to deny that such a link was "plausible." Under the circumstances, the risk of making a wrong decision could appropriately be shifted to the defendant.[88]

III. Discovery and the Litigation Matrix

A. Discovery and the Substantive-Procedural Balance

Discovery, as I have long argued, is reminiscent of the invention of fire. Like fire, if used with proper restraint discovery can be enormously valuable to achievement of the goals of the litigation matrix. But also like fire, if used carelessly or recklessly discovery can give rise to serious harm and destruction.

As explained in earlier sections, the concepts of revised pleading and discovery went hand in hand in the procedural model implemented in the original Federal Rules. It was quite clear that by simultaneously reducing the barriers imposed by the fact pleading requirement and establishing a complex set of court-enforced information gathering devices, Judge Clark and the rest of the Advisory Committee were attempting both to increase procedural fairness to plaintiffs and employ procedure as a more effective means of implementing substantive law. To illustrate, one need only point to the facts of *Conley*.[89] Recall that the plaintiffs there could never have satisfied the requirements of fact pleading; they of course had no way of knowing the specific details of their union's conspiratorial involvement with their employer at the outset of the litigation, since such conspiracies are by their nature secret. Yet to have denied them the opportunity to employ the discovery devices provided for in the Federal Rules under those circumstances would almost surely have been fundamentally unfair to those plaintiffs and a disruption of the enforcement of applicable substantive law. The availability of discovery, then, was essential to the procedural vindication of the *Conley* plaintiffs' substantive rights; the process thus restored the substantive-procedural balance, which rightfully plays such an important role in the shaping of the underlying litigation matrix of values.

We have already seen, however, that because of the potential burdens of discovery, an unduly lax pleading standard cannot be imposed without seriously skewing the substantive-procedural balance toward pathological over-enforcement of the substantive law.[90] As important a role as discovery is designed to play—and does play—in the effective procedural implementation of substantive law, the very same danger of pathological over-enforcement exists with regard to discovery.

B. Discovery and the Remaining Elements of the Litigation Matrix

As significant as a focus on the substantive-procedural balance is in order to fully understand the impact of the litigation matrix on discovery, it is equally important to recognize the relevance of the remaining elements of the litigation matrix to the analysis of discovery's potential risks and rewards. No matter how restrictively employed, of course, discovery will come at a cost: in the narrow and immediate sense, the costs imposed on the litigants and the system will be greater when discovery is employed than when it is not. The decision that such inherent costs were to be borne, of course, was made by the drafters of the Rules in 1938 when they inserted Rules 26-36. As a purely theoretical matter, at least, this seems to have been a wise call. Court-enforced information gathering devices can play a vitally-important role in enabling individuals intended to be protected by the substantive law to enforce their rights; simultaneously, they may also play a significant role in enabling defendants to protect themselves against illegitimate judgments or over-enforcement of substantive restrictions on their behavior.

It does not necessarily follow, however, that discovery should be free from significant restriction or control. To the contrary, if left wholly unregulated discovery can give rise to numerous procedural and substantive pathologies. In its most extreme form, intentionally abusive discovery effectively transforms the adjudicatory system into a means of facilitating legalized blackmail and extortion. The very threat of costly discovery likely induces rationally self-interested defendants to settle even non-meritorious suits for an amount smaller than the projected costs of discovery. Requests for wholly unnecessary discovery could thus function in an extortionate manner, financially coercing a defendant into settling unjustified claims. It should therefore hardly be surprising

that abusive discovery contravenes virtually all of the elements of the litigation matrix—fundamental fairness, efficiency, and maintenance of the substantive-procedural balance. It is fundamentally unfair to a defendant to allow the adjudicatory system to be employed against him as a weapon of coercion.

No one, presumably, would openly sanction or condone what is unambiguously and intentionally abusive discovery. The problem, of course, is to find ways to control such pathological discovery without either effectively destroying the beneficial effects of the discovery process or establishing control methods that are as economically inefficient as the abusive discovery itself. This has proven to be far more difficult a task than one might have hoped. It is to consideration of this difficult issue that the analysis now turns.

C. Controlling Discovery: The Alternative Models

The most frustrating aspect of the long and sad history of discovery control, however, is the Rule makers' total failure even to recognize, much less implement, a model of discovery control that (1) would curb not only intentionally-abusive discovery, but the probably more pervasive category of "excessive" discovery as well, and (2) would do so with only a relatively limited increase in the impact on the level of procedural costs and burdens imposed on the adjudicatory system. This method of discovery control, which is most appropriately called "the cost-allocation" model, is for the most part a self-executing system. In this sense, the model can be contrasted to four other variants, what can be called the "direct interventionist" model, the "direct restrictive" model, the "interventionist prophylactic" model and the "automatic prophylactic" model. The first and second of these alternative models involve direct restrictions on litigants' ability to engage in discovery, while the third and fourth alternative models involve efforts to prevent discovery abuse before it happens. The third alternative model is designed to deter discovery abuse through the establishment of judicially managed structure, while the fourth alternative seeks to prophylacticly prevent abuse through the use of purely litigant-based procedures.

1. The "Direct Interventionist" Model

The "direct interventionist" model can be described as "managerial," in the sense that it requires the court to directly involve itself in the control of discovery. The model is manifested in two different ways in the Federal Rules. One version, generally referred to as the "proportionality" requirement, is currently implemented through Rule 26(b)(2). The provision requires the court—either on motion of a party or on its own—to restrict discovery when, on the basis of a balancing of specified factors, it determines that discovery is unwarranted.[91] The second version of this model is embodied in Rule 26(c)'s authorization of a judicial protective order, designed to prevent or stop specific abusive practices.[92] The latter provision is the one method of discovery control that has existed since the Rules' original adoption in 1938. It stands as the last line of defense against specific instances of abusive or unjustified discovery. It was designed to provide trial courts with an additional method of intervening to prevent unjustified discovery.

Both of the rule-based versions of this model arguably perform legitimate roles in the control of unwarranted discovery. Rule 26(c) does so by leaving the court with virtually unlimited discretion to make individualized judgment calls. Rule 26(b)(2), on the other hand, is arguably more problematic, because it purports to provide a level of objectivity that simply fails to comport with the realities of the test. Far from providing any sort of objective guidance, the test necessarily requires the court to balance factors that are inherently subjective, without the slightest guidance as to how they are to be measured or how they are to be weighed against each other. At the very least, the process threatens to undermine the predictability element of the litigation matrix.[93] In a sense, both versions of the model pose a prima facie threat to the internal efficiency element of the litigation matrix,[94] because they are inherently labor-intensive—for the court, as well as for the litigants.

The primary problem with this method of discovery control, however, is neither its lack of predictability nor its potential inefficiencies. It is, rather, simply its failure to do an effective job of controlling discovery abuse. This does not mean that the methods should be abandoned. To the contrary, at the very least the Rule 26(c) protective order provision provides an enormously valuable method of controlling abusive discovery in the individual instance. It means, rather, that the direct interventionist model must be significantly supplemented if unwarranted discovery is to be controlled effectively.

2. The "Direct Restrictive" Model

At first glance, the direct restrictive model appears to be far less labor-intensive than the various versions of the direct interventionist model, because its different manifestations are, on their face, self-executing. These rule-based manifestations include the certification requirement of Rule 26(g), which requires certification of all discovery requests and responses, indicating that they are not interposed for improper purposes,[95] as well as the presumptive limitations imposed on the amount of discovery, expressly included in the rules describing the particular discovery devices—limitations which the court has discretion to alter in an individual instance. In both instances, however, as a result of these limitations, both the court and the litigants may become involved in potentially burdensome satellite litigation concerning either possible sanctions for violation of the certification requirement or the possible need to alter the presumptive limitations expressly imposed in the specific discovery rules.

Once again, the primary difficulty with this method of discovery control is probably not the potential burdens and inefficiencies of resultant satellite litigation, however real those dangers may or may not be. It is, rather, the inherently clumsy nature of this form of restriction. The Rule 26(g) certification requirement, for example, gives rise simultaneously to serious risks of over-protection and under-protection. On the one hand, it is far from inconceivable that risk-averse litigants, for fear of possible sanctions, will refrain from making discovery requests that, with perfect knowledge, they would have known would be perfectly legitimate. On the other hand, parties

operating in bad faith may comply with the certification requirement in the belief that they will be able to circumvent sanctions. Absent effective enforcement of the certification requirement, those parties will be able to undermine the salutary purposes served by that requirement. The problem with the presumptive limitations imposed on the amount of discovery, in contrast, is the "one size fits all" nature of those limitations. It is, of course, true (as already noted) that the court has authority to alter those limits in the individual instance, but that option inherently brings with it arguably unnecessary internal transaction costs necessarily involved in making the decision whether to authorize the alteration.

As in the case of the direct interventionist model, it does not necessarily follow that these forms of discovery control should be abandoned. It means, rather, that they need to be supplemented in some way in order to achieve the goal of assuring the discovery process's compliance with the dictates of the litigation matrix.

3. The Prophylactic Models of Discovery Control

a. The "Interventionist Prophylactic" Model

In contrast to the direct models, a prophylactic model of discovery control seeks to prevent or deter discovery excesses before they occur. What I refer to as the interventionist version of the prophylactic model involves the use of the discovery conference methodology that, in one form or another, has been around since the 1980 amendment to the Federal Rules.[96] While the procedure neither directly restricts discovery as a whole nor provides for intervention into specific situations in order to stop particular discovery abuse, its rationale is that by ordering the substance of the discovery process from the outset the model may deter pathological aberrations later on.

Though it is difficult to make definitive empirical assessments, it is possible that use of the discovery conference has had some beneficial impact on the discovery process. Even assuming that to be true, however, there is no doubt that the benefit comes at a cost, in terms of both judicial and attorney time. Moreover, because the process does not involve direct attacks on discovery abuse, it is very difficult to ascertain the true benefits the methodology brings about. That problem, after all, is inherent in the use of any prophylactic method.

b. The "Automatic Prophylactic" Model

The automatic disclosure device, originally adopted (in a more controversial version) in 1993 and currently embodied in Rule 26(a)(1),[97] seeks to avoid the burdens and confrontations that often accompany the discovery process by imposing on the litigants the obligation to automatically disclose certain basic information which, most likely, would have been requested in any event.

There are likely marginal benefits of efficiency derived from this anticipatory process, but it is difficult to see how it can deter or avoid the problems of inefficiency and distortion threatened by unwarranted or abusive discovery. It is therefore necessary to search for an alternative means of discovery control, one that functions differently from the currently-existing four regulatory models.

4. The Cost Allocation Model and the Control of "Excessive" Discovery

One can readily see why these alternative models of discovery control, either standing alone or in combination, fail to control discovery in an effective and efficient manner. They all create the risk of being over-effective, under-effective, economically inefficient, or even all three at once. This does not mean that they fail to serve any legitimate role in the overall scheme of discovery control. At the very least, however, it does mean that something more is needed in order to ensure that discovery in large or complex litigation is not permitted to degenerate into a pathological process of procedural inefficiency or substantive distortion. It is for this reason that it is necessary to turn to an alternative method of discovery control that has mysteriously been all but ignored since the very inception of the Federal Rules: the allocation of the costs of discovery not to the responding party (the overwhelmingly accepted practice) but rather to the requesting party.

From the outset, it is important to understand that I am not here advocating a process of cost *shifting;* indeed, the very use of that word would necessarily concede that the inertia of cost allocation appropriately belongs on the responding party, and must be *shifted* in order to have discovery costs attributed to the requesting party. Yet at no point has anyone—including those who drafted the Federal Rules in the first place—even attempted to rationalize the respondent-centric model of cost allocation that has dominated since the Rules' original promulgation. Were one actually to consider the issue afresh, it would be difficult to understand the assumptions inherent in such a model. It is true, of course, that in the crudest, most concrete sense the cost is immediately incurred by the responding party, not the requesting party. But that fact, standing alone, in no way necessarily implies that even at that point the cost is appropriately to be attributed to the responding party.

One may best understand the point by consideration of a simple analogy. Assume a co-worker asks you to do him a favor and pick up lunch for him. You do so, paying the $15 that the lunch costs. You then bring the lunch to your co-worker; unless he was raised by wolves, he will immediately thank you and reimburse you for your $15 expenditure on his behalf. Is such reimbursement appropriately characterized as "cost shifting" in anything but the most concrete, technical and immediate sense? At any point in this hypothetical transaction, was the cost of that lunch appropriately viewed, morally or conceptually, as *your* cost, rather than your co-worker's cost? Long-established principles of quantum meruit would readily answer that question in the negative.[98] You performed work on behalf of another, who was aware both that you were performing that work on his behalf and, as a result, incurring costs on his behalf. The law of quasi-contract unambiguously dictates that in such a situation the cost is deemed that of the party on behalf of whom the work was done, not of the party who performed the work.[99] In fundamental ways, the discovery process is identical to this hypothetical situation. The only differences are that in the case of discovery, the performing party is usually performing

the work not out of the goodness of his heart but rather due to the coercive threat of court sanction if he fails to do so. Moreover, the work performed by the responding party will not only help the requesting party but often actually harm the interests of the responding party himself. These differences, however, make even more bizarre the seemingly universal but wholly unsupported assumption that discovery costs are appropriately attributed to the responding party, rather than to the requesting party.

It should be clear that as both a legal and moral matter, the costs of discovery are properly attributable, in the first instance, to the requesting party. By imposing the costs of discovery on the responding party, then, our system has effectively required the responding party to provide a subsidy to the requesting party. To be sure, assuming no constitutional problems,[100] the system may choose to order such a subsidy. But because those who created the system implicitly—and inaccurately—assumed that the costs of discovery was properly seen as a cost to be borne by the responding party, our system has provided for a hidden subsidy, one recognized by no one. At the very least, democracy demands that the decisions of those who make fundamental choices of social policy make clear what those choices actually are, so a transparent debate of whether it is fair to impose such a subsidy may finally take place. This has never been done in the case of discovery costs.

Wholly apart from this complete lack of transparency, the implicit assumption that the costs of discovery are to be attributed to the responding party makes little sense, from any theoretical or practical perspective, particularly when coupled with the broad scope of discovery in the age of informational technology. In addition to its moral and legal bases, attribution of the costs of discovery to the discovering party, rather than the responding party, is likely to have significant instrumental benefits, because it would cure what has long been a fundamental economic pathology plaguing the discovery process: the externality inherent in the choice to invoke discovery. Simply put, under the prevailing practice the cost-benefit decision whether or not to invoke the discovery process is made by a party who risks incurring no cost, only benefit, even though it is quite conceivable that the choice will impose a significant cost on others. This lack of economic disincentive underscores what may well be a far greater harm to the system than intentionally abusive discovery: what can be most appropriately labeled "excessive" discovery. This concept includes discovery which, while not consciously interposed for purposes of delay or harassment, nevertheless gives rise to costs greater than its benefits in finding truth. Recall that in the foundational litigation matrix, the value of finding truth cannot be considered in a vacuum, wholly divorced from the costs to which the effort gives rise.[101] Some rough judgment must always be made by some decision maker whether the likely benefit to come from the effort justifies the effort's costs. Yet when the responding party, rather than the requesting party, bears the costs of the process, the requesting party has absolutely no economic disincentive not to make the request, regardless of its costs. Indeed, given that it is the requesting party's opponent who will bear that cost, one might even suggest that in a perverse sense, the higher the cost the greater the incentive to invoke the discovery process.

This focus on the subtle but important differences between "abusive" and "excessive" discovery underscores the manner in which a reversal in the ex ante presumption of discovery cost attribution can function in a symbiotic manner with both the direct and prophylactic methods of discovery control.[102] While those more judicially-driven practices are more likely to punish or deter abusive discovery, it is the self-executing shift in discovery cost allocation that is far more likely to deter the practice of excessive discovery.

The key social problem to which imposition of discovery costs on the requesting party might give rise derives from its inherently regressive nature: the poor will be more immediately and seriously impacted by such costs than will the rich. To be sure, this is also true of all litigation costs, though this fact has never caused us to shift all of the poor's litigation costs to the wealthier party. Moreover, particularly in the case of complex class action lawsuits, the real party in interest will not be the individual plaintiff but rather the plaintiff's attorneys, for whom the funding of such suits is simply a cost of doing business. In these cases, it would be wrong to see this alteration in discovery cost allocation as an inherently regressive practice. In any event, if there are particular substantive rights which the governmental body decides require procedural subsidization, that body may say so at the time it creates those rights. Therefore, even if one were to find the regressive impact of this reversal in cost allocation to be a matter of concern, a wholesale rejection of the cost allocation model would not be justified.

Even if society were to decide to subsidize a poorer litigant's discovery in particular suits, it hardly makes sense to impose that cost on his opponent, rather than on society as a whole. Indeed, to allow a private individual's unilateral filing of a lawsuit to justify imposition of discovery costs on the defendant gives rise to serious constitutional concerns of due process. The Supreme Court has long held that due process prohibits the deprivation of a defendant's property absent meaningful judicial involvement in the determination of that defendant's culpability.[103]

A conceivable objection to the reversal of the current cost allocation model might be that such a practice would simply shift the externality, for under the new model the responding party will have no incentive to keep costs down. But it is the discovering party who sets the contours of the response by the scope of its inquiries or production requests. In an important sense, then, the outer limits of the costs that the responding party will incur are set out by the requesting party. In any event, there always exists the possibility of judicial intervention to determine that the submitted costs are excessive. While it might be responded that such intervention would significantly increase the systemic burdens of the discovery process, it is highly unlikely that judicial intervention would be required in many instances. If the responding party knows that any excessive costs it incurs may well not be reimbursed, it is unlikely to risk incurring them in the first place.

IV. Pleading, Discovery, and the Revision of the Federal Rules

It should be clear by this point that an exclusive focus on the concern that plaintiffs be able to vindicate their substantive rights in court myopically ignores many of the most important elements of the foundational litigation matrix. While the danger of under-enforcement is surely to be avoided wherever feasible, the fundamental values of efficiency, fundamental fairness and maintenance of the substantive-procedural balance dictate the need to avoid both wasteful systemic costs and the substantive economic skewing that inevitably results from over-deterrence. Simply put, an understanding of the foundational normative precepts of modern procedural theory demand that pleading requirements impose some meaningful restraint on litigants' ability to invoke the elaborate discovery devices. Otherwise, it will be all but impossible to prevent parties who have suffered no legally cognizable injury from wastefully increasing both the internal costs of the adjudicatory system and the external costs of products and services in the marketplace. Moreover, once a litigant is permitted to get past the pleading stage to the discovery process, it is essential that the costs of that system are attributed in a manner consistent with the dictates of fundamental fairness and economic efficiency, in order to avoid the wasteful and inefficient misuse of that system.

Once all agree on these fundamental normative contours of the procedural system, the question naturally arises whether those goals may be achieved within the existing framework established by the Federal Rules of Civil Procedure, or whether instead fundamental changes in that framework are now required. Quite clearly, if the reversal in cost allocation presumption were to be imposed, an amendment adding this directive would need to be adopted. Of course, nothing in the current version of the Federal Rules expressly prohibits a court from shifting the costs of discovery from responding party to requesting party, and it is well-accepted that a court possesses discretion to shift costs under its broad powers given it by Rule 26(c).[104] But absent a provision in the Federal Rules expressly dictating that presumptively the costs of discovery are to be imposed on the requesting party, it appears clear that as a general matter courts will fail to allocate discovery costs in this manner. Thus, it is vitally important that the Federal Rules be amended to reflect such a change in traditional practice.

In contrast, the language of Rule 8(a) is in no way necessarily inconsistent with a plausibility standard. The Court in both *Twombly* and *Iqbal* has already construed the provision's text to implement this standard, and the provision's wording is sufficiently flexible to countenance such an interpretation, purely as a matter of textual construction. Indeed, it is highly likely that while the drafters of the original Federal Rules (in most cases)[105] sought to break away from the unduly high barriers to suit set by the code pleading standard for required factual detail, it is difficult to imagine that they intended to allow the pleading of vague and conclusory assertions of legal wrongdoing to enable a plaintiff to invoke the costly and burdensome discovery process absent some showing that the case was more than fanciful. Otherwise, defendants would regularly be at the mercy of any plaintiff who chose to sue them, because the mere filing of a complaint alleging a legal wrong would force defendant to suffer the costs and burdens of discovery. Absent overwhelming evidence to the contrary, one should not assume the Rules' drafters would have intended so untenable a result.[106] Thus, even absent an explicit amendment to Rule 8(a), courts applying that provision may and should reasonably construe it in accordance with the dictates of the plausibility standard.

The fact remains that in its present form Rule 8(a) is sufficiently ambiguous that it is subject to constructions different from that standard. In order to remove any conceivable ambiguity, therefore, the Advisory Committee would be well-advised to revise the provision's language in a manner that explicitly invokes plausibility as the standard that must be satisfied before a party may proceed to the discovery process. Were the Advisory Committee to track the language of the standard articulated in *Twombly,* the revision of Rule 8(a) would send a very clear message to all concerned that while in most cases a complaint need not satisfy the high bar imposed by a fact pleading standard,[107] the mere assertion of a vague and conclusory claim will not permit a litigant to proceed past the pleading stage. Something more is required in a complaint: the allegation of non-conclusory facts sufficient to give rise to the reasonable suspicion that a violation of plaintiff's rights has occurred. Thus, while perhaps as a doctrinal matter formal adoption of such an amendment may not be essential to restoring the proper balance to the pleading standard, doing so would avoid any further confusion among the courts as to what the controlling standard is. More importantly, adoption of such an amendment would stand as a social and political reaffirmance of the need for an economically balanced approach to the competing interests involved in the pleading context.

While formal amendment of the pleading rule may not be essential, the same is not true of the discovery process. As explained earlier, the key to taming the discovery process is to understand that, in the first instance, the costs of discovery are appropriately seen as costs attributable to the requesting party, rather than the responding party. While in its current form Rule 26(c), authorizing the issuance of protective orders, is framed in a manner that vests broad discretion in the district court's hands to "shift" costs, such a power is only rarely employed. In any event, the point of the amendment would not be merely to *authorize* the court to *shift* costs, but rather expressly to attribute the costs, in the first instance, to the requesting party. Rule 26 should therefore be amended to state unambiguously that discovery costs are attributable to the requesting party, unless applicable substantive law provides to the contrary or the court finds that a compelling reason for shifting the costs to the responding party exists.[108]

Conclusion: Assuring that the Genius of 1938 Survives in the Twenty-First Century (With a Little Tweek Every Now and Then)

There is much to celebrate as the Federal Rules of Civil Procedure rapidly approach their 75th birthday. The genius of Charles Clark was his effort to walk the tightrope of the substantive-procedural balance. The goal of Clark and his

colleagues was to assure that the rules of procedure neither over-enforce nor under-enforce the substantive law being enforced. They wisely saw that the barriers to suit imposed by the stringent standards of fact pleading failed that test, and therefore needed substantial revision. But to construe their abandonment of the fact pleading standard as an intended shift to *no standard at all* would be to commit the same sin of all-or-nothing clumsiness that had plagued the standard they sought to replace. First-year law students have long been taught that law is not simple; there are invariably conceptual and practical complexities that must be carefully balanced. Though it is perhaps difficult for us now to see it, the genius of the drafters of the original Federal Rules was their ability to recognize those complexities and to seek carefully to balance the competing needs as a means of achieving a solution that takes all of those complexities into account. Today, there are many who—in the name of the Rules' original drafters—urge that we impose an extremely lax pleading standard that allows plaintiffs to trigger the burden and costs of the discovery process by nothing more than a cryptic and conclusory assertion of a legal wrong. But now to characterize what the drafters did as the equivalent of a bull in a china shop, destroying everything in its path, would be to do them an injustice. The goal today should be to implement their genius under modern conditions. The Court in *Twombly* and *Iqbal* basically did just that, though one could justifiably question the extent to which it adequately explained its conclusion and rationale. Our goal today should therefore not be to sweep away the important insights of those pleading decisions, but rather to use them as a basis for a deeper intellectual exploration of the moral, social and economic foundations of modern procedure.

The drafters of the Rules, of course, were only human, and humans make mistakes—especially in the process of revolutionizing an entire system. In the discovery process, their first mistake was their failure even to consider the question of to whom discovery costs were to be appropriately attributed in the first instance. Their second mistake was their flawed implicit assumption that the costs were properly to be attributed not to the party who is best able to economically internalize the costs and benefits of discovery, but to the party who has little or no control over those decisions. Just as we have already corrected some of their failures in the discovery process over the years, it is now time to correct their errors—and then wish them a happy birthday.

Endnotes

1 *See generally* Michael E. Smith, *Judge Charles E. Clark and the Federal Rules of Civil Procedure,* 85 Yale L.J. 914 (1976).

2 *See* Hickman v. Taylor, 329 U.S. 495, 500-501 (1947).

3 *Id.* at 507.

4 *See generally* Charles E. Clark, *The Nebraska Rules of Civil Procedure,* 21 Neb. L. Rev. 307 (1942).

5 *See* Martin H. Redish, *Electronic Discovery and the Litigation Matrix,* 51 Duke L.J. 561, 593-600 (2001).

6 550 U.S. 544 (2007).

7 129 S. Ct. 1937 (2009).

8 *See, e.g.,* Kenneth S. Klein, *Ashcroft v. Iqbal Crashes Rule 8 Pleading Standards on to Unconstitutional Shores,* 88 Neb. L. Rev. 261 (2009); Suja A. Thomas, *The New Summary Judgment Motion: The Motion to Dismiss Under* Twombly*,* 14 Lewis & Clark L. Rev. 15 (2010); A. Benjamin Spencer, *Plausible Pleading,* 49 B.C. L. Rev. 431 (2008).

9 355 U.S. 41 (1957); *see* discussion *infra* at 149-150.

10 *See* discussion *infra* at 156.

11 One example is Robert G. Bone, *Plausibility Pleading Revisited and Revised: A Comment on* Ashcroft v. Iqbal, 85 Notre Dame L. Rev. 849 (2010).

12 *See* discussion *infra* at 146-147.

13 *See* discussion *infra* at 147.

14 *See* Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 559 (2007) ("It is no answer to say that a claim just shy of a plausible entitlement to relief can, if groundless, be weeded out early in the discovery process through 'careful case management,' *post* at 1975, given the common lament that the success of judicial supervision in checking discovery abuse has been on the modest side. See, *e.g.*, Easterbrook, Discovery as Abuse, 69 B.U.L.Rev. 635, 638 (1989) ('Judges can do little about impositional discovery when parties control the legal claims to be presented and conduct the discovery themselves'). And it is self-evident that the problem of discovery abuse cannot be solved by 'careful scrutiny of evidence at the summary judgment stage,' much less 'lucid instructions to juries,' *post,* at 1975; the threat of discovery expense will push cost-conscious defendants to settle even anemic cases before reaching those proceedings. Probably, then, it is only by taking care to require allegations that reach the level suggesting conspiracy that we can hope to avoid the potentially enormous expense of discovery in cases with no '"reasonably founded hope that the [discovery] process will reveal relevant evidence"' to support a § 1 claim.").

15 Notice Restoration Act of 2009, S. 1504, 111th Cong. (2009).

16 Website of the Advisory Committee on Civil Rules' 2010 Conference on Civil Litigation, Duke Law School, May 10-11, 2010, http://civilconference.uscourts.gov/LotusQuickr/dcc/Main.nsf/h_RoomHome/4df38292d748069d052567080016721 2/?OpenDocument (last visited Sept. 1, 2010).

17 *See* Martin H. Redish, *Summary Judgment and the Vanishing Trial: Implications of the Litigation Matrix,* 57 Stan. L. Rev. 1329 (2005). *See generally* Redish, *supra* note 5.

18 *See infra* Section II.

19 *See infra* Section III.

20 *Id.*

21 Redish, *supra* note 5, at 593-94.

22 *Id.* at 594.

23 See my elaboration of the role of private enforcement of systemic policies in Martin H. Redish, Wholesale Justice: Constitutional Democracy and the Problem of the Class Action Lawsuit 29-42 (2009).

24 Redish, *supra* note 5, at 595-96.

25 Mathews v. Eldridge, 424 U.S. 319, 335 (1976); *see also* Connecticut v. Doehr, 501 U.S. 1, 11-18 (1991).

26 *See* Hanna v. Plumer, 380 U.S. 460, 476 (1965) (Harlan, J., concurring).

27 550 U.S. 544 (2007).

28 *See* discussion *infra* at 150-151.

29 129 S. Ct. 1937 (2009).

30 4 William Blackstone, Commentaries 358; *see also* Alexander Volokh, *Guilty Men*, 146 U. Pa. L. Rev. 173 (1997).

31 It is, of course, true that whatever pleading standard is adopted applies to *both* complaints *and* answers. Thus, to the extent plaintiffs are required to satisfy a fact pleading standard, any affirmative defenses pled by defendants would have to satisfy a similar standard. As a categorical matter, however, there can be little doubt that a more demanding pleading standard will have

a more significant impact on plaintiffs. If plaintiffs are unable to satisfy fact pleading in their complaint, defendants will not even need to plead affirmative defenses.

32 Richard Marcus, Martin Redish, Edward Sherman & James Pfander, Civil Procedure: A Modern Approach 125 (5th ed., West Publishing 2009).

33 *Id.*

34 The extent to which either Dean Clark or the other rule makers fully grasped the truly dramatic nature of the change is unclear. On occasion, Clark suggested that the changes brought about by the Federal Rules were relatively minimal, and merely a natural evolution of existing pleading rules. *See* Clark, *supra* note 4, at 308. It is, of course, conceivable that understating the impact of the changes of the Federal Rules was merely an element of a strategy to disarm opponents to the adoption of the Federal Rules' pleading system. It is clear, however, that those opposed to the Rules' alteration of the pleading system fully grasped the magnitude of the changes.

35 Fed. R. Civ. P. 8(a).

36 Fed. R. Civ. P. 12(b)(6). Judge Clark himself would have preferred to dispense with the motion to dismiss altogether, instead relying solely on the summary judgment motion as a means of filtering suits prior to trial. However, his preferred approach was not adopted by the Advisory Committee, which chose to retain the motion to dismiss.

37 The discovery devices are presently governed by Rules 26-37 of the Federal Rules of Civil Procedure.

38 Fed. R. Civ. P. 37.

39 Fed. R. Civ. P. 9(b).

40 *See, e.g.*, Edward R. Finch, *Some Fundamental and Practical Objections to the Preliminary Draft of Rules of Civil Procedure for the District Courts of the United States,* 22 Am. Bar Ass'n J. 809, 809 (1936) (arguing that the Rules, "if finally adopted[,] will increase litigation and subject those against whom claims are asserted to an increasing number of suits based on unfounded claims," and would increase "to an appreciable extent the volume of litigation . . . . because the preliminary draft [of the Rules] adds greatly to the nuisance value of litigation brought without just cause" and "also increase so-called speculative litigation or litigation based on suspicion rather than facts, with the hope that such fishing may reveal a good cause of action . . . .").

41 Justice Souter expressly acknowledged this fact in *Twombly*. *See* Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 559 (2007).

42 Discovery reform of one form or another was undertaken by the Advisory Committee in 1980, 1983, 1991, and 2000.

43 Justice Souter's assertion is borne out by the relevant data. *See* Litigation Cost Survey of Major Companies 8 fig.4 (2010), *available at* http://tinyurl.com/LitCostSurvey-Major-Companies.

44 Use of the term "plausibility" did not begin until the Supreme Court's relatively recent decision in *Bell Atlantic Corp. v. Twombly*, 550 U.S. 544 (2007). *See* discussion *infra* at 19-20. However, it is my position that the Court's "plausibility" standard is identical to the standard originally intended to be embodied in Rule 8(a) and the standard the Court has traditionally applied. *See* discussion *infra* at 38-39.

45 *See* Fed. R. Civ. P. 9(b) (imposing fact pleading standard on claims of fraud or mistake).

46 Fed. R. Civ. P. 9(b).

47 355 U.S. 41 (1957).

48 *Id.* at 43.

49 *See* Richard L. Marcus, *The Revival of Fact Pleading Under the Federal Rules of Civil Procedure*, 86 Colum. L. Rev. 433, 433-434 (1986).

50 *See* Gillispie v. Goodyear Serv. Stores, 128 S.E.2d 762 (N.C. 1963) ("[T]he facts constituting a cause of action, rather than the conclusions of the pleader, must be set out in the complaint, so as to disclose the issuable facts determinative of the plaintiff's right to relief.").

51 355 U.S. 41 (1957).

52 Such a standard is consistent with the assertion by respected commentators that because "speculation" in a complaint's allegations is insufficient to satisfy a motion to dismiss, allegations that raise nothing more than mere "suspicion" is insufficient. Charles Alan Wright & Arthur Miller, Federal Practice and Procedure. "Speculation" amounts to an allegation that renders the defendant's unlawful behavior *possible*; on the other hand, a "plausible claim" standard requires a reasonable basis for believing that, if the allegation is ultimately supported by evidence, a reasonable fact finder could consider the allegedly unlawful behavior on the defendant's part more likely than not.

53 *See supra* Section I.

54 *See* Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 553-54 (2007) (quoting Brooke Group Ltd. v. Brown & Williamson Tobacco Corp., 509 U.S. 209, 227 (1993)).

55 550 U.S. at 555.

56 *Id.* at 556 (footnote omitted).

57 *Id.* at 557.

58 *Id.*

59 *Id.* at 566.

60 Ashcroft v. Iqbal, 129 S. Ct. 1937, 1939 (2009).

61 *Id.* at 1940.

62 *Id.* at 1949 (quoting *Twombly*, 550 U.S. at 556-57).

63 *Id.*

64 *Id.* at 1951.

65 *See* discussion *supra* at 146.

66 *See* discussion *supra* at 147-148.

67 *See* discussion *supra* at 149.

68 *See* discussion *supra* at 149-150.

69 139 F.2d 774 (2d Cir. 1944).

70 *Id.* at 774.

71 *Id.* at 775-76.

72 534 U.S. 506 (2002).

73 42 U.S.C. § 2000e *et seq.*

74 29 U.S.C. § 621 *et seq.*

75 534 U.S. at 509 (citing the district court).

76 Fed. R. Civ. P. 9(b); *see* 534 U.S. at 513.

77 534 U.S. at 508 (quoting Rule 8(a)).

78 *Id.* at 512.

79 *Id.* at 514-515.

80 *See* Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 569-70 (2007).

81 *See* discussion *supra* at 150-151.

82 411 U.S. 792 (1973).

83 *See, e.g.*, Swierkiewicz v. Sorema, N.A., 534 U.S. 506, 510 (2002) ("In subsequent cases, this Court has reiterated that the prima facie case relates to the employee's burden of presenting evidence that raises an inference of discrimination.").

84 *See, e.g.*, Lathram v. Snow, 366 F.3d 1085 (D.C. Cir. 2003) (Issues of material fact existed as to whether a female government employee was eligible for a higher-level position for which an agency hired an outside male candidate, and whether the employee was substantially more qualified for the job, precluding summary judgment on the employee's Title VII claim of discriminatory nonpromotion.); Scheidemantle v. Slippery Rock Univ. State System of Higher Educ., 470 F.3d 535 (3d Cir. 2006) (A genuine issue of material fact existed as to whether a female employee was more qualified for a university's locksmith position, precluding summary judgment in favor of the female in her Title VII claim of failure to promote due to gender discrimination.); *see also* Wilson v. BIE Aerospace, Inc., 376 F.3d 1079 (11th Cir. 2004); Erickson v. Pardus, 551 U.S. 89 (2007).

85 *Erickson*, 551 U.S. at 89-91.

86 It is true that the Court pointed to the plaintiff's special status as a *pro se* litigant. *Erickson*, 551 U.S. at 93-94. However, it did so *in addition to,* rather than as a rationale for, its description of the generally lax pleading standard.

87 *See* discussion *supra* at 150-151.

88 *See* discussion *supra* at 150.

89 *See* discussion *supra* at 149.

90 *See* discussion *supra* at 148.

91 Fed. R. Civ. P. 26(b)(2)(C).

92 Fed. R. Civ. P. 26(c).

93 *See* discussion *supra* at 147.

94 *See id.*

95 Fed. R. Civ. P. 26(g).

96 Amendment to Rule 26 (1980).

97 Fed. R. Civ. P. 26(a)(1).

98 Martin H. Redish & Colleen McNamara, *Back to the Future: Discovery Cost Allocation and Modern Procedural Theory*, 79 Geo. Wash. L. Rev. (forthcoming 2011) (manuscript at 14-18), *available at* http://ssrn.com/abstract=1621944.

99 *Id.*

100 *But see* discussion *infra* at 156.

101 *See* discussion *supra* at 153-154; *see also* Mathews v. Eldridge, 424 U.S. 319 (1976) (adopting a test that balances systemic costs against goal of accuracy in determining procedural due process).

102 *See* discussion *supra* at 154-155.

103 *See, e.g.*, Fuentes v. Shevin, 407 U.S. 67 (1972).

104 It is interesting to note that the current version of the Federal Rules expressly provides for such cost shifting in the case of expert witnesses. *See* Fed. R. Civ. P. 26(b)(4)(C). And, the Committee Note to Rule 26(b)(2) made clear that the conditions the Court may impose on ordering discovery include payment by the requesting party of all or part of the reasonable costs of obtaining the information. *See* Fed. R. Civ. P. 26(b)(2), Committee Note, at 17, *available at* http://www.uscourts.gov/uscourts/RulesAndPolicies/rules/EDiscovery_w_Notes.pdf.

105 In Fed. R. Civ. P. 9(b), the drafters established an island of fact pleading when the issues of fraud or mistake were to be pled.

106 *See* discussion *supra* at 148.

107 It should once again be emphasized that, for whatever reason, the drafters of the original Federal Rules chose in Rule 9(b) to impose a fact pleading standard in cases of fraud or mistake, and there appears to be no movement to alter that exception.

108 Beyond this amendment, it would also make sense for the Advisory Committee to consider possible alternative methods of directly controlling discovery. One such method that has been suggested is restriction of the scope of available discovery. For example, respected organizations have suggested: "Discovery in general and document discovery in particular should be limited to documents or information that would enable a party to prove or disprove a claim or defense or enable a party to impeach a witness." *See, e.g.*, Am. College of Trial Lawyers & Inst. for the Advancement of the Am. Legal Syst., Final Report 7 (2009); Final Report on the Joint Project of the American College of Trial Lawyers Task Force on Discovery and the Institute for the Advancement of the American Legal System 8, 14 (2009). That question, however, is an issue on which this Article takes no position.

Religious Liberties

Perry v. Schwarzenegger: Is Traditional Marriage Unconstitutional?

*By George W. Dent, Jr.**

Note from the Editor:

This article and the article that follows by Mark Strasser provide different perspectives on the main issue at stake in the Ninth Circuit case *Perry v. Schwarzenegger*, namely whether California Proposition 8, an amendment to the state's constitution providing that a marriage is a union between a man and a woman, violates the U.S. Constitution. We hope that publishing these articles helps contribute to the debate over this and other challenges to state constitutional provisions. The Federalist Society takes no position on particular legal or public policy initiatives. We welcome your responses to these articles; to join the debate, you can e-mail us at info@fed-soc.org.

Until 2000 the legal institution of civil marriage was understood to be available only to one man and one woman. In 2000 Californians passed an initiative statute (Proposition 22) reaffirming that understanding. The California legislature then enacted a law authorizing domestic partnerships for same-sex couples that offer the same legal treatment as marriage under a different name.[1] In 2008 the California Supreme Court nullified Proposition 22 and construed the state constitution to mandate that marriage be redefined to be available to same-sex couples.[2]

At the next opportunity, just five months later, the people of California approved Proposition 8, which added to the state constitution: "Only marriage between a man and a woman is valid or recognized in California." The initiative did not affect domestic partnerships.

Two same-sex couples who were denied marriage licenses after passage of Proposition 8 sued, challenging its constitutionality. The Governor, Attorney General, and other state officials refused to defend the law. Sponsors of Proposition 8 intervened to defend it. Judge Vaughn Walker of the U.S. District Court for the Northern District of California held that the intervenors had standing to defend the law and that Proposition 8 violates both the Due Process and Equal Protection Clauses of the U.S. Constitution.[3]

The defendant-intervenors appealed to the U.S. Court of Appeals for the Ninth Circuit. A three-judge panel of that court heard oral argument on the case in December 2010.[4]

I. Defendants' Standing

An initial question is the standing of the defendant-intervenors. In *Arizonans for Official English v. Arizona*[5] the majority opinion by Justice Ginsburg expressed in dictum "grave doubts" whether sponsors of a ballot initiative have standing to defend it if elected officials refuse to do so. However, the purpose of ballot initiatives is to enable voters to enact laws that government officials refuse to adopt. To deny sponsors of initiatives standing to defend them would in effect privilege officials to nullify this democratic process. It is unlikely that the court of appeals or Supreme Court will allow such nullification.

II. Findings of Fact

In reaching his decision Judge Walker made several crucial—essentially dispositive—determinations that he labeled findings of fact. Ordinary findings of fact are reversed only if found on appeal to be clearly erroneous.[6] However, the legislative and executive branches of government must constantly make findings of fact in order to formulate and enforce laws and regulations, and these "legislative facts" cannot be ignored by a trial court and are not subject to the "clearly erroneous" standard but to de novo review.

"Legislative choice is not subject to courtroom factfinding and may be based on rational speculation unsupported by evidence or empirical data."[7] The burden is "on the one attacking the legislative arrangement to negative every conceivable basis which might support it, whether or not the basis has a foundation in the record."[8] The issue, then, is whether the law satisfies the relevant standard of review, and this is an issue to be decided de novo by the appellate court, with due deference to democratic processes.

III. Issues Specific to Proposition 8

All but a few states have laws limiting marriage to one man and one woman. This suit certainly could have bearing on all those laws. Hesitant to overreach, the plaintiffs have struggled to identify particulars to differentiate Proposition 8 from other state laws and thereby narrow the scope (and the threat) of a ruling in their favor. One such particular is that Proposition 8 was adopted after the California Supreme Court mandated recognition of same-sex marriages. Thus, it is claimed, Proposition 8 differs from other state marriage laws because it deprived same-sex couples of an existing right rather than simply withholding a right they never had. However, if a right is not constitutionally mandated, how can it be unconstitutional for a state that has granted the right to change its mind and withdraw it?

In a few cases, the Supreme Court has overturned laws that withdrew constitutionally-discretionary rights because the Court found that the laws were impelled by an impermissible motive. In *Romer v. Evans*[9] for example, the Court struck down a Colorado constitutional amendment adopted by voter initiative that withdrew from the state legislature and local governments the power to enact laws against sexual-orientation discrimination. The majority said that the law was motivated by animus—a bare desire to harm—because it was not "directed

* *Schott-van den Eynden Professor of Law, Case Western Reserve University School of Law.*

to any identifiable legitimate purpose or discrete objective."[10] The majority further objected to the law because it amended the state constitution and thereby precluded those seeking laws against sexual-orientation discrimination from attaining them through ordinary legislation.

As the minority pointed out, the very purpose of constitutional provisions is to erect a barrier against ordinary legislation. Further, in many states constitutional provisions are the only laws that can be adopted by voter initiative. Therefore, to nullify such an initiative in effect deprives citizens of any power to act on a particular matter, even though the goal they seek is permissible under the Federal Constitution. *Romer* seems to make much of state constitutions unconstitutional.

The minority in *Romer* also noted that the Colorado initiative merely overrode local laws that were not constitutionally mandatory. However, the holding that the initiative had no "legitimate purpose" seems to mean that there would also be no legitimate purpose for *not* having laws against sexual-orientation discrimination to begin with, which would mean that such laws *are* constitutionally mandatory. Neither *Romer* nor any other Supreme Court decision, though, has so held.

Moreover, it is easy to find legitimate purposes for the initiative in *Romer*. In a free society, people are generally free to choose with whom to deal, even if others might consider one's choices irrational or improper. Discrimination is forbidden only on a few select grounds. The people of Colorado might plausibly have believed that discrimination based on sexual orientation is tolerable or that any problems it creates were not serious enough to require the heavy burden of government intrusion through antidiscrimination laws.

One cannot claim that the whole structure of marriage, recognized by every civilization throughout history, was contrived solely to harm homosexuals. As discussed below, it is also easy to find a legitimate purpose for Proposition 8. However, *Romer* seems to be a constitutional wild card—a precedent with no firm meaning that can simply be played whenever five Justices feel like striking down a law they do not like but in which there is no constitutional flaw.

IV. Standard of Constitutional Scrutiny

The central issue of *Perry* is the constitutional validity of laws restricting legal marriage to a woman and a man. A key subsidiary issue is the standard of constitutional scrutiny by which such laws should be reviewed under the Equal Protection and Due Process Clauses. All laws create distinctions and thus treat people unequally, and all laws limit rights either by forbidding some kind of behavior or by granting benefits for some persons or conduct and not others. In general, however, a law satisfies the Equal Protection and Due Process Clauses if the distinctions it makes have a rational basis. As noted above, this means only that there is some conceivable rational basis, even if that basis is not found in the record.

However, a few areas—those involving fundamental rights or distinctions that create a "suspect class"—are subject to strict scrutiny, requiring the state to show that the law serves a compelling interest that cannot be achieved by less discriminatory means.[11] The paradigm category of strict scrutiny is supposed to be race[12] because the Equal Protection Clause was adopted in the wake of the Civil War to protect the rights of former slaves who were being reduced to virtual serfdom by racially-discriminatory laws adopted in the former Confederate states.

Strict scrutiny is not necessarily fatal to discriminatory state action. For example, the Court has upheld racial preferences in university admissions on the factually dubious ground that racial diversity improves the quality of education.[13]

A strong argument can be made that traditional marriage serves a compelling state interest. The family is society's most basic institution, and traditional marriage has always been considered crucial to the successful functioning of the family. The suffering of children in our neighborhoods where marriage has lost its prestige and has ceased to be the norm certainly argues for a compelling need to retain and promote traditional marriage.

However, strict scrutiny should not be the applicable standard. The Supreme Court has never applied anything more stringent than the rational basis standard to sexual orientation. The Court's only decisions overturning laws based on sexual orientation are *Romer*, where the Court found no legitimate purpose for the law, and *Lawrence v. Texas*, which held only that disapproval of homosexual acts could not be enforced "through operation of the criminal law."[14] In both *Lawrence* and *Romer* the Court applied the rational basis standard. The Court in *Lawrence* said that Texas's criminal sodomy law "furthers no legitimate state interest which can justify its intrusion into the personal and private life of the individual."[15] *Perry* involves no "intrusion into the personal and private li[ves]" of homosexuals. The Court in *Lawrence* said expressly that the case did "not involve whether the government must give formal recognition to any relationship that homosexual persons seek to enter."[16]

The Equal Protection Clause was not intended originally to protect homosexuality. *Lawrence* might be defended on the ground that there had evolved a national consensus to tolerate private homosexual acts. Most states had repealed their criminal sodomy laws, and the few remaining laws were rarely enforced. In effect, these laws had ceased to be expressions of public morality and had become tools of arbitrary police harassment. It is not surprising that the Court would find such laws irrational.

Public attitudes about homosexual marriage are very different. Even if the meaning of equal protection can change with public consensus, that has not occurred here. Referenda on this issue have been held in thirty-one states, and in every one traditional marriage has been affirmed—usually by a large margin. Interracial marriage offers an instructive contrast. Prior to the Supreme Court's decision in *Loving v. Virginia* (discussed further below), some state courts had struck down antimiscegenation laws. In none of these states was there a serious effort to restore the law by ballot initiative.

The Court has sometimes suggested that strict scrutiny applies to groups with "an immutable characteristic determined solely by birth."[17] This could be a very broad concept. Intelligence, for example, is to some significant extent hereditary, yet many state actions (like college admissions and matriculation) discriminate on the basis of intelligence. Even if homosexual orientation is "an immutable characteristic,"

traditional marriage is intended to encourage responsible procreation, a purpose that is clearly irrelevant to homosexual conduct.

Further, the concept arguably does not apply to homosexual marriage. First, it is unclear that sexual orientation is "determined solely by birth." The American Psychiatric Association says "there are no replicated scientific studies supporting any specific biological etiology for homosexuality."[18] In at least some cases sexual inclinations may be influenced by experience.

Second, the issue in *Perry* is not sexual *orientation* but sexual *behavior*, which is not immutable. Some cultures have condoned some forms of homosexual activity,[19] and in these cultures such activity has been more common than in cultures where that activity is severely condemned. Californians could reasonably decide that they do not want to encourage homosexual conduct by honoring same-sex marriages.

Strict scrutiny may also apply to laws that discriminate against "discrete and insular minorities" that are subject to widespread discrimination.[20] Homosexuals certainly have experienced discrimination, although there is some question how widespread this discrimination is now. Unlike African-Americans, for example, homosexuals do not have lower average incomes than other Americans. California's authorization of domestic partnerships with the same legal rights as marriage and the ability of opponents of Prop. 8 to raise more money for their campaign than its proponents did show further that in California homosexuals have significant political influence.

The passage of Prop. 8 alone does not establish that homosexuals in California are a powerless, oppressed minority needing constant judicial insulation from democracy. Every substantive law (and the rejection of every proposal for a new law) creates winners and losers. Supporters of many causes lose repeatedly, but not every such group is entitled to the privilege of strict scrutiny.

Most American marriage laws, for example, exclude not only same-sex couples, but also marriage of close relatives ("endogamy") and marriage of groups of more than two persons ("polygamy" or "polyamory"). Supporters of these forms of marriage have not succeeded in any state, nor have they attained approval of civil unions or domestic partnerships for their relationships. Most advocates of same-sex marriage (including the plaintiffs in *Perry*) have not argued that these groups are "discrete and insular minorities" whose exclusion from marriage demands strict scrutiny review, yet it is hard to see why they (whose practices find more support in other societies than does same-sex marriage) do not merit as much judicial solicitude as homosexuals.

This inconsistency points to a more fundamental problem with the equal protection claim here. All parties in *Perry* agree that marriage is a privileged status. Plaintiffs do not challenge that status; they simply want same-sex couples to be eligible for it. However, if traditional marriage is unconstitutionally discriminatory, can any privileged status for marriage be upheld? The defense of traditional marriage is that it promotes responsible bearing and raising of children. If that defense is constitutionally inadequate, what constitutional justification is there for privileging marriage at all? Proponents of same-sex marriage do not answer this question.

V. The Fundamental Right to Marry

Judge Walker held that Proposition 8 violates the Due Process Clause by denying homosexuals a fundamental right of people to marry as they please. The Supreme Court has sometimes recognized a constitutional right to marry.[21] However, this right has always been limited, and the Court has never held or even hinted in dictum that the right extended to same-sex couples.

Same-sex marriage is very different from the cases where the Court has recognized a right to marry. In *Loving v. Virginia*,[22] the Court overturned a law forbidding interracial marriage. However, California does not forbid homosexual marriage; it simply does not license it, but leaves it as a private matter. Further, California offers homosexuals all the legal benefits of marriage, withholding only the label. The Court has never suggested that there is a fundamental right to the label "marriage."

As noted, the limitation of marriage to opposite-sex couples is only one of several traditional restrictions on marriage. The parties must be unmarried, i.e., no polygamy. The parties must not be too closely-related, i.e., no endogamy. And the parties must be adults, i.e., no child marriage. Unlike the same-sex requirement, all these practices have been condoned in many societies. If there is a fundamental right to same-sex marriage, *a fortiori* all the other practices must be permitted. It seems unlikely that the Supreme Court wants to take such a step.

The fundamental right to marry should mean the right to enter into a relationship that falls within the traditional definition of marriage, not to legal recognition of whatever arrangement some person or group of people wants to label marriage. The law struck down in *Loving* was a sharp departure from the traditional definition of marriage in Western civilization, which never forbade interracial marriages.[23] If anything, then, *Loving* is a precedent for adhering to, rather than nullifying, the exclusion of same-sex marriage because, unlike interracial marriages, Western civilization has never recognized same-sex relationships as marriages.

VI. The Case for (Traditional) Marriage

Judge Walker held that the "purported rationales" for the non-recognition of same-sex marriage "are nothing more than post-hoc justifications" by Prop. 8's proponents. As with his finding that Prop. 8 was motivated by a desire to harm homosexuals, this conclusion seems to rest on the premise that the institution of marriage was fabricated in every culture throughout history for the sole purpose of stigmatizing homosexuals.

Our society generally leaves adults free to arrange their own affairs. However, a woman and a man can create children who cannot protect their own interests. Marriage practices have varied among cultures in myriad ways. However, whatever else marriage is about (e.g., caste or property), it has always been centrally concerned with the bearing and raising of children. As Bertrand Russell said: "But for children, there would be no need for any institution concerned with sex. . . . [I]t is through children alone that sexual relations become of importance to society."[24]

Marriage both memorializes and solemnizes the relationship of a man and woman and provides the basis for an

enforceable legal commitment among them and their children. Marriage both reminds the parties and informs the world that they have entered into a relationship with responsibilities to each other and to the human lives they may create. In so doing, it encourages them to plan for responsible procreation, which includes not only conception but everything that might affect children. We know, for example, that married men work longer hours, commit fewer crimes, and abuse drugs and alcohol less than unmarried men. In other words, marriage works.

The district court held that "same-sex parents and opposite-sex parents are of equal quality."[25] It is not clear exactly what this means or what is the basis for the statement. However, there have been *no* studies comparing same-sex parents with married, biological parents. If the district court's statement means that the two are the same, it has no basis in fact. If it means something else, it is not relevant to the constitutionality of marriage.

Moreover, there are good reasons to believe that the finding is inaccurate. Innumerable studies have found the traditional family to be better for children than families with a single parent or cohabiting couples. If the district court is right, then cohabiting same-sex couples are better parents than cohabiting opposite-sex couples. The district court pointed to no studies purporting to support such a finding.

Same-sex couples can be allowed to adopt, but adoption is a legal event, not a biological act as is reproduction. Adopted children can be better protected through adoption proceedings and custody regulation than by fitting the square peg of same-sex relationships into the round hole of marriage.

The district court declared that "[p]ermitting same-sex couples to marry will not affect the number of opposite-sex couples who marry, divorce, cohabit, have children outside of marriage or otherwise affect the stability of opposite-sex marriages."[26] The only empirical basis the court gave for that finding was that recognition of same-sex marriages in Massachusetts supposedly has not affected rates of marriage and divorce.

This is not a firm basis on which to brand traditional marriage as irrational bigotry. First, Judge Walker's empirical finding may not be correct. Defendant-intervenors cited studies showing that divorce rates rose and marriage rates fell in Massachusetts from 2004 to 2007.[27] Similarly, marriage rates have declined and illegitimacy rates have risen in the Netherlands since it recognized same-sex marriages.

Further, the Massachusetts law has been in effect only briefly. More time is needed to determine its long-term effects, such as whether it will influence the raising of children and the use of artificial reproduction. Moreover, Massachusetts is a small state in a large country, in nearly all of which traditional marriage still prevails. Massachusetts may be atypical. Even its own residents may consider its law an aberration, not a general change of the meaning of marriage.

There are also substantial reasons to think that recognizing homosexual marriage would impair the social prestige of marriage. At trial both plaintiffs and defendant-intervenors introduced expert statements that validating same-sex marriage would radically alter the institution.[28] Some gay activists support same-sex marriage for the express purpose of destroying its social standing.[29]

Recognizing same-sex marriage would transform marriage from its immemorial function as an arrangement centrally concerned with children to one primarily for the gratification of adults. As discussed above, thirty-one states have held referenda on the issue, and in every one the voters favored traditional marriage, usually by large margins. It is hard to believe that a Supreme Court decision branding the majority of Americans (and, indeed, virtually all human beings who have ever lived) as irrational bigots because they believe there is something special about the ability of a woman and a man to create human life would not diminish public respect for marriage.

Much of the legal benefit of marriage is achieved through the expressive function of law—the effect of the law in promoting certain norms by the law's symbolic support. Perhaps the esteem for marriage generated by the law's symbolic support would be impaired by extending it to intrinsically sterile relationships, but this esteem may be less impaired if a different label is used. Whether one thinks that California domestic partnerships go too far or not far enough in recognizing same-sex relationships, that approach is not irrational.

VII. Future Proceedings

If, as expected, the Ninth Circuit panel affirms the district court's decision, the defendants could seek an en banc rehearing, or head for the Supreme Court as quickly as possible. The latter would set the stage for a Supreme Court hearing and decision in its 2011-12 Term, thus making the case a potential issue in the 2012 presidential and other elections.

Conclusion

The Constitution confers no right to legal validation of same-sex marriage. As Judge Richard Posner has said, "If there is such a right, it will have to be manufactured by the justices out of whole cloth."[30] For the Supreme Court to do so would gravely damage its legitimacy and invite efforts to change the composition of the Court. However justified the public anger at the obliteration of traditional marriage, such moves would create a dangerous precedent. It is hoped that the Court will not provoke such action.

Endnotes

1 Cal. Fam. Code § 297.5.

2 *In re* Marriage Cases, 183 P.3d 384 (Cal. 2008).

3 Perry v. Schwarzenegger, 704 F. Supp. 2d 921 (N.D. Cal. 2010).

4 Defendant-intervenors argue that the case is governed by *Baker v. Nelson*, 409 U.S. 810 (1972), in which the Court summarily rejected an appeal from a decision upholding the Minnesota marriage law that excluded homosexual couples. The Supreme Court generally gives little precedential force to such summary denials of appeals. *See* John E. Nowak & Ronald D. Rotunda, Constitutional Law § 2.5(c) (8th ed. 2010).

5 520 U.S. 43 (1997).

6 Fed. R. Civ. Pro. 52(a)(6).

7 Heller v. Doe, 509 U.S. 312, 320 (1993).

8 *Id.* at 320-21.

9 517 U.S. 620 (1996).

10 *Id.* at 636.

11 United States v. Carolene Products Co., 304 U.S. 144, 152 n.4 (1938).

12 *See* Nowak & Rotunda, *supra* note 4, § 14.3(a)(iii).

13 Grutter v. Bollinger, 539 U.S. 306 (2003).

14 539 U.S. 558, 571 (2003).

15 *Id*. at 578.

16 *Id*.

17 Frontiero v. Richardson, 411 U.S. 677, 686 (1973) (plurality op.).

18 American Psychiatric Association, Gay/Lesbian/Bisexuals (2009).

19 Significantly, however, the homosexual conduct condoned has usually been the sodomizing of boys by adult men or of slaves or other social inferiors by their masters or social superiors. Almost never have same-sex relationships between equal adults been approved.

20 United States v. Carolene Products Co., 304 U.S. 144 (1938).

21 *E.g.,* Zablocki v. Redhail, 434 U.S. 374 (1978) (overturning a law forbidding marriage to persons who are in violation of a court order of child support).

22 388 U.S. 1 (1967).

23 *See* Irving G. Tragen, *Statutory Prohibitions Against Interracial Marriage*, 32 Cal. L. Rev. 269, 269 & n.2 (1944) ("[A]t common law there was no ban on interracial marriage . . . .").

24 Bertrand Russell, Marriage and Morals 77, 156 (1929).

25 Perry v. Schwarzenegger, 704 F. Supp. 2d 921, slip op. at 127 (N.D. Cal. 2010).

26 *Id.* at 83.

27 Ctr. for Disease Control, Divorce Rates by State: 1990, 1995, and 1999-2009, *available at* http://www.cdc.gov/nchs/data/nvss/marriage_rates_90_95_99-09.pdf; Ctr. for Disease Control, Marriage Rates by State: 1990, 1995, and 199-2009, *available at* http://www.cdc.gov/nchs/data/nvss/marriage_rates_90_95_99-09.pdf.

28 *See, e.g.*, Witherspoon Inst., Marriage and the Public Good: Ten Principles 18-19 (2008).

29 *See, e.g.*, Michelangelo Signorile, *Bridal Wave*, Out Mag., Dec./Jan. 1994.

30 Richard Posner, *Wedding Bell Blues*, New Republic Online, Dec. 19, 2003, available at http://www.freerepublic.com/focus/f-news/1043523/posts.

Perry, Same-Sex Marriage, and Federal Constitutional Guarantees

*By Mark Strasser**

In *Perry v. Schwarzenegger*,[1] a California district court struck down California's Proposition 8, which amended the state constitution to preclude same-sex marriage. The opinion is of interest for a number of reasons, some of which have limited applicability outside of the California context and others of which have more general application. This essay focuses on the points in *Perry* that are of wider application, some of which would seem applicable to all states banning same-sex marriage and others of which are relevant to a subset of those states.

Perry suggests that Proposition 8 violated both due process and equal protection guarantees contained within the United States Constitution, and at least one issue involves the degree to which the analyses offered in *Perry* should or will have constitutional force in other parts of the country. It is useful, then, to consider *Perry*'s arguments in detail.

The right to privacy protects a constellation of rights connected with family—the rights to marry, procreate, and raise one's child are all included within the right to privacy. The rights thereby protected are not absolute—privacy rights can be overridden by a statute that is narrowly tailored to promote compelling state interests. Nonetheless, the state must bear a heavy burden to justify infringing on an interest protected by the right to privacy.

It might be thought, then, that the right to marry a same-sex partner can only be limited if the state can carry its heavy burden of justification, because the right to marry is one of the rights protected under the right to privacy. However, many courts have suggested that the right to marry a same-sex partner should be treated as a separate and distinct right. In contrast, the *Perry* court refused to characterize the plaintiffs as seeking to establish their "right to marry a same-sex partner" but, instead, as seeking to vindicate their "right to marry." This dispute is not merely a matter of semantics, because the plaintiffs are then not seeking to have a new right recognized but, instead, are seeking to enforce a right that has already been recognized.[2]

Consider *Loving v. Virginia*,[3] where Virginia's anti-miscegenation statute was challenged as violating federal constitutional guarantees. Richard Loving and Mildred Jeter did not seek to a have a new right recognized—"the right to marry someone of another race." Further, when the United States Supreme Court recognized that Virginia's anti-miscegenation law violated the Lovings' fundamental rights, the Court did not recognize a special right to marry outside of one's race but, instead, a more generalized right to marry.

Some commentators suggest that the right to marry recognized in *Loving* was only meant to include those who could have children though their union. But that cannot be correct, for it suggests that different-sex couples who cannot procreate through their union do not have a fundamental right to marry. Such a characterization of the right might exclude the elderly, those who are sterile, and those with certain physical handicaps. No court has held that the right to marry is contingent on individuals having the ability and desire to procreate, and Justice Scalia has recognized the implausibility of the procreation argument.[4]

Some courts have suggested that the reason people who cannot procreate nonetheless have the right to marry is that the right to privacy protects individuals from having to establish their ability or willingness to procreate. But this, too, is incorrect. In fact, some states prohibit certain people (first cousins) from marrying unless they can establish their *inability* to procreate through their union.[5] Two lessons might be learned from such statutes. First, states can and do condition marriage on procreation concerns, so it is not as if states could not do so, assuming no independent bar to the state's precluding the marriage at issue.[6] Second, the state and individual interests in marriage are not limited to those involving children, because the state is limiting first-cousin marriages to those that likely will not involve children.

Loving is important to consider for several reasons. First, the *Loving* Court was striking down a marriage prohibition that had been in existence since before the Nation's founding. Further, several states in addition to Virginia had anti-miscegenation laws, sometimes within their constitutions.[7] Thus, it could not be said that the substantive due process right to marry that was recognized in *Loving* was deeply-rooted in the Nation's history or tradition or was so rooted in the traditions and conscience of our people as to be ranked fundamental.[8] On the contrary, there was a long history in some states of prohibiting such unions, so the history and traditions test could not have been the basis of the holding.[9]

It might be argued that the right to marry is deeply-rooted in the Nation's history, even if the right to marry someone of another race is not. But an analogous point might be made here, namely, that the right to marry is deeply-rooted, even if the right to marry a same-sex partner is not. In both cases, the individuals have the right to marry, and the question is whether the Federal Constitution permits a state to limit that choice on the basis of the races or sexes of the parties.

Those who argue that *Loving* only recognized the right to marry for individuals capable of reproducing through their union seem not to appreciate that the *Loving* opinion nowhere even mentions children. The Court's reticence on this subject is quite understandable when one considers Virginia's justification for its anti-miscegenation policy, namely, that the state wanted to preclude interracial marriage for the sake of the children that might be born of such unions.

Loving shifted the focus of the constitutional debate from the interests of children to the interests of the adults. "The freedom to marry has long been recognized as one of the vital personal rights essential to the orderly pursuit of happiness by free men."[10] Marriage promotes the interests of adults, and, as the *Perry* court notes, that is true whether the marital partners are of the same sex or of different sexes.[11]

** Trustees Professor of Law, Capital University Law School, Columbus, Ohio.*

While the *Loving* Court did not focus on children, it did mention that marriage is fundamental to our existence and survival. The Court did not develop the point, although a few possible explanations might be offered. For example, a variety of benefits accrue to the partners in a marriage.[12] Insofar as the marital partners are benefited and become more productive members of society, society benefits as well.

Another way to understand why marriage is fundamental to our existence and survival involves the benefits of marriage for the next generation, which was also a point suggested in *Zablocki v. Redhail*.[13] *Zablocki* involved a challenge to a Wisconsin law by Roger Redhail, who was prevented from marrying his pregnant fiancé. Redhail had fathered a child (with a different woman) out of wedlock while he was in high school, and Wisconsin law prevented noncustodial parents from marrying if they had children from a previous relationship who were not receiving the required child support. The *Zablocki* Court commented:

> It is not surprising that the decision to marry has been placed on the same level of importance as decisions relating to procreation, childbirth, child rearing, and family relationships. As the facts of this case illustrate, it would make little sense to recognize a right of privacy with respect to other matters of family life and not with respect to the decision to enter the relationship that is the foundation of the family in our society.[14]

A few points might be made about what the Court is saying here. When describing marriage as the foundation of family, the Court is not suggesting that children are only born into marriages. Redhail had already fathered one child out of wedlock and might well be fathering another out of wedlock, given the Wisconsin law that precluded him from marrying his fiancé. Rather, the Court was suggesting that there are a variety of rights associated with family, and that it does not make sense to recognize those rights while at the same time not recognizing the right to marry. When describing marriage as the foundation of family, the Court was likely further suggesting that marriage provides a setting in which children might prosper.

The *Zablocki* Court's suggestion that marriage provides a setting in which children might flourish may also be what the *Loving* Court had in mind when saying that marriage is necessary for the survival of humankind. Neither Court believed that children are only produced within wedlock, as the facts of *Zablocki* clearly illustrate. Rather, both seemed to appreciate that marriage provides a setting in which children can be loved, cherished, taught, and helped to thrive.

Which children can be helped in such a setting? Certainly, children who are biologically-related to both parents can benefit from living in such a setting, but so can children who are biologically-related to only one or perhaps to neither of the parents. Many families in the United States, whether the parents are of the same sex or of different sexes, involve children who are not biologically-related to both adults. The children might have been adopted or might have been the product of a previous relationship of one of the adult partners. But today's demographics demand recognition that many children are being raised in such homes, and if the survival of humankind is dependent on children being raised by two parents, each of whom is biologically-related to the children, then humankind, and our Nation in particular, may be facing some difficult times.

The *Perry* court noted that same-sex couples are raising children, and that the children are doing quite well.[15] Further, gay and lesbian parents have the right to raise their children just as other parents do. But *Zablocki* suggests that this is a reason to recognize same-sex marriage—it makes little sense to recognize a right of privacy with respect to other family matters but not recognize a right to marry. Children, whether raised by same-sex or different-sex couples, can benefit from the stability that marriage can bring.

While there are no appreciable differences between children raised by same-sex parents and those raised by different-sex parents,[16] that should not be the focus of the discussion when examining the right to marry. (We do not limit marriage to those who would be optimific parents.) Whether or not same-sex marriage is recognized, children will be raised by same-sex parents. The question at hand is whether the children raised by such parents should be able to benefit from the increased stability and other benefits associated with marriage or whether, instead, they should be forced to suffer the different opportunity costs associated with their parents having been precluded by law from marrying.

If the right to marry a same-sex partner falls within the right to marry, then it seems unlikely that the state will be able to justify refusing to recognize such marriages. Indeed, the *Perry* court held that there was no legitimate basis to refuse to recognize such marriages.[17]

To understand whether same-sex marriage bans promote legitimate state interests, it is important to consider what happens when a same-sex marriage ban is struck down by the courts or is repealed by a legislature. Traditional marriages are not thereby held unconstitutional or somehow denied legal recognition. Rather, those marriages are recognized, and other marriages are recognized as well. Indeed, it is somewhat difficult to specify what legitimate interests are promoted by refusing to recognize same-sex marriages. It is not as if such bans make it more likely that different-sex couples will marry or remain married. Instead, such bans merely impose a burden on same-sex couples and their families without bringing about any offsetting benefits for anyone else. Further, even when same-sex marriage is recognized by the state, religious groups do not have to permit such marriages to be celebrated if such unions contravene religious beliefs.[18] Of course, those religious groups that do recognize same-sex marriage would then be able to confer both religious and civil significance on those unions.

It might be thought that prohibiting same-sex marriage somehow promotes morality. But that is exactly the kind of argument that is precluded by *Lawrence v. Texas*,[19] in which the Court struck down a Texas law barring same-sex sodomy. While recognizing that some individuals sincerely believe same-sex relations immoral, the *Lawrence* Court suggested that the majority could not "use the power of the State to enforce these views on the whole society through operation of the criminal law."[20]

The *Lawrence* Court explained that the "Texas [sodomy] statute furthers no legitimate state interest which can justify its

intrusion into the personal and private life of the individual."[21] But it is helpful to look closely at the language employed. The Court did not say that the statute furthered no legitimate interest at all but, instead, that it furthered no legitimate interest that would justify the intrusion. Why mention this? Because the Court's comments are consistent with its using a rational basis test—the state had no legitimate interest implicated at all—but also consistent with a higher level of scrutiny being used—while the state had some legitimate interests implicated, those interests were not sufficient to justify the burden that the state was imposing.

There are other reasons to think that the *Lawrence* Court was using a higher level of scrutiny than rational basis, Justice Scalia's comments in dissent to the contrary notwithstanding.[22] The Court cited a variety of cases within the right to privacy jurisprudence as support for striking down the Texas law—*Griswold*, *Eisenstadt*, *Roe*, *Carey*, and *Casey*—and one might wonder why those cases would be cited in an opinion in which the right to privacy was not at issue.

Traditionally, the Court has privileged relationships such as marriage over sexual relations. In *Griswold v. Connecticut*,[23] Justice Goldberg explained in his concurrence that marriage could not be regulated even though non-marital sexual relations could be.[24] Privileging same-sex relations but not same-sex relationships inverts the traditional priorities. That said, there is reason to read *Lawrence* as recognizing that same-sex relationships themselves have value, because part of the rationale for protecting same-sex relations was that the sexual "conduct can be but one element in a personal bond that is more enduring."[25] The claim here is not that the *Lawrence* Court held same-sex marriage constitutionally protected. On the contrary, the Court expressly refused to address that issue,[26] just as it had expressly refused to address whether interracial marriage was protected when striking down a law more severely punishing interracial, non-marital relations than intra-racial, non-marital relations.[27]

A separate issue is whether same-sex marriage bans are unconstitutional because they violate equal protection guarantees. As an initial point, there is some confusion with respect to the basis of the classification at issue. Consider a statute that says, "A man can marry a woman but not a man; a woman can marry a man but not a woman." Such a statute expressly classifies on the basis of sex.[28] A separate issue is whether such a classification can be justified, but the statute itself is a facial, sex-based classification. Why is that important? Because facial sex-based classifications trigger heightened scrutiny, just as facial race-based classifications trigger strict scrutiny. When a facial race-based classification is at issue, there is no need to show in addition that one race would be adversely affected more than another in order for strict scrutiny to be imposed. For example, almost forty years before *Lawrence* was decided, the Court examined and struck down a statute more severely punishing interracial non-marital sexual relations than intra-racial non-marital sexual relations.[29] It was not necessary to show that the statute adversely impacted one race more than another in order for close scrutiny to be triggered. So, too, where a statute facially discriminates on the basis of sex, there is no need to show that one sex is adversely affected more than another.

One might contrast the kind of classification at issue in a same-sex marriage ban with the kind of classification at issue in *Romer v. Evans*,[30] which involved a Colorado constitutional amendment enacted by referendum that precluded affording antidiscrimination protections on the basis of sexual orientation.[31] On its face, this was orientation rather than sex discrimination, and the Court suggested that this kind of classification could not even pass the lowest level of scrutiny.[32]

Justice O'Connor suggested in her *Lawrence* concurrence that laws targeting on the basis of sexual orientation had to be given "a more searching form of rational basis review."[33] It might be noted that statutes might expressly classify on the basis of sex but be intended to target on the basis of orientation. Such statutes should receive heightened scrutiny because of their facial basis. However, if there were some way to avoid triggering heightened scrutiny notwithstanding the facial classification, one would presumably still trigger this heightened form of rational basis review that was used in *Romer* and, perhaps, was used in *Lawrence*. Or, perhaps orientation discrimination itself should receive heightened review, as some of the state supreme courts have suggested in light of their state constitutional guarantees.[34]

Whatever the appropriate level of scrutiny, states must offer some justification for the discrimination. One justification that has had some success involves the point that same-sex couples cannot conceive accidentally—they have to plan in order to have children through adoption or through the use of advanced reproductive techniques. In contrast, different-sex couples may conceive on the spur of the moment. For this reason, it is thought that states have a greater interest in encouraging different-sex couples to marry than same-sex couples to marry—that way unplanned pregnancies are more likely to occur within the context of a marriage.[35]

Yet, such an argument implies that the difficult part of parenting is in producing children rather than raising them. But, for most parents, the exact opposite is true. In any event, states that are deciding whether to permit same-sex couples to marry are not choosing between on the one hand permitting same-sex couples to marry and on the other permitting different-sex couples to marry. Rather, they are choosing between permitting different-sex couples to marry on the one hand and permitting both same-sex and different-sex couples to marry on the other. As Chief Judge Kaye pointed out in her dissent in *Hernandez v. Robles*, there "are enough marriage licenses to go around for everyone."[36] Absent reason to think that some different-sex couples would refuse to marry because same-sex couples were allowed to marry,[37] same-sex marriage bans do not encourage different-sex couples to marry (and thus have children within wedlock); all they do is prevent same-sex couples from marrying, which means that any children that they have will necessarily be out of wedlock. If children tend to do better when raised within a marriage than when raised outside of one, those supporting same-sex marriage bans do not end up helping children raised by different-sex parents (because the parents' decision to marry simply will not be affected by whether the state permits same-sex marriage), but will harm some of the children raised by same-sex parents who would

have married if they could have.

Some of *Perry* is directed toward particulars in California. For example, California recognizes domestic partnerships, so at least one question is what legitimate interests are served by creating a separate status for same-sex couples. Such a separate status seems stigmatizing,[38] for it suggests that same-sex couples would somehow sully marriage if permitted to marry.[39] Further, domestic partnerships are not treated as marriages for federal purposes,[40] so if the Defense of Marriage Act is struck down or repealed, same-sex domestic partnerships (or civil unions for that matter) would be inferior to same-sex marriages with respect to tangible benefits as well as with respect to symbolism.

California, like many states, permits second-parent adoptions, so it permits each member of a same-sex couple to be legally-related to the child whom they are raising.[41] The state is not questioning the ability of same-sex couples to parent, but is nonetheless imposing opportunity costs on the very children whom the state may have entrusted to those parents though adoption. It is difficult to understand how a state could do this and claim to be interested in the welfare of children.

Perry suggests that same-sex marriage bans do not serve a legitimate state interest. If that analysis is accepted by other courts, then same-sex marriage bans as a general matter are constitutionally vulnerable. Further, *Perry* offers reasons to suggest that such bans should be examined with heightened or strict scrutiny. If those reasons are taken seriously by other federal courts, then it is even less likely that same-sex marriage bans in other states will be held to pass constitutional muster. While it is unclear what the United States Supreme Court will ultimately say with respect to the kinds of arguments offered by *Perry*, they are careful, weighty, and worthy of serious consideration.

Endnotes

1 704 F. Supp. 2d 921 (N.D. Cal. 2010).

2 *Id.* at 993 ("Plaintiffs do not seek recognition of a new right. . . . Rather, plaintiffs ask California to recognize their relationships for what they are: marriages.").

3 388 U.S. 1 (1967).

4 *See* Lawrence v. Texas, 539 U.S. 558, 605 (2003) (Scalia, J., dissenting) (rejecting the encouragement of procreation argument, "since the sterile and the elderly are allowed to marry").

5 *See, e.g.*, Ind. Code 31-11-1-2 (only first cousins over age sixty-five may marry); Wisc. Stat. Ann. § 765.03 (first cousins may marry if woman over age fifty-five or there is a signed affidavit by a physician stating that one of the parties is permanently sterile).

6 Because there is no right to marry someone too closely-related by blood, the states can create an exception permitting some to marry among those who would otherwise be precluded from marrying because within the state's incest limitation. The state could not as a general matter limit marriage to those able (or for that matter only to those unable) to produce a child. Limitations on marriage, e.g., with respect to incest limitations, are permissible where the state has compelling interests at stake and the limitation is narrowly tailored to promote that interest.

7 *Cf. Perry*, 704 F. Supp. 2d at 957 (noting that several states had anti-miscegenation laws).

8 *See* Washington v. Glucksberg, 521 U.S. 702, 720-21 (1997).

9 It might be noted that the Court had refused to strike down Virginia's anti-miscegenation law a mere eleven years before the *Loving* opinion was issued. *See* Naim v. Naim, 350 U.S. 985 (1956).

10 Loving v. Virginia, 388 U.S. 1, 12 (1967).

11 *Perry*, 704 F. Supp. 2d at 962 (discussing ways that marriage benefits both partners).

12 *Id.* at 969 ("Same-sex couples receive the same tangible and intangible benefits from marriage that opposite-sex couples receive.").

13 434 U.S. 374 (1978).

14 *Id.* at 386.

15 *Perry*, 704 F. Supp. 2d at 980 ("Children raised by gay or lesbian parents are as likely as children raised by heterosexual parents to be healthy, successful and well-adjusted.").

16 *See id.*

17 *Id.* at 997 ("[E]xcluding same-sex couples from marriage is simply not rationally related to a legitimate state interest.").

18 *Id.* at 976 (noting that even when same-sex marriage was recognized, "no religious group was required to recognize marriage for same-sex couples").

19 539 U.S. 558 (2003).

20 *Id.* at 571.

21 *Id.* at 578.

22 *See id.* at 586 (suggesting that the Court had used "rational-basis review").

23 381 U.S. 479 (1965).

24 *See id.* at 499 (Goldberg, J., concurring).

25 *Lawrence*, 539 U.S. at 567.

26 *See id.* at 578 ("The present case . . . does not involve whether the government must give formal recognition to any relationship that homosexual persons seek to enter.").

27 *See* McLaughlin v. Florida, 379 U.S. 184, 195 (1964) (striking down the Florida statute but refusing to reach the "question of the validity of the State's prohibition against interracial marriage").

28 *See* Perry v. Schwarzenegger, 704 F. Supp. 2d 921, 996 (N.D. Cal. 2010) ("Proposition 8 operates to restrict Perry's choice of marital partner because of her sex.").

29 *See* McLaughlin v. Florida, 379 U.S. 184 (1964).

30 517 U.S. 620 (1996).

31 *Cf. Perry*, 704 F. Supp. 2d at 996 ("Proposition 8 also operates to restrict Perry's choice of marital partner because of her sexual orientation.").

32 *See Romer*, 517 U.S. at 632 ("[I]t lacks a rational relationship to legitimate state interests. . . .").

33 *See* Lawrence v. Texas, 539 U.S. 558, 580 (2003).

34 *See, e.g.*, Kerrigan v. Comm'r of Pub. Health, 957 A.2d 407 (Conn. 2008) (subjecting orientation discrimination to heightened scrutiny).

35 *See, e.g.*, Hernandez v. Robles, 855 N.E.2d 1, 7 (N.Y. 2006).

36 *Id.* at 30 (Kaye, C.J., dissenting).

37 Perry v. Schwarzenegger, 704 F. Supp. 2d 921, 972 (N.D. Cal. 2010) ("Permitting same-sex couples to marry will not affect the number of opposite-sex couples who marry, divorce, cohabit, have children outside of marriage or otherwise affect the stability of opposite-sex marriages.").

38 *Id.* at 973 ("Proposition 8 places the force of law behind stigmas against gays and lesbians.").

39 *Id.* at 971 (noting that "the cultural meaning of marriage and its associated benefits are intentionally withheld from same-sex couples in domestic partnerships").

40 *Id.* at 970 ("California domestic partnerships may not be recognized in other states and are not recognized by the federal government.").

41 *Cf. id.* at 968 (noting that California promotes gay and lesbian parenting).

Does Neutrality Equal Secularism? The European Court of Human Rights Decides *Lautsi v. Italy*

*By William L. Saunders**

Introduction

Religion can be an intensely personal activity. However, the idea that religion is *only* a private, personal devotion with no public political consequences is relatively new. For many nations in Europe, religion, in particular Catholicism, exerted an important influence over government and politics for centuries. The remnants of this influence still remain in anthems, oaths, and ideologies, not to mention architecture. However, with the rise of an ideology of "strict separation of church and state" in the European Union and the Council of Europe, it has been unclear how countries may incorporate their religious influences and histories into public life and expression. The case of *Lautsi v. Italy* in the European Court of Human Rights illustrates this struggle between secular ideology and religious faith and affiliation in the European context. The ultimate decision in the case acknowledges that "freedom of religion" need not result in, as the late Richard John Neuhaus put it, the naked public square.[1]

The Italian Case

The case was originally filed in Italy by an Italian national, Ms. Soile Lautsi, who sued on her own behalf and on the behalf of her two school-age sons, who were students at an Italian public school. The school, like all public schools in Italy, had a crucifix prominently displayed in its classroom. This concerned Ms. Lautsi because Italy had a program of mandatory education. She believed that the display of the crucifix was depriving her of the right to raise her children as she believed best.

The crucifix has been displayed in Italian schools for more than a hundred years. Though the practice is not enshrined in the Italian Constitution, a number of decrees and circulars have mandated the practice. For instance, in 1861, Article 140 of the Kingdom of Piedmont-Sardinia's Royal Decree no. 4336 required all schools to display a crucifix. In 1871, Law no. 214 of 13 May 1871 created a formal relationship between Rome and the Kingdom of Italy and granted the Catholic Church a number of rights and privileges. The fascist and monarchical governments also propagated a number of similar rules between 1922 and 1929. The Italian State in 1985 stripped the Catholic Church of its title as the official religion with Law no. 121. However, Italy did not repeal any of the previously-mentioned laws requiring the display of crucifixes in schools.[2]

Ms. Lautsi brought her concerns to school officials at a meeting in 2002. She demanded that the officials remove the crucifixes from the classrooms. The school officials denied her request. Several years later this decision would be officially codified with the 2007 Ministry of Education Directive no. 2666, which recommended that all classrooms display crucifixes.

Because her request was denied, Ms. Lautsi brought her case to the Veneto Regional Administrative Court. The court, however, felt that the case dealt with a constitutional question. Therefore, it referred the case to the Italian Constitutional Court. The Constitutional Court chose not to rule on the case because the regulations on crucifixes were contained in *statutes*, rather than the constitution of Italy.[3]

Thus, the case was remanded to the Administrative Court. In its ruling, the Administrative Court determined that, as symbols of Italian history and principles, it was reasonable to display the crucifix in classrooms. Though the court did note that the crucifix had a religious connotation, it deemed the crucifix's use as a secular representation of Italian history and culture outweighed any possibly oppressive religious influence. The court also determined that the presence of the crucifix in no way constituted indoctrination or hindered Ms. Lautsi's ability to raise her children as a secular parent.[4] Ms. Lautsi's subsequent appeal to the Consiglio di Stato, or Supreme Administrative Court, resulted in a similar decision.[5] At the conclusion of this case, Ms. Lautsi had exhausted all courses of action available to her under Italian law.

Lautsi v. Italy (I)

Under the European Convention on Human Rights (hereafter, the Convention), the European Court of Human Rights (hereafter, the ECHR or the Court) has jurisdiction over cases dealing with religious freedom. One limitation of the Court's power lies in its jurisdictional reach. Cases can only be brought to it if all other legal options *within* the member state have been exhausted. Because Ms. Lautsi had concluded her case in Italy with the Supreme Administrative Court, she was able to bring her case to the ECHR.

Ms. Lautsi rested her argument on Article 2 of Protocol No. 1 of the Convention, which reads in part, "No person shall be denied the right to education. In the exercise of any functions which it assumes in relation to education and to teaching, the state shall respect the rights of parents to ensure such education and teaching in conformity with their own religious and philosophical convictions."[6] She also argued that display of the crucifix was a violation of Article 9. That article reads as follows:

> 1) [E]veryone has the right to freedom of thought, conscience, and religion; this right includes freedom to change his religion or belief, and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching practice and observance. 2) Freedom to manifest one's religion or beliefs shall be subject only to such limitations are as prescribed by law and are necessary in a democratic society in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others.[7]

**William L Saunders is a human rights attorney working in Washington, D.C. with a nonprofit organization. He is chair of the Federalist Society's Religious Liberties Practice Group.*

Ms. Lautsi argued that the crucifix represented solely a religious symbol with no other historical value. Thus, by displaying the crucifix in public schools, Italy was giving a preference to Christianity and hindering other religious perspectives among children. In response, the government argued that crucifixes as they were used in the schools were not primarily reflections of religion but, instead, represented the cultural heritage of the Italian nation. The government also argued that the placement of the cross could not influence schoolchildren in any manner that constituted a violation of the Convention because, at its heart, it was a passive symbol.

In this first argument before the European Court there was one notable third party that was granted leave to intervene. The Greek Helsinki Monitor argued for Ms. Lautsi.[8] The main thrust of its argument was that the sign of the cross could only be considered a religious symbol that existed as an implicit "teaching of religion." The Monitor argued that this display constituted indoctrination because students in schools could feel that the government supported one religion over others.

The ruling by a panel, or Section, of the ECHR was in Ms. Lautsi's favor. First, the Court determined that "the State [was] forbidden to pursue an aim of indoctrination that might be considered as not respecting parents' religious and philosophical convictions."[9] The panel also held that this required that the member States recognize

> an obligation on the State's part to refrain from imposing beliefs, even indirectly in places where persons are dependent on it or in places where they are particularly vulnerable. The schooling of children is a particularly sensitive area in which the compelling power of the State is imposed on minds which still lack (depending on the child's level of maturity) the critical capacity which would enable them to keep their distance from the message derived from a preference manifested by the State in religious matters.[10]

However, even though the Court recognized the principle that European states were not allowed to impose beliefs, it still had to determine whether Italy's display of the crucifix constituted an imposition of a particular religious belief. However, this the Court did with relative ease. It determined that "the presence of the crucifix may easily be interpreted by pupils of all ages as a religious sign, and they will feel that they have been brought up in a school environment marked by a particular religion."[11] The Court further held that it

> considers that the compulsory display of a symbol of a particular faith in the exercise of public authority in relation to specific situations subject to governmental supervision, particularly in classrooms, restricts the right of parents to educate their children in conformity with their convictions and the right of schoolchildren to believe or not believe.[12]

The Court also addressed the argument of Italy that the crucifixes were passive symbols. It said that "[the Court] cannot see how the display in state-school classrooms of a symbol that it is reasonable to associate with Catholicism (the majority religion in Italy) could serve the education pluralism which is essential for the preservation of 'democratic society' within the Convention meaning of that term."[13] Essentially, the Court dismissed the idea that symbols can be publicly displayed without also actively indoctrinating students in a classroom.

The more difficult argument the Court had to deal with was whether the crucifix was predominantly a cultural symbol. The Court acknowledged that it had approved the use of certain religious symbols, and other paraphernalia, when the item was part of a nation's cultural heritage. However, in this case, "the court considers that the presence of the crucifix in classrooms goes beyond the use of symbols in specific historical contexts."[14]

The Court distinguished between religious symbols with significant historical relevance and religious symbols whose historical value was outweighed by their religious significance. To explain this distinction, the Court referenced another case of a similar nature: *Buscarini v. San Marino*. In that case, members of the parliament of the Republic of San Marino, a State member of the Council of Europe, were required to take an oath in which they swore "upon the holy gospels." The Court determined that this violated the Convention: "[T]he traditional nature, in the social and historical sense, of a text used by members of parliament when swearing loyalty did not deprive the oath to be sworn of its religious nature."[15] Like the oath used in the Parliament in the *Buscarini* case, the crucifix in Italy could not be emptied of its essentially religious nature regardless of their tradition of use.

The Court concluded that what was required was "confessional neutrality." According to the Court, "confessional neutrality" implied that all displays of an overtly religious symbol in public schools by the state violated the religious protections of the Convention.[16] As would be noted in reaction to the decision, this principle would seem, logically, to require the removal of all religious symbols in public places.[17]

Interestingly, the Court ignored the concept of "margin of appreciation." This doctrine is traditionally applied by the ECHR to member states and allows the laws and regulations within states, particularly with regard to religion and social policy, to differ. However, in the *Lautsi (I)* case, the Court did not address the question of the "margin of appreciation" to be granted to Italy in this matter, and instead imposed a blanket ban on crucifix display.[18]

Reaction to *Lautsi (I)*

The European Court of Human Rights is dependent on the governments of the member states to uphold and enforce its rulings. However, the negative response to the *Lautsi (I)* decision was vigorous, marked, and widespread. A number of prominent Italian politicians, family organizations, and religious groups responded immediately. The Vatican's response noted, "[I]t seems as if the court wanted to ignore the role of Christianity in forming Europe's identity, which was and remains essential. . . . The crucifix has always been a sign of God's love, unity and hospitality to all humanity."[19]

Poland, Lithuania, and Slovakia openly decried the decision. The Greek Orthodox Church and the Romanian Orthodox Church condemned the Court's decision. Kyrill, Patriarch of Moscow, released a statement in which he said:

Christian religious symbols present in Europe's public space are part of the European identity, without which the past, the present and the future of this continent are unthinkable. The guaranteeing of a secular nature of the state must not be used as a pretext for infusing an anti religious ideology that conspicuously breaches peace in society and discriminates against Europe's religious majority—Christians.[20]

Each of these churches encouraged their affiliated organizations to speak out strongly for rules allowing the displays of religious symbols in schools. Some groups created signs, and others petitioned governments and lobbied for legislation protecting religious symbols in schools.

The Italian government also responded vehemently to the decision of the Court. A number of ministers and high-ranking officials, including Education Minister Mariastell Gelmini, spoke openly about their opposition to the decision.[21] Some state and local officials openly refused to consider removing the crucifixes. Local governmental bodies also spoke about the influence of Catholic teaching and practices in everyday Italian culture and deplored what they saw as the effort of the ECHR to remove this influence.

The Italian judiciary's response to the decision is also of note. A serious legal criticism of the decision was that it was not adequately deferential to state practices under the margin of appreciation (explained above). The Constitutional Court of Italy stated that any decision by an international court that was clearly opposed to the practices, policies, and heritage of the Italian state would not be binding upon the nation.[22]

Nine more countries, including Albania, Austria, Croatia, Hungary, Moldova, Poland, Serbia, Slovakia, and Ukraine, also "openly criticized the initial judgment and petitioned that the Court remember that it must respect the national identities and religious traditions of each of the 47 member States."[23] Including Italy, nearly half of the member states of the Council of Europe supported the crucifixes and presented an unparalleled unification in support of traditional religious practices and symbols.[24]

Lautsi v. Italy (II)

The Court had been unprepared for the response their decision elicited. The Italian government petitioned for a rehearing of the *Lautsi* case by the Grand Chamber.[25] This petition was more than a simple request for a rehearing; it indicated a formal and decisive dissent by the Italian government. Perhaps unsurprisingly, given the strong reaction to the first decision by the ECHR, the Grand Chamber granted the request for rehearing quickly.[26]

Lautsi II, as the case came to be known, witnessed an unprecedented number of amicus briefs filed. The Court granted requests to intervene to thirty-three members of the European Parliament, as well as to non-governmental organizations, the Greek Helsinki Monitor, the Associazion nazionale de libero Pensiero (National Association of Free Thought), the European Center for Law and Justice, Eurojuris, the International Committee of Jurists, Interights and Human Rights Watch, the Zentralkomittee der Deutschen Katholiken (Committee for German Catholicism), Semaines sociales de France (Social weeks of France), and the Associazioni cristiane lavoratori italini (Italian Christian Workers Association). The ECHR also granted leave to file a joint brief to the governments of Armenia, Bulgaria, Cyprus, the Russian Federation, Greece, Lithuania, Malta, Monaco and the Republic of San Marino.[27] In total, thirty briefs were filed, thirteen on behalf of Lautsi and seventeen on behalf of the Italian government.

The number of amicus briefs was the most ever filed in the Court. Commentators noted the *Lautsi* case, which had "such a large and unified reaction from the Member States[,] [was] simply unprecedented at the Court."[28]

At the hearing, the government argued that neutrality, as required by the Convention, did not equate with secularism. The government suggested that "the Court should acknowledge and protect national traditions and the prevailing popular feeling, and leave each State to maintain a balance between opposing interests."[29] The government emphasized that a number of other faith-based symbols were welcomed in Italian schools. For example, headscarves are freely worn.[30] Therefore, according to Italy's position, no student or their parents were prohibited from practicing, or not practicing, any religion that they desired.

Lautsi, by contrast, argued that the state was showing an overt preference for Catholicism by requiring that crucifixes be displayed in public schools. She again claimed that this preference kept her from educating her children in accordance "with her own philosophical convictions."

The arguments made by the amicus briefs on the side of Italy mostly stated that the lower chamber had misunderstood the concept of neutrality, which the chamber had "confused with secularism. . . . State symbols inevitably had a place in state education. . . . Many of these had a religious origin, the Cross—which was both a national and a religious symbol—being the most visible example." Additionally, many of the briefs signaled the belief that the Court had "exceeded the scope of the application and limits of its jurisdiction" by creating an "obligation to ensure that the educational environment was entirely secular."

One portion of the oral arguments of *Lautsi II* was particularly memorable. Joseph Weiler, who argued on behalf of the States, including Armenia, Bulgaria, Cyprus, Greece, Lithuania, Malta, Russia, and San Marino, that intervened in the case on the side of Italy, did so wearing a yarmulke.[31] He argued on behalf of the eight nations that "[i]n all our countries freedom of religion and freedom from religion must be respected. However, it is counterbalanced with considerable liberty which the convention system allows, as to the place of religion and religious heritage and religion symbols in the definition of the collective identity of the nation and the state and its public spaces."[32] He also noted that, as a state with an established state church and members of the legislature which are also members of the church, "England would appear to violate the strictures of the chambers. For how could we say that all those symbols, which I mentioned, the head of state, the head of the church, the cross, the anthem etc. do not represent some kind of assessment of the legitimacy of religious belief?"[33] He emphatically noted that it was for the people of a country to choose to remove religion from the public sphere, and not for the European Court of Human Rights. Both his own public

display of religious affiliation and arguments were well-received by the Court.

By a vote of fifteen to two, the Grand Chamber overturned the prior decision. Unlike the lower chamber, the Grand Chamber focused heavily on the idea of giving the States a broad margin of appreciation. The Court noted that States "enjoy a wide margin of appreciation in determining the steps to be taken to ensure compliance with the Convention with due regard to the needs and resources of the community and of individuals."[34] The Court stated: "[I]n principle it is not for the Court to rule on such questions, as the solutions may legitimately vary according to the country and the era."[35]

Additionally, many of Lautsi's arguments were based on her subjective belief that the crucifixes had infringed her rights. However, the Court noted that subjective feelings of infringement alone were not enough to constitute actual violation of the Convention. The decision stated: "[T]he applicant's subjective perception is not in itself sufficient to establish a breach."[36] Instead, the Court required that there be actual and tangible violations of a parent's right to bring their child up in the religion (or non religion) of their choice.[37] Since it appeared Lautsi had not been hindered in teaching secularism to her children, there had been no tangible violation.

Though the Court could have stopped here, it went further. It stated:

> [T]here is no evidence before the Court that the display of a religious symbol on classroom walls may have an influence on pupils and so it cannot reasonably be asserted that it does or does not have an effect on young persons whose convictions are still in the process of being formed.[38]

To reach this conclusion, the Court relied heavily on the evidence presented by the government to indicate the tolerant nature of the Italian schools.[39] It was the opinion of the Court that, since the schools were openly tolerant of other religious symbols and celebrated other religious holidays, the crucifixes were merely an additional expression of a religion that featured prominently in Italy's history. Therefore, because of the religious toleration in the classrooms, the crucifixes could not infringe on Lautsi's rights as a parent to educate her children in any way she saw fit.[40]

Finally, the Court was quick to find a historical value to the crucifixes. The court stated, "Beyond its religious meaning, the crucifix symbolized the principles and values which formed the foundation of democracy and western civilization."[41]

Concurring and Dissenting Opinions

There were a number of concurring opinions to this decision. Perhaps the most interesting was that of Judge Giovanni Bonello. His opinion was a scathing retort to the lower chamber and Ms. Lautsi. He stated, "[T]he Court ought to be ever cautious in taking liberties with other peoples' liberties, including the liberty of cherishing their own cultural imprinting. Whatever that is, it is unrepeatable. Nations do not fashion their histories on the spur of the moment."[42] He noted that it was unreasonable to remove all forms of religious influence on schools in favor of secularism. For example, he noted that the school calendar was based around religious holidays like the Lord's Day, Christmas, and Easter. He suggested that it would be outrageous to "suppress" the school calendar merely because it was determined around days of religious observation.[43] He noted that, like the crucifixes, the school calendar had never shown a tendency toward indoctrinating students or negatively influencing the practice of their religions.[44]

Judge Bonello agreed with the main argument of the government: the lower chamber wrongly equated "freedom of religion" with "secularism." For example, Ms. Lautsi had the freedom to behave and profess, or not profess, any religion she saw fit, and it was this freedom, rather than secularism, which was protected by the articles she appealed to in the Convention. Bonello also rejected the idea that the removal of the crucifix would be "neutral":

> The crucifix purge promoted by Ms. Lautsi would not in any way be a measure to ensure neutrality in the classroom. It would be an imposition of the crucifix-hostile philosophy of the parents of one pupil, over the crucifix-receptive philosophy of the parents of all the other twenty-nine. If the parents of one pupil claim the right to have their child raised in the absence of a crucifix, the parents of the other twenty-nine should well be able to claim an equal right to its presence, whether as a traditional Christian emblem or even solely as a cultural souvenir.[45]

The Future of *Lautsi*

Two lessons can be drawn from the *Lautsi* decisions. First, there is a sizeable number of European countries and peoples that support displays of religious affiliation in public spaces. These nations consider their majority religion and its related symbols to be a part of their cultural heritage. They joined the Council of Europe based upon the understanding that member States were permitted to follow these traditions (under the margin of appreciation). Their view is that freedom of religion, as guaranteed in the Convention, rather than undermining this view, supports it. They do not equate, as our Supreme Court has, "neutrality" with secularism or "the naked public square." This has obvious significance for discussions and debates in other fora as to the meaning of the international human right of religious freedom.

Second, the strong negative reaction by many member States to the decision in *Lautsi I* has important implications for the force and effect of future pronouncements of the European Court of Human Rights. These member States have signaled that they will not acquiesce in the kind of judicial decisions that have often been made in the United States.

Endnotes

1 Richard John Neuhaus, The Naked Public Square (2d ed. 1997).

2 A full discussion of the laws, regulations, and Italian case law relevant to the *Lautsi* Decision are available at the European Court of Human Rights HUDOC website: *Lautsi v. Italy*, http://cmiskp.echr.coe.int/tkp197/view.asp?item=2&portal=hbkm&action=html&highlight=lautsi&sessionid=74121335&skin=hudoc-en.

3 Corte Cost, 15 Dicembre 2004, n. 389 (It.).

4 Giun. Pro. Ammin, 17 Marcia 2005, n. 1110 (It.).

5 Cons. Stato, 13 Febbraio 2006, (It.).

6 Convention for the Protection of Human Rights and Fundamental Freedoms Protocol 1 art. 2, opened for signature Mar. 20, 1952, *available at* http://conventions.coe.int/treaty/en/treaties/html/009.htm.

7 Convention for the Protection of Human Rights and Fundamental Freedoms art. 9, opened for signature Nov. 4, 1950, 213 U.N>T.S. 221, Europ T.S. No. 5., *available at* http://www.hri.org/docs/ECHR50.html#C.Art 9.

8 A group created to protect human rights in Europe, in particular to oversee any violations of the Helsinki Final Act.

9 Lautsi v. Italy, App. No. 30814/06, Eur. Ct. H.R. 47 (2009) [hereinafter *Lautsi I*], *available at* http://www.echr.coe.int/ECHR/homepage_en [hereinafter Lautsi (I)].

10 *Id.* at 48.

11 *Id.* at 55.

12 *Id.* at 57.

13 *Id.* at 56.

14 *Id.* at 52.

15 *Id.* at 52 (citing Buscarini v. San Marino [GC], no. 24645/94, Eur. Ct. H.R. 1999-I).

16 *Id.* at 56.

17 *Archbishop Peter Smith Speaks on* Lautsi *Case*, Indep. Catholic News, June 23, 2010, *available at* http://www.indcatholicnews.com/news.php?viewStory=16363.

18 The Government of Italy did address the question of "margin of appreciation." *Lautsi I*, App. No. 30814/06, at 37-41.

19 Gabriel Andreescue & Liviu Andreescue, *The European Court of Human Rights'* Lautsi *Decision: Context, Contents, Consequences,* J. for Study Religions & Ideologies, Summer 2010, at 47 [hereinafter Andreescue] (citing *Italy School Crucifixes "Barred": The European Court of Human Rights Has Ruled Against the Use of Crucifixes in Classrooms in Italy,* BBC News, Nov. 3, 2009, *available at* http://news.bbc.co.uk/2/hi/8340411.stm.

20 *Russian Patriarch Protests Court Ruling to Ban Cross from Italian Schools,* Interfax (Religion), Nov. 26, 2009, *available at* http://www.interfax-religion.com/?act=news&div=6675.

21 Thiago Alves, *A Tale of Two Courts—Part Two: The Reactions to the* Lautsi v. Italy *Chamber Judgment*, Hum. Rts. F., May 2, 2011, *available at* https://humanrightsforum.wordpress.com/2011/05/02/a-tale-of-two-courts-%E2%80%93-part-two-the-reactions-to-the-lautsi-v-italy-chamber-judgment/.

22 Lautsi v. Italy*: Crucifixes in the Classroom and "Dialogue" with Strasbourg (Italian Style),* Education Law Blog, http://www.education11kbw.com/?p=263 (Mar. 29, 2011, 16:18 GMT).

23 *ECHR Crucifix Case: 20 European Countries Support the Crucifix*, European Centre for Law and Justice, europeNEWS, July 21, 2010, *available at* http://eclj.org/Releases/Read.aspx?GUID=983c3dd3-9c17-4b70-a016-37851446ec0e&s=eur.

24 *Id.*

25 An appeals process within the European Court of Human Rights.

26 Since the Grand Chamber is composed of all judges on the Court while a panel is composed of a lesser number, this review by the Grand Chamber is essentially equivalent to an en banc re-hearing of a panel's decision by all members of a U.S. circuit court.

27 These countries were also granted the right to intervene in oral arguments.

28 Gregor Puppinck, Director of the European Centre for Law and Justice, Presentation Prepared for the 2010 Annual International Law and Religion Symposium (Oct. 3-6, 2010).

29 Lautsi v. Italy, App. No. 30814/06, Eur. Ct. H.R. 37(2011), *available at* http://cmiskp.echr.coe.int/tkp197/view.asp?item=1&portal=hbkm&action=html&highlight=Lautsi%20%7C%20v.%20%7C%20Italy&sessionid=74114939&skin=hudoc-en [hereinafter *Lautsi II*].

30 *Id.* at 39.

31 Austen Ivereigh, *Waiting on* Lautsi, Am.: Nat'l Catholic Wkly., July 23, 2011, *available at* http://www.americamagazine.org/blog/entry.cfm?entry_id=4030.

32 *Joseph Weiler's Testimony Before the European Court of Human Rights*, http://dotsub.com/view/65bc5332-aa10-4b8c-bc50-d051e8f4fcc7 (last visited Oct. 28, 2011).

33 *Id.*

34 *Lautsi II*, App. No. 30814/06 at 61.

35 *Id.* at 62.

36 *Id.* at 61.

37 *Id.* at 66.

38 *Id.*

39 *Id.* at 74.

40 *Id.*

41 *Id.* at 67.

42 Lautsi v. Italy, App. No. 30814/06, Concurrence of Judge Giovanni Bonello, Eur. Ct. H.R. 1.5 (2011), *available at* http://cmiskp.echr.coe.int/tkp197/view.asp?item=1&portal=hbkm&action=html&highlight=Lautsi%20%7C%20v.%20%7C%20Italy&sessionid=74114939&skin=hudoc-en.

43 *Id.* at 1.6.

44 *Id.*

45 *Id.* at 3.6.

Book Reviews

Rebuilding the Ark: New Perspectives on Endangered Species Act Reform

Jonathan H. Adler, Editor

*Reviewed by James S. Burling**

In 2005, Congressman Richard Pombo engineered the passage of the most sweeping reform of the Endangered Species Act (ESA) since it was passed in 1973. HR 3824 would have required more workable habitat restoration and better peer review science for listings. Most intriguingly, it contained a compensation mechanism that would have rewarded landowners for maintaining endangered species habitats rather than the current practice of punishing landowners with a massive devaluation of their land values. While it passed the House with bipartisan support, it failed in the Republican-controlled Senate. To thank Representative Pombo for his efforts, the environmental community labeled Pombo an "eco-thug" and flooded his district with attack ads and volunteers in order to ensure his defeat at the 2006 election.

In many segments of the environmental community, the notion of touching the ESA is akin to skinning baby harp seals alive. So it is with some boldness that Jonathan Adler, a Professor of Law and Director of the Center for Business Law and Regulation at Case Western Reserve University School of Law, has pulled together a collection of essays centered around proposals to reform the ESA.

Why reform the ESA? After all, it has been around for decades, and several industries, most notably in the Pacific Northwest and California's central valley, have complained that it has been used as a tool to their destruction. The Act also supports a cottage industry of environmental lawyers—both those in favor of returning the earth to Gaia by any means necessary and those who have a more anthropocentric world view. But aside from these dubious accomplishments, has it actually saved any species? Is it doing more harm than good for plants and animals it is supposed to protect? Do we actually have a clue what the real state of most threatened and endangered species is? Is whatever good it is doing worth the "at any cost" mandate of the Act.[1] Is there a better way?

The essays in this volume attempt to answer these questions, especially the last one. Sadly, we really do not know the answers to most of these questions. Whether the ESA has saved any species may depend on what we mean by "saved." Has the ESA allowed the "recovery" of a meaningful number, or at least a nonzero number, of species? Or has it prevented the slide of species into the abyss of extinction? By the recovery standard, most acknowledge that the ESA hasn't done much. But the ESA's defenders posit that it has met the "slide into the abyss" standard—though this is more through supposition than any hard evidence. After all, we don't have a spare Earth handy to test the efficacy of the ESA against the parallel universe Earth that lacks an ESA.

On whether the ESA does any *harm* to endangered and threatened species, there have always been whispered, but for obvious reasons, largely unverified tales of landowners who deal with their endangered species "problem" with the "shoot, shovel, and shut-up" trifecta. But there are more plausible, and documented, stories of landowners "preplanning" for the arrival of endangered species by rendering land unfit for nonhuman habitation. Owners of Southern pine plantations are thought to be harvesting trees early and before the trees are mature enough to develop cavities that red-cockaded woodpeckers are wont to interpret to be an "open house" invitation.

As to whether we even know enough about the state of most species, the answer is clearly not because the science is incomplete and access to much of America's land is restricted. Various levels of government own over one-half of the nation's land mass, where access by government biologists is reasonably easy. But, for now, that leaves one-half of the nation's land mass—and habitat—in private hands. And those private hands are not very keen on inviting NGO and government biologists onto their property to look for species that are or might become endangered—after all, a positive finding could ultimately sterilize the use and value of the land. More importantly, if we are to be serious about protecting species, then protecting them on private land is essential. But landowners are reluctant to cooperate so long as that means drastic and uncompensated reductions in the use and value of their land.

So, what to do? The nine essays in this compilation each focus on a different problem and a potential solution. While in agreement that the current regime is lacking in its efficacy, the range of solutions is diverse. Ranging from tax relief to free market reforms and to more workable regulatory programs, the common theme of most of the essays is that there must be a better way than regulatory fear and loathing to encourage landowners to preserve and even improve habitats for endangered species.

Northwestern Professor David Dana suggests we improve the process of creating Habitat Conservation Plans (HCPs). HCPs began as a reform from the Clinton Administration that sought landowner cooperation in preserving ecosystems for multiple species in exchange for regulatory certainty. It was then, and remains today, a creative interpretation of the ESA, and any major changes will need statutory authorization. Large-scale HCPs have often been beset with political controversy as multiple landowners have sought to protect their interests, sometimes at the expense of other landowners. Dana's primary criticism of the current HCP process is that the process is less than transparent and there is no standard or reliable measurement of success or even compliance. Congress should, Dana contends, at a minimum mandate a complete database on existing HCPs, mandate the collection of meaningful information, and mandate compliance reporting. Next there should be mandatory review by a scientific advisory board. He proposes that in order to encourage landowners to agree to meaningful biological goals, we should institute an insurance program to protect against a "conservation-failure." Finally,

* *James S. Burling is the Director of Litigation for Pacific Legal Foundation based in Sacramento, California. For more, see www.pacificlegal.org.*

where a smaller scale program is needed, Dana suggests conservation banking could be a more viable alternative. What Dana does not address, is whether the HCP process, existing or as imagined, will provide enough incentives for landowners to voluntarily and readily enter into the process. It is one thing to "encourage" landowners to join because of a fear that the heavy hand of government could become heavier; it is another to actually provide enough incentives so that landowners will actually desire to join HCPs.

Texas A&M Professor of Wildlife Neal Wilkins picks up on the need to provide more landowner incentives. He points to the example of landowners in Texas who may wish to contribute to efforts to preserve the lesser prairie chicken, but may have reservations because of an ongoing boom in wind farms. Reforms could include more in the way of "recovery crediting" wherein landowners who make positive contributions to a species' recovery can be rewarded by landowners who need to affect other habitat. Land use lawyers are quite familiar with the concept of transferable development credits—including the fact that many of them are little more than glorified shell games where some landowners are required to compensate others for takings that might otherwise be assessed to government. If recovery crediting is to be a meaningful reform, it will need to avoid the skepticism engendered by TDR programs.

Wilkins has some additional innovative suggestions. In order to foster more landowners' cooperation with information gathering, he suggests that enforcement functions of government be separated from the science, monitoring, and recovery duties. He also suggests that NGO third parties be authorized to work with private landowners. While not all landowners trust the NGOs, they may well trust some NGOs more than government agents. Other reforms Wilkins proposes are more in the way of market-based conservation programs and more defined recovery goals when species are listed.

In the wake of the Tellico Dam controversy, the ESA was amended to allow for a so-called cabinet level "God Squad" to grant exemptions and "incidental-take permits" to allow for some activities to proceed, even if they might impact an endangered species. Pennsylvania State law professor Jamison Colburn characterizes these amendments, designed to add some flexibility in the ESA, as "notorious," as is pretty much anything that requires meaningful consideration of costs. Colburn suggests instead some alteration in our understandings of the line between permits and property. However, it is uncertain that Colburn's ideas will readily translate into policy prescriptions (assuming that were a desirable outcome) in an essay replete with sentences like, "Yet, even supposing unprecedented computational or coordinative breakthroughs were to make globally scaled cognition practicable, we will still face the normative frictions generated when political power is limited by a polity's democratic traditions and geographic boundaries."[2] Not only is the rhetoric obtuse, but the suggestion leaking through these words—that to save species we must transcend democracy and national sovereignty—is not likely to gain traction in the near term.

Another commonly used mechanism for enlisting landowners' cooperation in species protection is through tax-deductible donations of conservation easements. But there well may be an inefficient allocation of resources with this practice. To a rancher, losing the ability to use 100 acres through a conservation easement may have the same economic consequences whether the habitat is extraordinarily valuable to a critter or simply of marginal biological utility. And because the economic consequences are the same, the government's tax expenditure in allowing the deduction will be the same. In other words, the rancher writes off the same amount in each case. While a receiving entity will be happy to take both marginal and valuable habitat, should government pay the same amount for both?

Emory University School of Law Professor Jonathan Remy Nash has a better idea: "[T]he value of the donation of a conservation easement [should be] based not upon the economic value of the donated easement but rather upon the value of the easement to the ecosystem."[3] This would skew the incentives such that landowners may have added incentive to improve habitat in order to increase its value *to the landowner*. Shoot, shovel, and shut-up could be replaced by restore, improve, and donate. While Nash admits that the valuation of land from economic utility to ecosystem utility may be difficult, it should not prove to be impossible. As with any new proposal, Nash also admits that it may be difficult to craft a program that isn't too costly or that doesn't have unintended consequences.

Unasked and unanswered by Nash is the related vexing and somewhat philosophical question of how much land should ultimately be encumbered. We are entering a brave new world where the utility of vast holdings of land are being stripped from the fee in perpetuity (for to be tax-deductible, easements must be perpetual). While Nash's proposal makes great sense in terms of better targeting government tax expenditures, and it beats the notion that oppressive land use regulation is the best way to achieve ecosystem preservation, it leaves unanswered what the final destination of this journey ought to be. How much land can the nation afford to remove forever from productive use? Further, the common law has always been resistant to attempts of one generation to control the resources of future generations. Will this attempt fare any better?

Today in the Central Valley of California there is a new water war. In the Klamath Basin there has been a water war for over a decade. Unlike previous water wars between ranchers and farmers, or between rural and urban interests, this one is between fish and people. Or, perhaps more accurately, there is war between people who value fish for ecological and commercial purposes and people who value water more for urban and agricultural purposes. Unlike prior water wars fought with guns or *Chinatown* intrigue, this one is being fought with biological opinions and lawsuits. Professor James Huffman at Lewis and Clark Law School understands well the difficulties of creating positive ecosystem incentives among water users who, at present, are feeling rather put upon. And the challenges of water rights, creatures of state law (some would say archaic state laws) but respected by federal law and, more importantly, protected by the federal constitution, is fiendishly complex.

There are several water-rights based challenges to the implementation of the ESA being litigated now in the courts. Huffman argues that the Takings Clause is the most substantial challenge, but that "a strong takings clause does not necessarily

obstruct achievement of the species protection objectives of the ESA."[4] This is not because, Huffman argues, property rights in water are or ought to be malleable (meaning capable as some argue of being defined out of existence). Instead, Huffman suggests, there needs to be *better* understanding of water rights, an understanding that allows greater marketability—such as with water transfers and a greater ability to allocate water to conservation purposes without risking the loss of rights under the regime of "use it or lose it" that is common in many Western states. Huffman concludes that the magnitude of the water wars can be reduced—at least from "all-out warfare to isolated skirmishes—if both sides take a more practical and less principled approach."[5] So long as there is weather—and too much rain falls in one place and not enough in another—people will fight over water. Huffman is optimistic that out of today's controversy we will reach an accommodation that will serve both fish and man; let us hope he is right.

Science and politics are like the East and West. Rudyard Kipling once wrote of the East and West that "never the twain shall meet." But like the East and West in modern times, science and politics are inextricably entwined. The biological sciences are used to justify what are essentially political land use questions. And politics are used to determine whether science is "junk" or gold-plated and peer-reviewed. But because the stakes are so high, both landowners and species advocates have tremendous incentive to ensure that science falls their way. Science also has its limitations. We can only know so much given our current state of knowledge and availability of resources to put into science. In his short piece on science and the ESA, economic consultant Brian Mannix puts a face on the extraordinary burden being placed on science to answer essentially unanswerable questions. For example, EPA has an obligation to consider the impacts of pesticide registrations on endangered species that could "provide *millions* of potential obligations to consult with the [federal regulatory] Services—each, based on experience, taking as much as ten years."[6] Mannix has a few suggestions to get us out of this mess, first and foremost of which is to distinguish between science and policy. In other words, make the ESA more like the National Environmental Policy Act, which demands an analysis of the impacts of a federal action—but does not mandate what should be done with that information. Thus the result of an environmental impact statement is to give federal agencies an option to change course, not to determine the course. The same would be the case in Mannix's new ESA. While eminently sensible and practical, Mannix's proposal to change the basic structure of the ESA may be about forty years too late. No one in Congress wants to be the next Richard Pombo.

It has become an article of faith with many that planet Earth is entering an unprecedented epoch of warming and we must act, and act quickly, to reverse anthropogenic global warming. How this can be achieved without putting an end to Western civilization (and Eastern civilization as well) is anybody's guess. But one way that will not work according to Florida State School of Law Professor J.B. Ruhl is a full-court press played by team ESA. Ruhl has no doubt that the crisis is real, but plenty of doubt that the ESA provides a workable solution. As he puts it, over the years the ESA has proven to be the pit bull of environmental statutes. But when it comes to global warming, he says this pit bull won't fly. Yes, Ruhl says, global warming will have a profound and largely devastating effect on species around the world. But the legal tools of the ESA were simply never designed to shut down emissions of carbon dioxide.

Some of the inherent flaws in using the ESA to combat global warming are being fought tooth and claw with the polar bear listing. The Fish and Wildlife Service listed the bear as threatened because of the potential impact warming will have on sea ice, which the bear uses for summer foraging. Logically, any federal action in any part of the United States that causes an emission of a greenhouse gas could now be made subject to the "consultation" requirement of the ESA with the whole panoply of action-stopping consequences. But that, to the chagrin of the ESA lawyers, was a bridge too far for the Bush II Administration, and it issued a ruling that the listing could *not* be used to trigger consultations in any state but Alaska. But to prove Ruhl's point about the limitations of the ESA, this was not simply a product of the so-called anti-environment Bush Administration. When given a chance to reverse, President Obama did no such thing.

As Ruhl puts it, the stop-carbon "mitigation litigation charge is leading the ESA away from its central mission of conserving ecosystems."[7] Its mission is suited well enough for "what is happening on the ground and in the water . . . rather than being concerned with what is happening in the troposphere."[8] The ESA could be modified, Ruhl suggests, to play a more meaningful and realistic role in combating the effects of climate change. These would include a specific category of listing for climate-threatened species and replacing the goal of species recovery with one of assisting the transition to a warmer climate—recognizing that some species may do better at the expense of others during the transition. But unlike the ESA of the past, which Ruhl calls "both noble and arrogant," he suggests instead that "the ESA must become noble and humble if it is to have any chance of helping species through the era of climate change."

Michael De Alessi, currently a post-doc scholar at Stanford who has long experience in environmental policy battles, concludes this book with a look at the interrelationship between the ESA and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Evolved from a conservation effort at the turn of the last century designed to protect megafauna from decimation through unrestricted poaching and trade, CITES restricts or prohibits trade in species from around the globe. But it is not an all-or-nothing proposition. As De Alessi notes, there have been some great successes where CITES has allowed the commercial utilization of species on the brink such as the Nile crocodile and African elephants. Commercial ranching of these species has brought their numbers back from the brink. Of course, this is not a panacea. The replacement of natural ecosystems with ranches is not the end goal, but can serve as a placeholder while ecosystems are restored.

But the ESA has the ability to list any species anywhere in the world. And the federal agencies sometimes do. But is this useful to the species? De Alessi thinks not. While a

listing may stop trade into the United States, and it may, in theory, discourage agencies from funding habitat-harming infrastructure projects, a listing under the ESA has no legal effect outside our borders. But it can hurt species. De Alessi notes that there once used to be 100,000 green sea turtles being ranched on the Cayman Islands for export to Europe. But after they were listed, it became illegal to transship them through the United States, once a necessary step to reach Europe. The farm is no more, having been replaced by a small, government-run eco-tourist operation with far fewer turtles.

There are other examples. As De Alessi points out, once a species is listed export licenses will be denied unless it can be proven that a commercial operation will *enhance* a species. This standard has stopped captive breeders of three African antelope species which are endangered in their native ranges. The ESA does nothing to protect foreign species or habitat, De Alessi contends. Without providing native villages a legal economic incentive to coexist with endangered fauna, especially valuable fauna like black rhinos, villagers might as well poach them. After all, it is hard to instill an environmental ethic in people who are many miles south of the poverty line.

Professor Adler has done a marvelous job collecting the essays in this book. Some are provocative, some are practical, and all are necessary to the debate about where we should go next with the protection of threatened and endangered species. The status quo has been played out. If the protection of species is to advance, the rules of the game need to be changed. And we'd better start recognizing that so long as a substantial percentage of habitat is on private land, landowners need to be encouraged rather than bludgeoned into working for the betterment of species.

Endnotes

1 Tennessee Valley Authority v. Hill, 437 U.S. 153 (1978).

2 Pg. 100.

3 Pg. 123-24.

4 Pg. 143.

5 Pg. 159.

6 Pg. 168-69.

7 Pg. 191.

8 *Id.*

Practice Group Publications Committees

Administrative Law & Regulation
Christian Vergonis, *Chairman*
Daren Bakst, *Vice-Chairman*

Civil Rights
Brian T. Fitzpatrick, *Chairman*

Corporations, Securities & Antitrust
Michael Fransella, *Chairman*
Carrie Darling, Robert T. Miller
& Robert Witwer, *Vice Chairmen*

Criminal Law & Procedure
Ronald J. Rychlak, *Chairman*
William A. Hall, Jr. & Peter Thomson, *Vice Chairmen*

Environmental Law & Property Rights
Donald Kochan, *Chairman*

Federalism & Separation of Powers
Ilya Somin, *Chairman*
Geoffrey W. Hymans, *Vice Chairman*

Financial Services & E-Commerce
Charles M. Miller, *Chairman*

Free Speech & Election Law
Reid Alan Cox, *Chairman*
Francis J. Menton, Jr., *Vice Chairman*

Intellectual Property
Mark Schultz, *Chairman*
Steven M. Tepp & Adam Mossoff, *Vice Chairmen*

International & National Security Law
Nathan Sales, *Chairman*
Soraya Rudofsky, *Vice Chairman*

Labor & Employment Law
James A. Haynes, *Chairman*
Raymond J. LaJeunesse, Jr., *Vice Chairman*

Litigation
Margaret A. Little & Christopher C. Wang
Co-Chairmen

Professional Responsibility & Legal Education
John J. Park, Jr., *Chairman*

Religious Liberties
William Saunders, *Chairman*

Telecommunications & Electronic Media
Raymond L. Gifford, *Chairman*

The Federalist Society for Law and Public Policy Studies is an organization of 40,000 lawyers, law students, scholars, and other individuals, located in every state and law school in the nation, who are interested in the current state of the legal order. The Federalist Society takes no position on particular legal or public policy questions, but is founded on the principles that the state exists to preserve freedom, that the separation of governmental powers is central to our constitution, and that it is emphatically the province and duty of the judiciary to say what the law is, not what it should be.

The Federalist Society takes seriously its responsibility as a non-partisan institution engaged in fostering a serious dialogue about legal issues in the public square. We publish original scholarship on timely and contentious issues in the legal and public policy world in an effort to widen understanding of the facts and principles involved and to continue that dialogue.

Positions taken on specific issues in publications, however, are those of the author, not reflective of an organization stance. Engage presents articles, white papers, speeches, reprints, and panels on a number of important issues, but these are contributions to larger ongoing conversations. We invite readers to submit opposing perspectives or views to be considered for publication, and to share their general responses, thoughts, and criticisms by writing to us at info@fed-soc.org. Additionally, we happily consider letters to the editor.

For more information about The Federalist Society, please visit our website: www.fed-soc.org.