E
N
G
A
G
E

Project of the
E.L. Wiegand
Practice Groups

The Journal of the Federalist Society Practice Groups

Volume 9, Issue 2 June 2008

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.

Engage

Volume 9, Issue 2

Letter from the Editor...

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 is all too often found in today's legal community. We expect that members will find some of the articles in these pages controversial. We expect disagreement on some issues, and welcome the controversy. But we also welcome responses. Articles can be sent to the editor for review at paigner@fed-soc.org.

Here we would also like to draw attention to the newest feature on our website (www.fed-soc.org): "Global Governance Watch" (www.globalgovernancewatch.org). Launched by the International Law and Sovereignty Project, in partnership with the American Enterprise Institute, this site will focus on four areas: (1) National Security, or the need for increased transparency and accountability b ecause the United Nations and other international organizations are taking more positions relevant to national security; (2) Human Security, or the broadening definition of human security from "freedom from fear" to include "freedom from want," and the repercussions this has on national sovereignty; (3) Development, or the efforts of various international nongovernmental organizations as their focus shifts from economic development to include social development; and (4) Global Regulation, or the growing networks that are being used by the United Nations to regulate economic and social affairs on a global basis, and the consequences that this might have for national sovereignty. We hope members find that this new feature, as well as other recent additions to the website, such as "Originally Speaking" and "SCOTUScast," provide the same high standard of scholarship and excellence that people have come to expect from the Federalist Society.

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 9, Issue 2 ENGAGE *June 2008*

Administrative Law and Regulation

Swallows Holding Ltd. v. Commissioner: Limited Progress in Rejecting Tax Exceptionalism in Administrative Law

*By Kristin E. Hickman**

Administrative law jurisprudence is an acknowledged mess. Following its development and application involves a lot of banging one's head against the wall. Yet, while application of administrative law doctrines is often "enshrouded in considerable smog,"[1] many of the governing rules and standards are relatively settled. For example, there is no question that agencies promulgating legislative rules must follow the public notice and comment requirements of the Administrative Procedure Act (APA).[2] Courts and scholars struggle to define the precise boundaries of the legislative rule category,[3] but courts have little difficulty concluding that an agency rule with clear legal effect, binding regulated parties and the government alike, and carrying congressionally imposed penalties for non-compliance, qualifies as legislative.[4] Similarly, especially after the Supreme Court's decision in *United States v. Mead Corp.*,[5] there is no question that reviewing courts should apply the strong, mandatory deference doctrine of *Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc.*,[6] rather than the less deferential standard of *Skidmore v. Swift & Co.*,[7] when evaluating legislative rules,[8] even if courts and scholars debate *ad nauseum* the various parameters, facets, and contours of *Mead*, *Chevron*, and *Skidmore*.[9]

If only administrative law doctrine were so settled with respect to Treasury regulations interpreting the Internal Revenue Code ("I.R.C."). Treasury utilizes two types of delegated authority in promulgating Treasury regulations. Many substantive provisions of the I.R.C. authorize Treasury to issue regulations to accomplish particular, congressionally specified goals;[10] but most Treasury regulations are adopted through the exercise of a more general grant of rulemaking authority in I.R.C. § 7805(a), which authorizes Treasury to develop "all needful rules and regulations for the enforcement of" the I.R.C.[11] The tax community generally recognizes specific authority Treasury regulations as legislative in character, and thus subject to the procedural requirements of APA § 553 and entitled to *Chevron* deference. Yet, whatever authority Treasury exercises in promulgating regulations interpreting the I.R.C., taxpayers who fail to follow Treasury regulations in preparing tax returns and paying taxes are subject to congressionally imposed penalties.[12] Accordingly, virtually everyone in the tax community agrees that general as well as specific authority Treasury regulations carry the force and effect of law.[13] Further, the Internal Revenue Service (IRS) claims that it utilizes the APA's notice and comment requirements in promulgating Treasury regulations.[14] However, for reasons of tradition based on a now-anachronistic understanding of the non-delegation doctrine, the tax community routinely uses the legislative and interpretative labels to distinguish specific authority Treasury regulations from general authority ones.[15] Hence, the IRS also claims that most Treasury regulations are interpretative rules, exempt from notice-and-comment rulemaking;[16] the preambles to most Treasury regulations disclaim the applicability of APA § 553;[17] and Treasury and the IRS routinely fail actually to comply with APA rulemaking requirements.[18]

Meanwhile, courts and tax scholars are also divided over the appropriate standard for judicial review of general authority Treasury regulations. Shortly after *Mead*, the Sixth Circuit declared outright that *Chevron* deference applies to general as well as specific authority Treasury regulations.[19] Prior to *Mead*, however, other circuit courts declared only a less deferential, multi-factor standard articulated prior to *Chevron* in the tax-specific *National Muffler Dealers Association v. Commissioner*[20] appropriate for general authority Treasury regulations.[21] Like *Skidmore*, *National Muffler* lists several criteria for courts to consider in deciding whether to defer to Treasury and IRS interpretations of the I.R.C.; but like *Chevron* after it, *National Muffler* also emphasizes Congress's delegation of administrative authority over the I.R.C. to Treasury and the IRS.[22] Hence, still other circuit courts maintained that *Chevron* and *National Muffler* are indistinguishable.[23] Scholars in the area run the gamut as well, with some in favor of *Chevron*,[24] others for *Skidmore*,[25] and still others searching for compromise by way of some *Chevron*/*National Muffler* hybrid or otherwise modified *Chevron* for general authority Treasury regulations.[26]

Enter *Swallows Holding, Ltd. v. Commissioner* into this quagmire.[27] *Swallows Holding* is a classic case of Treasury and the IRS using their congressionally delegated but general rulemaking authority under I.R.C. § 7805(a) to promulgate a regulation interpreting ambiguous language in another, more substantive provision of the tax code. According to I.R.C. § 882, foreign corporations engaged in trade or business in the United States are taxed much like U.S. corporations on income connected with the conduct of that trade or business—at graduated rates, after the income is reduced by corresponding offsets for deductible expenses connected with such income, with possible alternative minimum tax exposure.[28] I.R.C. § 882(c)(2) allows a foreign corporation to claim deductions against its U.S.-sourced gross income "only by filing... a true and accurate return, in the manner prescribed in subtitle F" of the I.R.C., which subtitle contains procedural provisions governing the filing of tax returns.[29] Exercising its general rulemaking

---

* *Kristin E. Hickman is an Associate Professor of Law at the University of Minnesota Law School. Parts of this essay draw heavily from research documented in prior work. See generally* 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 (2007), *and* Kristin E. Hickman, *The Need for Mead: Rejecting Tax Exceptionalism in Judicial Deference*, 90 Minn. L. Rev. 1537 (2006).

authority under I.R.C. § 7805(a), Treasury promulgated Treas. Reg. § 1.882-4(a)(3)(i), which interprets I.R.C. § 882(c)(2) to allow a foreign corporation to claim offsetting deductions against its U.S.-sourced income only if the corporation files its tax return "in a timely manner," designated specifically as within 18 months of the return's due date.[30] In so doing, the regulation cross-references timing rules for filing tax returns contained in I.R.C. §§ 6072 and 6081 and related Treasury regulations.[31]

The taxpayer in *Swallows Holding* was a foreign corporation that realized rental income from real property that it held in San Diego, California, and elected to treat its rental activity as a U.S. trade or business.[32] The taxpayer filed tax returns claiming deductions under I.R.C. § 882(c)(2) after the 18 month filing period specified by Treas. Reg. § 1.882-4(a)(3)(i) had expired; the IRS denied the deductions, citing the regulation; and the taxpayer challenged the regulation's substantive validity.[33] The interpretive question was whether Congress's use of the word "manner" without a corresponding reference to "time" in I.R.C. § 882(c)(2) allowed Treasury to impose a limitation period for claiming deductions under I.R.C. § 882(c)(2).[34] The Tax Court concurred with the taxpayer in concluding that I.R.C. § 882(c)(2) as worded did not permit Treasury to impose the 18 month time limit on deduction claims.[35]

The judges of the U.S. Tax Court were sharply divided over how to evaluate the regulation at issue in *Swallows Holding*. A majority of the court concluded that the plain meaning of the statute precluded the timely filing requirement imposed by Treas. Reg. § 1.882-4(a)(3)(i). Nevertheless, the court also expounded at some length regarding the character of general authority Treasury regulations and the appropriate standard for reviewing them. Adhering to tradition, but with little further explanation, the majority labeled Treasury regulations promulgated pursuant to I.R.C. § 7805(a) as interpretative in character.[36] Correspondingly, though suggesting that the two standards are roughly equivalent with only "possible subtle distinctions," the majority concluded that the "traditional, *i.e., National Muffler* standard" rather than *Chevron* applied.[37] Considering various factors drawn from *National Muffler*, the Tax Court majority reached the alternative holding that Treasury's interpretation of I.R.C. § 882(c)(2) was unreasonable.[38]

Extensive dissenting opinions from Judges Swift, Halpern, and Holmes rejected the majority's conclusions regarding the standard to be applied and the deference due to Treas. Reg. § 1.882-4(a)(3)(i). All three judges found the language of I.R.C. § 882(c)(2) ambiguous and advocated applying the *Chevron* standard to defer to Treasury's interpretation thereof as reasonable.[39] With the agreement of Judges Halpern and Swift, Judge Holmes in particular argued that *National Muffler* and *Chevron* represent different standards, and he rejected the continued vitality of *National Muffler* in light of *Chevron*.[40] Furthermore, in reaching his conclusion, Judge Holmes offered substantial analysis debunking the significance of the tax community's historic practice of characterizing general authority Treasury regulations as interpretative: observing that such regulations "are intended to bind the public and have the force of law;" noting that the Court has acknowledged regulations promulgated by other agencies pursuant to similar general authority grants in other statutes as *Chevron* eligible; and concluding that it is not "possible to draw distinctions between the deference owed tax regulations issued under section 7805(a) and those issued under more specific authority."[41]

The Tax Court's majority and dissenting opinions in *Swallows Holding* thus neatly reflected the circuit court and scholarly debate over whether the tax-specific judicial deference standard articulated in *National Muffler* or the more general *Chevron* deference standard applies to general authority Treasury regulations. They also raised the question of the character of such regulations. On appeal, the Third Circuit addressed at least the first of these questions directly.

The Third Circuit's opinion in *Swallows Holding* took two clear and unequivocal positions regarding judicial deference in the tax context. First, the court makes plain its belief that *Chevron* and *National Muffler* represent distinct and, to some extent, incompatible standards. The court recognized that the Supreme Court and lower courts have cited *National Muffler* as requiring Treasury regulations and rules to be reasonable—"a proposition that is not at odds with *Chevron*'s core teachings."[42] Considering *National Muffler* more particularly as requiring judicial evaluation of several factors, however, the court rejected as "not mandatory or dispositive inquiries under *Chevron*" at least two of those factors—contemporaneity and congressional reenactment—along with giving weight to earlier judicial interpretations of ambiguous I.R.C. provisions.[43] Second, to the extent that *Chevron* and *National Muffler* yield different results, as the court indicated they would in this case, the court held that *Chevron* controls the outcome.[44] To reach the second of these conclusions, the Third Circuit applied the standard articulated by the Court in *Mead*: asking whether Treas. Reg. § 1.882-4(a)(3)(i) carries the force and effect of law.[45] It is at this pivotal point, however, that Third Circuit unfortunately truncated its analysis.

Consistent with the Tax Court's majority, the taxpayer argued that Treas. Reg. § 1.882-4(a)(3)(i) is an interpretative rule rather than a legislative one, and as such was per se ineligible for *Chevron* deference.[46] Under general, rather than tax-specific, deference principles, that would mean that the less deferential (and arguably more like *National Muffler*) *Skidmore* standard was appropriate.[47] However, rather than addressing the taxpayer's characterization of general authority Treasury regulations, as the different Tax Court opinions had done, the Third Circuit dodged that question. Instead, in applying *Mead* to decide between *Chevron* or *Skidmore*, the court decided that *Chevron* applied principally because the government put the regulation at issue through public notice and comment, "a move that is indicative of agency action that carries the force of law."[48] Thus ends the court's reasoning for why *Chevron* rather than *Skidmore* applies.

Because the *Mead* Court expressly mentioned notice-and-comment rulemaking as an indicator of *Chevron*'s applicability, the lower courts often seem to regard notice and comment as synonymous with *Chevron*'s applicability.[49] In most cases they are probably right. Agencies typically utilize the notice-and-comment process because the legal force of their regulations requires adherence to those procedures. Compliance with notice and comment thus often serves as a convenient proxy for *Mead*'s inquiry into whether regulations carry the force and effect of law. Yet, the *Mead* opinion clearly states that notice-and-comment rulemaking is only an indicator of the congressional delegation necessary for *Chevron* deference, and thus is neither an absolute precondition for *Chevron* deference nor a means of obtaining *Chevron* deference in the absence of the requisite delegation.[50] The real question under *Mead* is not whether regulations were promulgated using notice-and-comment rulemaking but rather whether they carry legal force. As noted, there is little doubt that Treasury regulations promulgated under I.R.C. §7805(a) regulations do. It is for this reason, rather than Treasury's utilization of notice and comment, that Treas. Reg. § 1.882-4(a)(3)(i) is entitled to *Chevron* deference.

It is also for this reason that Treasury and the IRS are wrong in their claim that most Treasury regulations are interpretative rules and, consequently, that notice and comment are optional therefore. By failing to address the Tax Court's disagreement over the characterization of general authority Treasury regulations generally or even Treas. Reg. § 1.882-4(a)(3)(i) specifically, the Third Circuit left unsettled as much or more than it resolved. Certainly, the question remains unresolved whether all Treasury regulations must satisfy APA rulemaking requirements. Given Treasury's position on that issue and its lousy record of compliance with APA rulemaking requirements, will *Chevron* deference apply to general authority Treasury regulations with APA compliance issues? Or will the courts evaluate the applicability of *Chevron* versus *Skidmore* to such regulations on a regulation-by-regulation basis? In the event a court holds that notice and comment are not required for general authority Treasury regulations, the Third Circuit's limited analysis in *Swallows Holding* provides an opening for taxpayers to argue, even before the Third Circuit, that courts should apply *Skidmore* rather than *Chevron* deference in reviewing the substantive validity of Treasury regulations.

Regardless of its flaws, however, the Third Circuit's opinion in *Swallows Holding* at least clearly repudiates the continued vitality of *National Muffler* as an independent, tax-specific evaluative standard for Treasury regulations. *Swallows Holding* thus represents a nail in the coffin of tax exceptionalism in judicial deference. In my view, that is progress.

Endnotes

1 Robert A. Anthony, *"Interpretive" Rules, "Legislative" Rules and "Spurious" Rules: Lifting the Smog*, 8 Admin. L.J. Am. U. 1, 4 n.13 (1994) (noting that courts' platitudinous use of the word "smog" as a catchword for certain administrative law perplexities).

2 *See* 5 U.S.C. § 553(b)-(c).

3 *See, e.g.*, Jacob E. Gersen, *Legislative Rules Revisited*, 74 U. Chi. L. Rev. 1705 (2007); John Manning, *Nonlegislative Rules*, 72 Geo. Wash. L. Rev. 893 (2004); Richard J. Pierce, Jr., *Distinguishing Legislative Rules from Interpretive Rules*, 52 Admin. L. Rev. 547, 556-57 (2000); Anthony, *supra* note 1.

4 *See, e.g.*, Sweet v. Sheahan, 235 F.3d 80, 91-93 (2d Cir. 2000); Hoctor v. U.S. Dep't of Ag., 82 F.3d 165, 169-72 (7th Cir. 1996). *Cf.* Am. Mining Congress v. Mine Safety & Health Admin., 995 F.2d 1106, 1109-11 (D.C. Cir. 1993) (designating criteria for assessing whether agency rule carries force of law and is thus legislative); Professionals & Patients for Customized Care v. Shalala, 56 F.3d 592, 595 (5th Cir. 1992) (articulating standard for identifying legislative rules based on binding effect on both public and agency).

5 533 U.S. 218 (2001).

6 467 U.S. 837 (1984) (holding that reviewing courts must defer to reasonable agency interpretations of ambiguous statutes under their administration).

7 323 U.S. 134 (1944) (articulating several factors that courts should consider in deciding whether to respect agency interpretations of statutes, including "the thoroughness evident in its consideration, the validity of its reasoning, its consistency with earlier and later pronouncements, and all those factors which give it power to persuade, if lacking power to control").

8 *See* Mead, 533 U.S. at 226-27, 230 n.12 (holding that *Chevron* provides the appropriate evaluative standard where Congress has given an agency the authority to bind regulated parties with "the force of law" and the agency has "exercised that authority," and identifying legislative rules as a prototypical example of *Chevron*-eligible agency action).

9 The judicial deference literature is simply too extensive to list. For just a few articles, *see* William N. Eskridge, Jr. & Lauren E. Baer, *The Continuum of Deference: Supreme Court Treatment of Agency Statutory Interpretations from* Chevron *to* Hamdan, 96 Geo. L.J. 1083 (2008); Kristin E. Hickman & Matthew D. Krueger, *In Search of the Modern* Skidmore *Doctrine*, 107 Colum. L. Rev. 1235 (2007); Thomas W. Merrill, *The* Mead *Doctrine: Rules and Standards, Meta-Rules and Meta-Standards*, 54 Admin. L. Rev. 807 (2002); Thomas W. Merrill & Kristin E. Hickman, Chevron*'s Domain*, 89 Geo. L.J. 833 (2001); Orin S. Kerr, *Shedding Light on* Chevron*: An Empirical Study of the* Chevron *Doctrine in the U.S. Courts of Appeals*, 15 Yale J. on Reg. 1 (1998); Ronald M. Levin, *The Anatomy of* Chevron*: Step Two Reconsidered*, 72 Chi.-Kent L. Rev. 1253 (1997); Cass Sunstein, *Law and Administration After* Chevron, 90 Colum. L. Rev. 2071 (1990); Antonin Scalia, *Judicial Deference to Administrative Interpretations of Law*, 1989 Duke L.J. 511.

10 *See, e.g.*, I.R.C. §§ 163(i)(5), 167(e)(6), 357(d)(3), 453(j)(1), 952(d), 1502 (2000). A search in the Westlaw FTX-USCA database for specific authority delegations derived from just one common phrasing, "'Secretary shall' /s prescribe /s regulations," resulted in 292 hits in May 2008.

11 I.R.C. § 7805(a).

12 *See* I.R.C. § 6662(a)-(b)(1) (imposing a 20% penalty when underpayment of tax is due to "negligence or disregard of rules or regulations"); Treas. Reg. § 1.6662-3(b)(2) (defining "rules or regulations" for this purpose as including all temporary or final Treasury regulations issued under the I.R.C.).

13 *See, e.g.*, Estate of Gerson v. Comm'r, 507 F.3d 435, 438 (6th Cir. 2007) (noting that both temporary and final general authority Treasury regulations are legally binding on taxpayers); Bankers Life & Cas. Co. v. United States, 142 F.3d 973, 979 (7th Cir. 1998) (observing that specific and general authority Treasury regulations have the force of law); Sheldon I. Banoff, *Dealing with the "Authorities": Determining Valid Legal Authority in Advising Clients, Rendering Opinions, Preparing Tax Returns and Avoiding Penalties*, 66 Tax Notes 1072, 1086, 1092 (1988) (noting legal effect of final and temporary Treasury regulations). *But see* Estate of Gerson v. Comm'r, 127 T.C. 139, 176 (2006) (Vasquez, J., dissenting) (opining that Treasury regulations issued under I.R.C. § 7805 do not carry the force of law).

14 *See, e.g.*, Internal Revenue Manual § 32.1.5.4.7.5.1(5); *id.* § 32.1.2.3(3).

15 *See* Kristin E. Hickman, *The Need for* Mead*: Rejecting Tax Exceptionalism in Judicial Deference*, 90 Minn. L. Rev. 1537, 1563-68 (2006) (tracing the historical origins of this practice).

16 *See id.* §§ 32.1.5.4.7.5.1(5); 32.1.2.3(3).

17 *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, 1749 (2007) (documenting results of three-year study of Treasury regulation projects and noting such disclaimers in 92.7% of the projects studied).

18 *See id.* at 1749 (noting that more than 40% of Treasury regulation projects over three-year period failed to satisfy APA rulemaking requirements). *Cf.* Michael Asimow, *Public Participation in the Adoption of Temporary Tax Regulations*, 44 Tax Law. 343, 369-70 (1991) (concluding that Treasury's routine adoption of temporary regulations with only post-promulgation notice and comment "leaves in doubt the validity of numerous temporary

and final regulations); Juan J. Lavilla, *The Good Cause Exemption to Notice and Comment Rulemaking Requirements Under the Administrative Procedure Act*, 3 Admin. L.J. Am. U. 317, 341 91989) (describing the IRS's comparative misuse of the good cause exception of APA § 553(b) as "particularly remarkable").

19 *See* Hosp. Corp. of Am. & Subsidiaries v. Comm'r, 348 F.3d 136, 140-41 (6th Cir. 2003).

20 440 U.S. 472 (1979). The Court's analysis in *National Muffler* reflects aspects of both *Chevron* and *Skidmore* deference, as the Court emphasized the significance of Congress's delegation of rulemaking authority to Treasury as well as listing several factors to be evaluated including Treasury's expertise and the need for consistency in tax law administration. *See id.* at 475-77.

21 *See, e.g.*, Snowa v. Comm'r, 123 F.3d 190, 197 (4th Cir. 1997); Nalle v. Comm'r, 997 F.2d 1134, 1138 (1993); *see also* Schuler Indus., Inc. v. United States, 109 F.3d 753, 754-55 (Fed. Cir. 1997) (calling *Chevron* more deferential than *National Muffler*).

22 *See National Muffler*, 440 U.S. at 475-77.

23 *See, e.g.*, Bankers Life & Cas. Co. v. United States, 142 F.3d 973, 978-83 (7th Cir. 1998) (acknowledging some differences but declaring the two standards practically indistinguishable); Tate & Lyle, Inc. v. Comm'r, 87 F.3d 99, 106 n.13 (3d Cir. 1996) (equating the two doctrines implicitly by citation).

24 *See* Hickman, *supra* note 15, at 1542. *Cf.* Mitchell M. Gans, *Deference and the End of Tax Practice*, 36 Real Prop. Prob. & Tr. J. 731, 792-93 (2002) (conceding *Chevron*'s applicability in light of *Mead* but advocating *Skidmore* for other reasons).

25 *See, e.g.*, John F. Coverdale, Chevron*'s Reduced Domain: Judicial Review of Treasury Regulations and Revenue Rulings after* Mead, 55 Admin. L. Rev. 39, 81-83 (2003).

26 *See, e.g.*, Irving Salem et al., *ABA Section on Taxation: Report of the Task Force on Judicial Deference*, 57 Tax Law. 717, 737-44 (2004); Ellen P. Aprill, *Muffled* Chevron*: Judicial Review of Tax Regulations*, 3 Fla. Tax Rev. 51, 82-84 (1996); Edward J. Schnee & W. Eugene Seago, *Deference Issues in the Tax Law:* Mead *Clarifies the* Chevron *Rule – Or Does It?*, 96 J. Tax'n 366, 371-72 (2002).

27 515 F.3d 162 (3d Cir. 2008); 126 T.C. 96 (2006).

28 I.R.C. § 882(a), (c)(1)(A).

29 I.R.C. § 882(c)(2).

30 Treas. Reg. § 1.882-4(a)(3)(i); *see also* Final Regulations, Untimely Filing of Income Tax Returns by Nonresident Alien Individuals and Foreign Corporations (T.D. 8322), 55 Fed. Reg. 50,827, 50,828 (Dec. 11, 1990), 1990-2 C.B. 172 (adopting relevant regulatory language).

31 *See* I.R.C. § 6072(c) (setting due date for filing income tax return of foreign corporation); *id.* § 6081(a) (authorizing rules and regulations for granting extensions of filing deadlings); Treas. Reg. § 1.6081-5(a)(3) (permitting six-month extension of time for filing income tax return of foreign corporation with office or place of business in the United States).

32 *See* Swallows Holding, 515 F.3d at 165; *see also* I.R.C. § 882(d)(1) (allowing a foreign corporation deriving income from real property in the United States to elect to treat such income as effectively connected with a U.S. trade or business).

33 *See* Swallows Holding, 515 F.3d at 165.

34 *See id.* at 165-66.

35 *See* Swallows Holding, Ltd. v. Comm'r, 126 T.C. 96, 132 (2006).

36 *See id.* at 129.

37 *Id.* at 131.

38 *See id.* at 137-43. Specifically, the majority drew from National Muffler six factors for its consideration in evaluating Treas. Reg. § 1.882-4(a)(3)(i): "(1) Whether the regulation is a substantially contemporaneous construction of the statute by those presumed to have been aware of congressional intent; (2) the manner in which a regulation dating from a later period evolved; (3) the length of time that the regulation has been in effect; (4) the reliance placed upon the regulation; (5) the consistency of the Secretary's interpretation; and (6) the degree of scrutiny Congress has devoted to the regulation during subsequent reenactments of the statute." *Id.* at 137.

39 *See id.* at 149 (Swift, J. dissenting); *id.* 157-62 (Halpern, J. dissenting); *id.* at 179-80 (Holmes, J. dissenting).

40 *See id.* at 162, 172-81 (Holmes, J. dissenting).

41 *Id.* at 176-80 (Holmes, J. dissenting).

42 *See* Swallows Holding, 515 F.3d at 168 n.6.

43 *See id.* at 167-68 & n.5.

44 *See id.* at 167-68.

45 *See id.* at 168.

46 *See id.* at 169.

47 *See id.*

48 *Id.* at 169-70.

49 *See, e.g.*, Wheatland Tube Co. v. United States, 495 F.3d 1355, 1360 (Fed. Cir. 2007); Sierra Club, Inc. v. Leavitt, 488 F.3d 904, 915 (11th Cir. 2007); Mount Royal Joint Venture v. Kempthorne, 477 F.3d 745, 754 (D.C. Cir. 2007). *Compare* BellSouth Telecommunications, Inc. v. Sanford, 494 F.3d 439, 447-48 (4th Cir. 2007) (applying *Skidmore* rather than *Chevron* deference to agency order promulgated with notice and comment because Congress did not delegate primary interpretive authority to issuing agency).

50 *See* United States v. Mead Corp., 533 U.S. 218, 229-30 (2001); *see also* Thomas W. Merrill, *The* Mead *Doctrine: Rules and Standards, Meta-Rules and Meta-Standards*, 54 Admin. L. Rev. 807, 814 (2002).

2008 Federalist Society Tax Policy Conference Panel: "How our Tax Laws Affect how Health is Paid for and Delivered"

Eileen J. O'Connor, Amy Monahan, Robert B. Helms & Michael F. Cannon

MS. O'CONNOR: Good morning everyone. I'm Eileen J. O'Connor. I'm a partner at Pillsbury Winthrop Shaw Pittman LLP, and it is my pleasure to welcome you this morning to the Federalist Society's tax policy conference, entitled "Our Nation's Founding Principles and Our Tax Code: Consistent or in Conflict?"

I will admit, those are actually my words, and the content of the conference is pretty much inspired by my experience over the past 30 years, as I have engaged in tax practice and have watched clients struggle with the Internal Revenue Code and how to comply with it, as they see other people try to figure out how to get out from under it, all the while weighing what the Internal Revenue Code really does against what, ideally, an Internal Revenue Code ought to do.

The 16th Amendment to the Constitution, ratified in 1913, says: "The Congress shall have the power to lay and collect taxes on incomes from whatever source derived, without apportionment among the several states, and without regard to any census or enumeration." That's pretty much all it says. But you don't have to be paying particularly close attention to the tax laws to appreciate that the Internal Revenue Code has become the repository not only of the rules for what revenues must be paid into the federal treasury, but also for many other rules, and we're going to talk about a few of those today.

Rather than coming right out and outlawing a behavior, lawmakers can provide a disincentive in the tax code. Similarly, rather than coming right out and providing a subsidy for a behavior, lawmakers can, and do, provide an incentive for it in the tax code.

Our first panel is about healthcare—how our tax laws affect how health care in our country is paid for and delivered. Our moderator is Professor Amy Monahan at the University of Missouri Law School, and she will introduce to you her panelists, Michael Cannon of the Cato Institute and Robert Helms of the American Enterprise Institute. Mark Pauly couldn't be with us today because of a death in the family, but he has some very valuable scholarship on his website, as he is a professor. So I encourage you, if this area interests you, to go there and read some of what he has written.

Our second panel today is about charity—whether and, if so, how our tax laws affect charitable activities, religious institutions, and free speech. Our moderator for that panel is Matthew Vadum, and our panelists will be Lee Goodman, Kevin Hasson and Anne Neil.

Our third and final panel is about tax expenditures. We will attempt to define that term, and discuss the wisdom and efficacy of using the Internal Revenue Code to implement social policy. I'll moderate that panel, and our panelists will be Lily Batchelder of the New York University School of Law, Leonard Berman of the Urban Institute, and Stephen Entin of the Institute for Research on the Economics of Taxation. [Video/audio recordings of the other panels can be found in our Multimedia Archive online at www.fed-soc.org.]

Without further ado, then, let us turn to our first panel. Our moderator is Amy Monahan, who is an Associate Professor of Law at the University of Missouri-Columbia School of Law. Ms. Monahan joined that faculty in 2004. Before she joined the faculty there, she had taught at Notre Dame and had also practiced law with Sidley Austin LLP in Chicago.

PROF. MONAHAN: First of all, thank you Lee and the Federalist Society for organizing this conference and having me here today. I'm really pleased that there's a panel at this conference on how the tax code affects health care, financing, and delivery. It's a fundamental issue, and I think one that all too often does not receive enough attention.

Before I turn things over to our two panelists, I'm going to give some brief introductory remarks. The topic of our panel today is how tax laws affect how health care is paid for and delivered. As many of you probably know, the answer is: they affect it fundamentally. The way we finance and deliver healthcare is really tax-code-driven, at least with respect to non-elderly Americans. I'm not going to go into the history of how we got where we are because Bob is going to cover that, but the regulatory system that we have in place now is one that few, if any, would defend in terms of regulatory theory. That said, it does have some benefits, which I'll talk about in a moment.

First: a brief summary of where we are today, for those that might not be engaged with the taxation of employer-provided health insurance.

There is a very significant tax preference given to employer-provided health insurance. Employers can deduct the cost of providing health insurance to their employees, just as they can deduct any other reasonable compensation expense. There's nothing special about that. But unlike other compensation that employers pay to employees, the amount an employer spends on employee health care is excluded from the employee's income for both federal income and payroll tax purposes. So, getting health care benefits from your employer is a much more advantageous arrangement than getting cash wages—because cash wages, of course, are taxed. As a result, employers are encouraged to contribute to the cost of health care they provide for employees: not just to set up a group plan but actually to contribute to it, instead of paying their employees cash wages.

There are other tax benefits as well. For example, the benefits paid from an employer-provided plan are excluded from the employee's income when paid. So, if they reimburse your hospital bill, that's not included in the employee's income. Employers can also set up a cafeteria plan under § 125 of the tax code to allow employees to pay their share of premiums for employer-provided health insurance on a pre-tax basis. As a result, both employer and employee contributions to health insurance can be excluded from an employee's income.

Author notes appear in text. This event took place on May 7, 2008.

Through a cafeteria plan, you can also set up a healthcare flexible spending account that allows you to pay out-of-pocket health care expenses on a pre-tax basis. Now, those are subject to various restrictions, the biggest one being the "use it or lose it" rule. You have to pre-commit to an amount that you set aside in your flexible spending account, and you either spend it on medical care or your employer gets that money back.

So, things are pretty good for employees that receive health care from their employer. Basically, everything can be tax-free. That stands in stark contrast to individuals who either do not receive health care through their employers or who are not employed. If you purchase health insurance outside of the employment context, you cannot deduct its cost unless you're self-employed. If you're self-employed, you can deduct 100% of the premiums.

Individuals can now also set up health savings accounts, which allow limited deductions for contributions to health savings accounts. Michael is going to speak in greater detail about health savings accounts, how they work, and how they might be reformed, but a health savings account requires, among other things, that an individual set up a high-deductible health plan and be covered only by that plan. However, the premiums associated with that plan cannot be paid on a pre-tax basis unless your employer is offering it, or unless you are self-employed. And finally, individuals can deduct their out-of-pocket medical expenses only to the extent they exceed 7.5% of the individual's adjusted gross income for the year. So there are significantly limited deductions for out-of-pocket health care expenses.

The bottom line is that anyone who has access to employer-provided health coverage usually elects that coverage if they want health insurance, because the tax code skews the economics of that decision in favor of employer-provided coverage. As a result, the majority of non-elderly Americans receive their health insurance through an employer. Additionally, because employer-provided coverage is subsidized by the federal government through tax provisions, we generally think that individuals end up with more health insurance than they would have, absent the federal tax preference. Because of the preference for employer-provided coverage over cash wages, employers offer more generous plans, and employees elect more generous plans, than they would absent the federal tax preference of those benefits.

Two big advantages of this system are based on the group purchasing model. First, group purchasing enjoys lower administrative costs. There's much lower overhead associated with purchasing health insurance as a group than in purchasing it as an individual, and this helps to control costs. The second significant advantage is the risk pooling function of employer groups. Employers are natural risk pooling groups for the most part, and that helps high-risk individuals afford coverage they otherwise would not be able to afford. There's a nice, kind of natural risk pooling there.

The third big advantage, and the one I will leave off with, is that it's easy for employees. One concern with insurance purchase on the individual market is that are tough decisions to make. If you're an individual looking at insurance policies, you have to evaluate a very wide range of factors and make decisions based on future, uncertain medical events. It's a difficult decision. The employer market is easy. It might not actually match your preferences, but it is easy. It's not cognitively taxing to choose an insurance plan from among the ones your employer offers. It's a nice, limited decision process, and we know from behavioral economics that can be a benefit, too.

But the list of disadvantages is longer, and that's probably why we're talking about this today. The disadvantages I would put into two main categories. There are both economic disadvantages and fairness issues. On the economic side, there are huge costs associated with the tax preference for employer-provided insurance. I think one of the handouts you received here today lists tax expenditures. You'll notice employer-provided health insurance is number one on that list. It is the most expensive tax preference item we have in the budget (which is why I think it's so important we're talking about that today). One thing to note is the handout you have refers only to federal income tax issues. We also lose money in the payroll tax system, because those benefits are exempt from payroll taxes. So it's a very big expenditure, a very big cost to us.

The other big economic disadvantage is that this tax preference encourages overspending in healthcare. There are incentives to elect generous insurance coverage. When people have generous insurance coverage, the rational thing to do is consume medical care. You paid for it through your premiums. Your plan is likely to have low cost sharing, so you have a small marginal cost when you decide to go to the doctor. A classic example is when you think you might have the flu. When deciding whether or not to go to the doctor—who you know probably can't help you—you are weighing a decision based on maybe a $20 co-pay. It could be as low as zero. In that case, you might as well go to the doctor—whereas, if you were evaluating that decision based on full cost of service, you might forgo it altogether. So, at least with respect to what I call discretionary medical services, the tax preference leads to over-consumption of medical care, and that's not good for anyone.

On the fairness side, there are several different issues. It's an exclusion from income, so obviously that's going to benefit those in the highest tax bracket. On the other hand, people with little or no income tax liability receive no benefit from the exclusion. And when we think about who we're trying to help or to subsidize—whose behavior we're trying to nudge if you will—it's hard to defend the exclusion on those grounds.

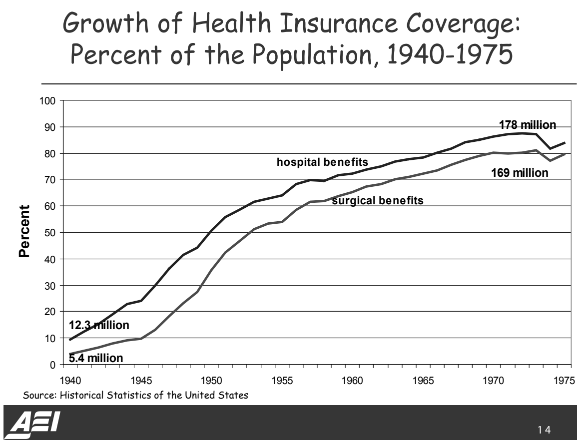
The last issue I want to mention is the disparate treatment. It's very hard to come up with an argument as to why people with the luck of working for an employer that offers health insurance should be able to purchase it on a pre-tax basis, but not those without such luck. This is the basis of an additional, and significant, fairness critique.

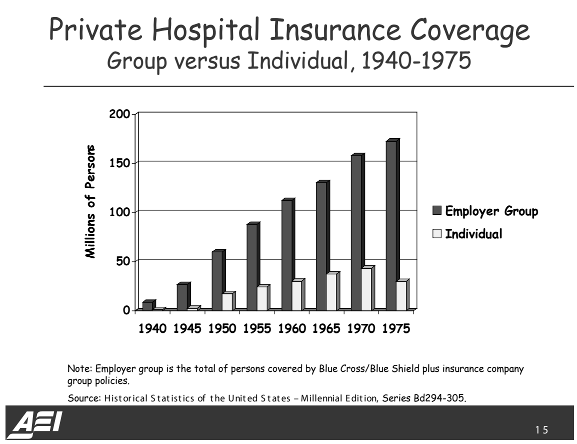
The list of disadvantages is significantly longer than the advantages. That isn't to say they necessarily outweigh them, but the result is that there's a great deal of interest in reforming how we tax health insurance. The big question, of course, is which path of reform one should take. Some people argue for leveling up—meaning, extending the favorable tax treatment to everyone, not limited to the employer context. Others argue for leveling down. Take away the preference from employer-provided insurance and put everyone on a level playing field.

Other proposals, (I would argue most proposals), lie somewhere in-between, and I think that's what we're going to see today.

There are no easy solutions here. First of all, the tax treatment of health insurance is actually pretty popular, because most people benefit from it. So there's a lot of political opposition to changing the tax treatment. And as we tinker with the tax treatment of health care, we have to contend very directly with how that affects state insurance markets, risk pooling functions of insurance, and uncertain outcomes.

We'll hear first from Bob Helms, Resident Scholar at the American Enterprise Institute. He's going to talk about the history of the tax treatment of health insurance, and some possible reforms. Bob has written and lectured extensively on health policy, health economics, and pharmaceutical economic issues. He has been widely published and has held various government positions in the healthcare industry. So without further ado, Bob Helms.





MR. HELMS: Thank you, Amy. That's was actually a very good introduction for my talk. I think this is an area that's greatly misunderstood—maybe it's more like ignored—in the broader health policy debate and in the media.

One little illustration of this is that almost every reporter you talk to about the tax treatment or health insurance uses the word "tax deduction" when what we're really talking about is a tax exclusion. So any time you read something in the paper about eliminating a tax deduction, you should question whether what they are talking about a tax deduction or a tax exclusion.

Health policy, even tax policy, is certainly in the political debate today. The last I heard, we still have two Democratic candidates. There is a lot of information about these plans on the Kaiser family website, "Health08." Hillary is proposing an individual mandate, and Senator Obama is proposing an individual mandate for children. Most of the experts in this area think that all of the candidates' proposals are still very vague. And whoever gets elected will have to face the separate issue of getting something passed. But I'll just point out that there are a lot of details in the Clinton and Obama plans already about heavy regulatory proposals for private health insurance.

McCain's proposal is even vaguer. So far, he's proposing to end the tax exclusion, but to give everybody, regardless of income, a tax credit: individuals $2,500 and families $5,000. Recently, he referred to the Guaranteed Access Plan. This is something he would have to work out with states, as kind of a fallback position that people could opt into. He also proposes to allow the interstate purchase of health insurance.

There are no official estimates of the cost of any of these plans. The estimates on the Clinton and Obama plans are really done by the campaigns themselves, not outside scorers.

I want to highlight what we health economists think we know about health and health insurance markets. First, we're talking about a service that a lot of people think is not a normal commodity—still, prices in this market matter. They matter to the buyers; they matter to the sellers. We also know that insurance, whether it's public or private, lowers the perceived price to the consumer. In other words, we have what's commonly called a moral hazard problem: increases in volume demanded when people don't pay the full price out-of-pocket.

On the supply side, the delivery of health care is very labor-intensive. It's mostly services. But we do have a lot of products that are always changing. So innovation and investment, etc., are important issues in this market. And we have large capital investments, particularly in hospitals, nursing homes, etc., which make quick adjustments more difficult.

But what's common about both public and private programs is that almost all the payment policies we have are open-ended in various ways. I'm not going to talk about Medicare and Medicaid; they have their own open-endedness. But an open-ended policy creates strong incentives to increase spending. In other words, you're spending somebody else's money, and so your incentive to be careful about it is much weaker. The result is what Clark Havighurst and others have called "flat of the curve medicine." Basically, we invest up to the point where the marginal returns are very low, sometimes negative.

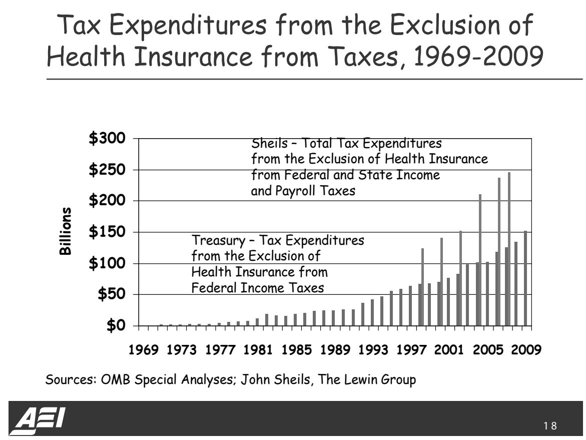
Now, I want to talk a little about history. The '30s, '40s, and early '50s were a period of great change in medical technology. Lots of new discoveries—in particular the development of penicillin, used with the troops in World War II, then made widely available after the war. This had the effect of making possible a lot of what we now consider to be routine surgery. Before that, it was more difficult because one could not keep down infection. So there were major advances not only on the drug side but also with regard to medical procedures. As a result, people began to value medical care more highly than they had in the '30s.

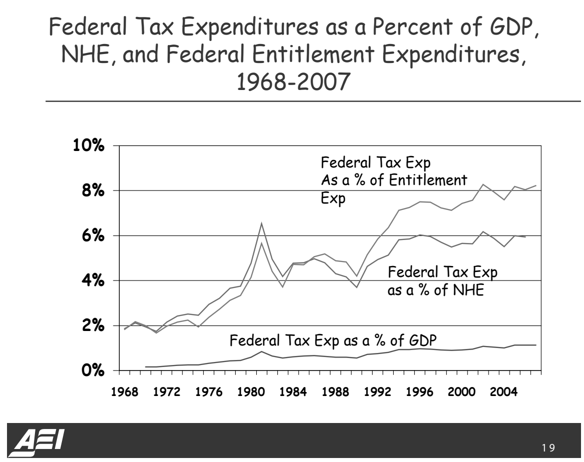
There are many statements in the literature that say something like, "In the 1930s the average physician could not affect the average condition of the average patient." That statement recognizes that there were lots of innovative things going on in medical schools, in the '30s. But the perception back then was, stay out of the hospital; it's a dangerous place. That's where people go to die. If you went there, you were probably going to get an infection. That changed, even by the '50s, so that people began to perceive of medical care as actually doing something good for you, even though it was expensive. This created a demand for insurance which allowed health insurance coverage to increase rapidly in the post-war period.

Early health insurance developed in the 1930s from early prepayment plans started in the Depression by hospitals seeking ways to get payment. These were organized by the American Hospital Association into the Blue Cross plans. Physician coverage developed later when the AMA helped to organize local plans into what became know as Blue Shield plans. The commercial insurance that we know today started in the private sector and were modeled after those plans. The AHA and the AMA went to the state legislatures and got a lot of legislation passed that made state Blue Cross and Blue Shield plans nonprofit and exempt from state premium taxes.

Then, World War II came along and with it a massive amount of economic planning and regulation. The government empowered a War Production Board to coordinate production of materials, and begin intensive government planning and price controls. The Office of Price Administration ran the rationing system and price controls of individual products. The National War Labor Board controlled wartime wages and attempted to settle labor disputes. In an effort to assure wartime production, they established many detailed rules and regulations regarding all aspects of hiring and paying employees. In 1943, the Board faced the problem of what to do about fringe benefits. In the end, they adopted the IRS rules and focused on controlling cash wages while exempting employer-provided fringe benefits, primarily pensions and health insurance. This was at a time when health insurance was relatively small and inexpensive.

After the war, the War Labor Board was disbanded but the IRS adopted the Board's policy of exempting fringe benefits from taxable income. It vacillated a little bit about whether employer-provided health insurance should be taxable or not, but the Congress finally stepped in and put this into legislation in 1954, saying that health insurance provided by the employer was to be excluded from taxable income. Now, that—the tax preference that Amy explained—combined with the increase in demand for health insurance led to rapid growth in health insurance coverage after the war. The post-war increase in the population, the influx of women into the marketplace, and income growth all combined to increase the demand for health insurance. The people who had hospital benefits increased from 12.3 million in 1940 to about 180 million in 1945; that is, nearly 90% of the population had health insurance of some kind (see chart 14).





Tax policy also influenced the form of health insurance. As you can see in chart 15, in terms of the millions of people covered by employer groups compared to individual coverage, the individual market did grow after the war. But the tax policy created a financial advantage for group, or employer-provided, coverage. Amy's reminder that there are advantages for risk pooling and so on also played a role in this. The tax policy was not the only reason for the growth of group coverage, but is provided the fertilizer to boost the growth of coverage in the employment sector.

Historical data from the 1960s to 2000 show that third party payments (also reflecting the growth of Medicare and Medicaid passed in 1965) increased relative to out-of-pocket payments. Out-of-pocket payments decreased from almost 60% to less than 20% from 1960 to 2000.

One of the conclusions I would like to leave you with today is that the present tax policy was somewhat of a historical accident. Still, it has played a large role in pushing health insurance into the employment sector and causing health insurance to evolve in an inefficient way.

Chart 18 shows two estimates of the growth of tax expenditures caused by federal tax policies that exclude employer-provided health insurance from taxable income. The measurement of tax expenditures is a controversial topic, but I use them because they are a convenient statistic and the only way I know to illustrate the effect of tax policy over time. The green bars plot the Treasury Department estimates over time and you can see they have grown. John Shiels at the Lewin Group has an alternative estimate that includes the effects of payroll taxes and state taxes. His estimate is almost up to $250 billion a year in 2007. Note that on the basis of the Sheils's estimates, total federal and state tax expenditures are increasing on average over $13 billion a year from 1998 throught 2006. If this lost revenue to the federal and state budgets had been from expenditure increases, it would most likely have received more attention.

Now, how does this compare to the economy? I've taken just those estimates of federal tax expenditures and compared them to three measures of economic activity (see chart 19). Tax expenditures are growing—relative to gross domestic product, as a percent of national health expenditures, and also as a percent of entitlement expenditures in the federal budget. So it's increasing relative to almost anything you put in the denominator. The effect of tax policy has been to push the demand for health services to a higher equilibrium than it would have been without the tax subsidy. To summarize, the effects of tax policy intensified the effects of increases in income, population, and medical technology; it expanded employment-based insurance relative to the individual insurance market; and it expanded insurance benefits, meaning it gave an incentive at the margin, so that people have more hospital coverage, more outpatient care, and eventually mental health, dental, and drug coverage, etc. Some people even include exercise programs now. And if you can get your employer to include these things, you get it tax-free. Another point to remember about the exclusion is that the higher your marginal tax rates, the higher the benefit is to you. Tax policy induced a higher level of costs, prices, and expenditures and created winners and losers.

Let me close with a comment on the health reform debate. This is oversimplified, but in my view you have two basic approaches to reform. You can take a tax reform approach—which, to me, is a necessary condition to the effective reform of incentives. If tax policy pushed us into an inefficient form of insurance, then it seems to me that we've got to change it somehow, maybe put a limit on it with a tax cap to give incentives for people to redesign these plans and put more emphasis on value and cost effectiveness. You could eliminate the tax exclusion, put on a tax cap, as in the 1940s during World War II. That was actually proposed under the Reagan administration, something I was involved in. Bush had a standard deduction proposal, and McCain has a tax credit proposal that would eliminate the tax exclusion but give people a deduction or a tax credit. Under this system you would have strong incentive to redesign the health insurance coverage. You'd have more research on cost-effectiveness. The market would then determine which kind of health policy would dominate, but you'd have, I think, more cost-effective options for small businesses, which is where a lot of the uninsured are now. This is a necessary, but not sufficient, condition for effective reform.

The alternative is the regulatory approach, more administered fee schedules like we have in Medicare, more global budgeting, as Hillary Clinton proposed in 1993. You'd have mandated benefits like both of the Democratic candidates are proposing. You could mandate coverage for individual employers. You'd have a lot more underwriting restrictions, price controls on insurance, and mandates on the pay-out from insurance, which to me is just a form of price controls. If we take the regulatory approach, it will be important to remember that it will be much easier to mandate insurance *coverage* than it will be to mandate that everyone have *access* to high quality and cost-effective care.

Thanks.

PROF. MONAHAN: Thank you, Bob. Our next panelist is Michael Cannon, the Cato Institute's Director of Health Policy Studies. Prior to joining Cato, he served as a domestic policy analyst for the U.S. Senate Republican Policy Committee and is widely published in health care policy.

MR. CANNON: Thank you, Amy. And Bob, I want to let you know that I'm on my own campaign to change the way people talk about not just health care but taxes generally—I'm trying to eliminate the term "tax expenditure" from our vocabulary. Tax expenditure is how we describe money the government leaves with you. We call it spending. They're also referred to as subsidies. I don't think you can spend money or subsidize someone unless you actually have that money in your hands and then give it to someone else. I fear the day the government begins to consider all of the money it lets us keep as a tax expenditure and then decides, "We're spending it here, let's just spend it somewhere else instead." Anyway, the campaign goes on.

I want to thank Amy and Bob for really laying out most of the important issues surrounding tax reform and health care. I would add that there are also exclusions for money put into a flexible spending account: a health reimbursement arrangement or a health savings account (HSA). These are all accounts that let the individual consumer control the money, but HSAs are really the only ones that create ownership because only with HSAs do you get a tax break for money that you own and can take with you from job to job. And there's also the tax deduction for qualified medical expenses in excess of 7.5% of adjusted gross income.

Once you add the loss of revenue under the payroll tax, the tax exclusion for employer-sponsored insurance is about two times that of the next biggest revenue loser, which is the mortgage interest deduction. The exclusion distorts the prices and costs when workers are making decisions about how much health insurance to buy, where to buy it, and how to pay for their medical needs generally.

It distorts the prices or the costs that workers face in three different ways. First, it distorts the relative cost of health versus non-health expenditures. You've got a tax break for employer-sponsored health insurance. That's going to reduce the cost of health care spending generally relative to non-health expenditures.

Second, it reduces the cost of third-party insurance relative to self-insurance, because if you decide to get less generous coverage through an employer—(your employer makes that decision)—and save a bit of money, or put aside those savings on premiums to help pay for your medical bills, traditionally that money was taxed. So it wasn't a level playing field between those two decisions, between premiums and savings. With HSAs and some of those other options, the playing field has been leveled somewhat, but not completely, because it only happens within a narrow range of premiums and deductibles, and it only works for people who purchase a qualified high-deductible health plan that's compatible with an HSA.

Third, the exclusion favors employer-sponsored insurance over other forms of third-party insurance. This could be insurance that you purchase on your own or through a group. The way that works is employer-sponsored insurance is favored to such an extent that if you decide to purchase insurance on your own rather than through an employer, depending on your health status and your tax bracket, you can end up paying twice as much for the same or less coverage—twice as much because you'd be paying higher taxes.

So what does this do to our health insurance market? Well, I think a lot of individuals end up getting stuck with insurance that doesn't meet their preferences. In the 1990s, as a response to the growing cost of health insurance, employers tried to move a lot of their workers into managed care. A lot of workers didn't like that. Right now, we're seeing a shift toward more cost-sharing in order to hold down the rising health insurance premiums. And I think we are seeing evidence of a similar backlash from consumers who would rather have less cost-sharing. So no matter what employers do they're going to step on some of their workers' toes.

As Bob and Amy mentioned, we end up encouraging people to obtain more coverage than they would if they were making decisions undistorted by the tax code. And they end up purchasing a lot of low value medical care. Economists have generally found that beyond a certain amount of health insurance, purchasing additional coverage doesn't deliver additional improvements in health. So that's one way of measuring value, but there are other ways that suggest that this excess coverage encourages consumers to purchase too much medical care.

It can also create inefficiencies by distorting people's labor market decisions. If you've ever heard of the term "job lock," that's when someone's got an employer-sponsored health insurance plan that they really need because they or someone in their family have a high cost condition, and they're afraid of moving to another job because that doesn't offer as generous coverage, or because they're afraid of retiring early and losing access to that coverage.

It also creates a lot of inequities, both horizontal and vertical. Say you've got two neighbors and they are identical in every way, except that one of them works for an employer that offers health insurance and the other does not. The one that does not ends up paying higher taxes because his health-insurance premiums are not tax-exempt. So it treats like people differently. A lot of people think the tax exclusion for employer-sponsored insurance creates vertical inequities as well. People with higher incomes get a larger tax break under the exclusion because they're in higher tax brackets. That's how deductions and exclusions work. People who are in lower tax brackets get smaller tax breaks, and because they're also less likely to work for an employer that offers employer-sponsored health insurance, they're less likely to get any tax break at all.

There is an interesting and underappreciated feature to the exclusion from employer- sponsored insurance. If you look at where the money comes from before it enters the hands of a doctor, insurance company, or anyone else operating in the health sector, almost half of it comes from the government, and over a quarter of it from employers, which means that, generally, employers are the ones controlling and deciding how it is used. The consumer share is only about a quarter. It's really less than that because that 26% includes things like Medicare Part B premiums, where the consumer doesn't really have much of a choice about how to spend it. Also, with regard to the employee portion of their employer-sponsored insurance premiums, if they decided to buy health insurance elsewhere, they could end up spending twice as much for less coverage. So the exclusion actually gives employers control over a large chunk of what is actually the employee's compensation. This is compensation that the employee earned. The economists tell us that the employer wasn't providing health insurance, that they would have to return that money to the employee in the form of cash wages—(or other benefits, but likely cash wages) —and that's a substantial chunk of earnings that the employer gets to control. An average family policy offered through an employer costs $12,000.

On average, the "employer portion" of that is $9,000. That's a substantial chunk of money for most families that the employer gets to control, and it's why, if you look at only the private sector, employers control about half the money that goes into health care from private payers. There are two reasons for that. The first is the wedge that the exclusion drives between the worker and their earnings, and the second is the fact that, by drawing so many people out of the individual health insurance market, it makes a much thinner market and one much less attractive to people who might like to purchase insurance that meets their own needs.

We've talked a little bit about possible reforms. There's leveling up and leveling down. Limiting health-related tax breaks would be leveling down, and there are a couple of ways to do that. You can cap the exclusion so that if you have a family policy with premiums of $12,000, and we placed a cap of $6,000 on the exclusion, then the worker would have to pay taxes on the $6,000 that exceed that count.

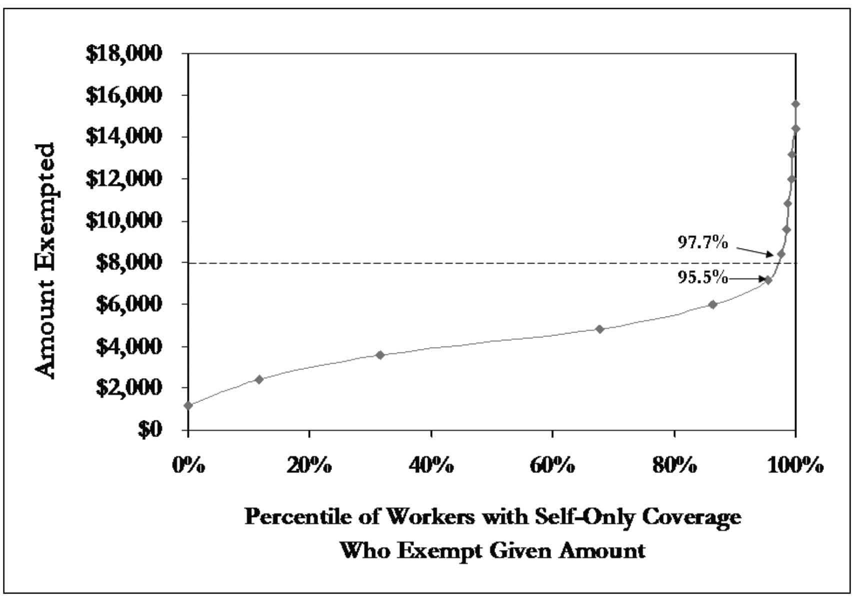
You could also eliminate the exclusion, which by itself would be a large tax increase because you'd be taxing a lot of previously untaxed activity. So usually when people talk about eliminating the exclusion they talk about reducing marginal tax rates so that it would be revenue-neutral.

You could also expand health-related tax breaks. Health savings accounts did that. Actually, they have done it twice so far: by creating a tax-free account, where you get a tax break for money the consumer controls; and again in 2006, they ramped up the amount of money that people can put into a health savings account tax-free. Both actions expanded health-related tax breaks.

Another option that has been proposed is full deductibility of all health spending. Now here I do mean deductibility. The authors of this idea propose allowing individuals to deduct from their income taxes every dollar that they spend on health insurance premiums and every dollar that they spend on medical care out-of-pocket. It also would expand the tax breaks as they exist.

Then there are some hybrid approaches that would broaden the existing tax breaks but still limit them in some way. Those include the tax credits Senator McCain has proposed and the standard health insurance deduction the President proposed.

I want to throw out on the table another option that actually builds on health savings accounts. This falls into the hybrid category. Like the others, it would expand the tax breaks, but cap them at the same time. So over time it would limit these tax breaks relative to what they are under the current law. I call this option large health savings accounts. Essentially, it would build on HSAs and replace the exclusion from employer-sponsored insurance. First, it would essentially triple the HSA contribution limits to $8,000 or $16,000 for individuals and families respectively (those aren't magic numbers; for political reasons they might need to be higher or lower). The second thing it would do to HSAs is remove the insurance requirement. Right now, you can only put money in an HSA tax free if you have a qualified high-deductible health insurance plan. There are reasons for removing the insurance requirement that I will get to. Lastly, it would allow people to purchase any kind of insurance from any source with their tax-free HSA funds. Currently, you can only use HSA funds for insurance premiums under limited circumstances.



Before I get to how they would do it, how would large health savings accounts work? For the most part, workers could take 100% of the money they currently exempt from income and payroll taxes as a tax-free large health savings account contribution. So the family that has a $12,000 policy through their employer could put the $9,000 their employer was paying into a large HSA, and put $3000 in, and the taxes would not go up. They could even put in $4,000. They could adjust that amount, as workers can with flexible spending accounts now. And, as I mentioned, they could purchase insurance from any source, or no insurance. There are people who would not be able to arrange this sort of payroll deduction, but there's a way to give those people an equivalent tax break. The President laid out that option solution when he proposed standard health insurance deduction.

So, if we made these three changes, how would it change the tax treatment of employer-sponsored health insurance? Well, you remember the three price distortions I mentioned earlier. It would completely eliminate the last two. It would retain a distortion between health and non-health uses of income, but because you would get that tax break for putting the money aside for your health care needs, it would eliminate the price distortions between third-party insurance and saving, and it would eliminate any distortion between consumers' decisions to purchase employer-sponsored insurance or insurance from another source. And like I said, it would broaden and cap the tax breaks. It would cap them because those contributions would effectively act as the cap on what is currently an unlimited exclusion. And if those were held constant in nominal terms or even in real terms, those contribution limits would reduce the tax break over time.

Large HSAs would allow people to purchase the mix of insurance and saving that is right for them. It would allow them to purchase the type of insurance that meets their needs, and allow them to choose between high-deductible health plans or health plans with lower deductibles, health maintenance organizations, preferred provider organizations, and fee-for-service or prepayment. And because when people are actually facing the cost of the premiums they are purchasing, they're more likely to reduce the amount of health insurance they purchase, that would reduce the consumption of the low-value care we mentioned before. It would also reduce labor market distortions, because there would be a level playing field for individual insurance. Fewer people would be stuck in jobs because their insurance would stay with them.

In terms of horizontal equity, it would eliminate the tax penalty currently imposed on people who don't get employer-sponsored insurance, so people would no longer be penalized based on the place of employment or the quantity of coverage they purchase and where they purchase it.

It is a little less clear what it would mean in terms of vertical equity, but I would argue that for those who are very concerned about vertical equity, large HSAs would be an improvement. First, they would cap the exclusion so that the wealthy would be less able to take advantage of these large tax breaks for health insurance and would extend a tax break for health care to low-income workers who currently get none.

Importantly, almost every proposal to reform the tax treatment of health insurance focuses on providing a different tax break for health *insurance*. What does that mean if you're uninsurable? If you're only providing a tax break for health insurance, people who are uninsurable don't get any tax break. This is true of the President's proposal for a standard health insurance deduction, proposals to cap the exclusion for employer-sponsored insurance, and McCain's tax credit proposal. One benefit of a large HSA approach is that it actually provides a tax break for health *savings*, rather than health *insurance*, so that the uninsurable get a tax break, the same break the insurable get.

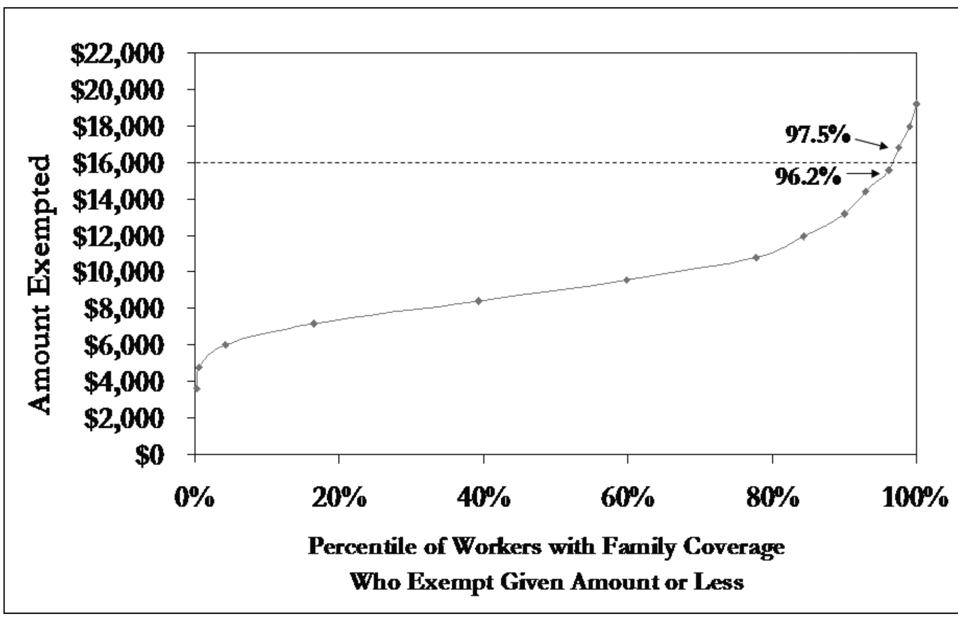
And I think it would be a more feasible way of capping the exclusion than most of the other proposal that we've seen. One of the recurrent obstacles to reforming the tax treatment of health insurance is that they pretty much all involve taxing previously untaxed economic activity. Large HSAs deal would let almost every worker get the same tax break they're currently getting. The first chart here (opposite) shows the cumulative distribution of how much of earnings workers exempted in the form of employer-sponsored health insurance in 2006. Now that line at $8,000 represents the proposed large HSA contribution limit. And you can see that about 97% of all workers exempted less than $8,000 in 2006 in the form of employer-sponsored insurance. What that means is that, by replacing the current exclusion with large health savings accounts, 97% of workers would see no tax increase. In fact, they may see a tax only if they wanted to purchase more generous insurance than they get right now. Only 3% of workers would see any possibility of a tax increase in the first year. Offsetting that potential tax increase, is the control they would get over the first $8,000 of their earnings, which is really a tax *cut*. The second chart (above) is the same, but for family coverage, and it shows basically the same thing, with a contribution limit of $16,000. Ninety-seven percent of workers who currently have employer-sponsored family coverage would see no increase in their taxes. And these are just another two ways of looking at that, showing the frequency distribution for those with self-only coverage and family coverage.

It can be likened to a tax cut, even for that 3%, because it gives them more control of the first $8,000 or $16,000 of their spending. And it would also make it easier to move toward a tax system that's completely neutral toward health expenditures.

There are some potential negatives, but those exist with all approaches for reforming the current approach to taxing health insurance. What is the effect on federal revenues? What are the effects on pooling? If you level the playing field between employer-sponsored insurance and individual market insurance, will that encourage low-risk people to leave the employer-sponsored pools, and therefore increase premiums for the high-cost people in those pools? Large HSAs create a potential problem with regard to free riding because people may decide they just don't want to purchase health insurance. Many of these potential negatives are smaller than they appear, and others can be mitigated by adjusting things like large HSA contribution limits. In my view, large HSAs would be less disruptive to people's health insurance, and would do more for the uninsurable than any other approach to reforming the tax treatment of health care. And I'll go ahead and stop there, and hopefully we can talk about those in the question-and-answer portion if there's any interest.

Thank you very much.

[Discussion and Q&A available in recording online.]



Civil Rights

Enhancing Disability Protection Without Abondoning Principle

*By Senator Orrin G. Hatch**

The ADA Restoration Act (ADARA)[1] states as its intention to reverse Supreme Court decisions that "narrowed the class of people who can invoke the protection from discrimination that the Americans with Disabilities Act of 1990 provides."[2] The bill's supporters claim that these decisions "ignored Congress's clear intent as to who should be protected,"[3] excluding "millions of people [Congress] intended to be protected under the ADA."[4] This article examines the ADA's basic principles, those Supreme Court decisions, and the ADARA's language and likely results. It concludes that, by abandoning the ADA's basic principles, the ADARA significantly expands, rather than restores, its intended coverage, and sets the ADA at odds with the rest of federal disability policy.

I. The Americans with Disabilities Act

Signed into law by President George H.W. Bush on July 26, 1990, the ADA has been called "the most extensive disability civil rights law ever enacted"[5] and "the most sweeping nondiscrimination legislation since the Civil Rights Act of 1964."[6] Senator Tom Harkin (D-IA), who introduced the bill,[7] called it the "emancipation proclamation" for disabled Americans, predicting that it will "change the way we live and the way we associate with one another in all aspects of our livelihood."[8] I was an ADA co-sponsor and described it as "a major, landmark piece of legislation"[9] that would "bring individuals with disabilities into the mainstream of the economic structure of this country."[10] To that end, the ADA prohibits discrimination against an individual on the basis of a present,[11] past,[12] or perceived[13] disability in employment,[14] state and local government services,[15] public accommodations,[16] and telecommunications.[17] In the employment context, it also requires provision of "reasonable accommodations to the known physical or mental limitations of an otherwise qualified individual with a disability who is an applicant or employee, unless… the accommodation would impose an undue hardship."[18]

A. Basic Principles

For more than three decades, Congress has built federal disability policy on basic principles such as individuality and functionality. The Rehabilitation Act of 1973, for example, states its purpose as authorizing programs to meet the needs of "handicapped individuals,"[19] and uses the word *individual* or its plural form 239 times. As one federal court put it, under the Rehabilitation Act, "[i]t is the impaired individual that must be examined, and not just the impairment in the abstract."[20] It defines the phrase *individual with a disability* in functional terms as "an individual who has a physical or mental impairment which substantially limits one or more… major life activities."[21] The Air Carrier Access Act of 1986[22] and the Fair Housing Act Amendments of 1988[23] define these terms by the same individual and functional principles.[24]

Against this backdrop, Congress implemented the principles of individuality and functionality in both the purpose and the provisions of the ADA. It pursues each of its four purposes on behalf of "individuals with disabilities."[25] The reports of the Senate Labor and Human Resources Committee[26] as well as the House Education and Labor Committee,[27] Judiciary Committee,[28] and Energy and Commerce Committee[29] similarly describe each of the ADA's purposes in terms of individuals with disabilities. The ADA's substantive provisions also implement the principle of individuality, defining the key term *disability* "with respect to an individual,"[30] and using the word *individual* or its plural form 297 times.

The ADA implements the principle of individuality in another important way. The House Education and Labor Committee report notes that disability discrimination "often results from false presumptions, generalizations"[31] and "stereotypical assumptions."[32] Generalizations and assumptions necessarily ignore individuals and their particular circumstances and abilities. In its ADA findings, Congress denounced restrictions and unequal treatment "resulting from stereotypic assumptions not truly indicative of the individual ability of such individuals to participate in, and contribute to, society."[33] As the Congressional Research Service explains, prohibiting discrimination based on perceived disability "has as its purpose the protection of individuals from stereotypical assumptions that do not reflect the individual's ability."[34]

The ADA also implements the related principle of functionality. The word dis*ability* itself reflects a focus on a person's ability and function.[35] Echoing previous disability statutes, the ADA defines a disability as "a physical or mental impairment that substantially limits one or more of the major life activities of [an] individual."[36] In other words, as each of the committee reports emphasizes,[37] while every disability is an impairment, only those impairments substantially limiting an individual's function are disabilities. Congress thus rejected a per se approach, based on assumptions and generalizations, that would automatically define any particular impairment or condition as a disability.

Instead, the definition of disability begins with the much broader category of impairments, defined in the committee reports[38] and subsequent regulations[39] to be "any… condition… affecting one or more… body systems" and "any mental or psychological disorder." Under this broad definition—unlimited by factors such as severity, symptoms, or duration—virtually every American is impaired. In the ADA, Congress emphasized that individuals with disabilities are a much smaller group,

** Senator Orrin G. Hatch represents the State of Utah and serves on the Health, Education, Labor, and Pensions Committee, which has jurisdiction over the Americans with Disabilities Act. He chaired the committee from 1981 to 1987 and was the committee's Ranking Member at the time of the ADA's passage in 1990.*

a "discrete and insular minority" of persons who have been "subjected to a history of purposeful unequal treatment."[40]

While the ADA, therefore, is similar to previous disability statutes, it is different than other civil rights statutes. The National Council on Disability (NCD) explains the difference this way: "Unlike prohibitions of discrimination according to race or gender, where one is automatically a member of a protected class by one's physical characteristics at birth, for one to be protected by the ADA one must qualify as a person with a disability."[41] The NCD lists this combination of individuality and functionality among the "key concepts in the ADA,[42] and these principles justify describing it as a "clear, balanced, and reasonable" approach to disability discrimination.[43] They keep the statute focused on the truly disabled, accommodate and balance various interests, and helped produce an overwhelming consensus supporting the ADA.

The basic principles of individuality and functionality mean that the "threshold issue in any ADA case is whether the individual alleging discrimination is an individual with a disability."[44] Because the statute itself does not define terms such as *impairment*, *substantially limits*, or *major life activity*, executive branch agencies and courts have developed criteria and standards. Disability advocates have highlighted several Supreme Court decisions which, they say, have narrowed the ADA's coverage and offer the ADARA as a legislative response.

B. Supreme Court Construction of the ADA

Sutton v. United Air Lines[45] was the first Supreme Court decision construing the ADA's employment provisions. United Airlines required uncorrected visual acuity and denied employment to twin sisters with severe myopia whose eyeglasses gave them 20/20 vision. Writing for a 7-2 majority, Justice Sandra Day O'Connor said that three ADA provisions support the conclusion that impairments should be examined in their mitigated state when determining the threshold issue of disability. First, the ADA requires that a person "be presently—not potentially or hypothetically—substantially limited in order to demonstrate a disability."[46] Second, "whether a person has a disability under the ADA is an individualized inquiry."[47] Third, considering all unmitigated impairments to be disabilities would expand the ADA's coverage far beyond its own estimate of 43 million disabled Americans.[48]

Murphy v. United Parcel Service[49] involved a plaintiff with hypertension. Federal regulations required that commercial vehicle drivers be medically certified to drive. In August 1994, Murphy was erroneously issued a medical certification when United Parcel Service (UPS) hired him as a mechanic, requiring that he drive commercial vehicles. When the error was discovered, UPS fired him "on the belief that his blood pressure exceeded the… requirements for drivers of commercial motor vehicles."[50] Murphy sued, claiming that he was fired because of either a present or a perceived disability. Writing for the same 7-2 majority, Justice O'Connor repeated *Sutton*'s holding that determining whether Murphy's impairment is a disability required "reference to the mitigating measures he employs."[51] The Court concluded that Murphy was not disabled because, as the appeals court found and Murphy did not challenge, "when medicated, [his] high blood pressure does not substantially limit him in any major life activity."[52] The Court also held that the major life activity of working involved "perform[ing] a class of jobs utilizing his skills,"[53] rather than the particular job he held.

Albertson's, Inc. v. Kirkingburg[54] involved a plaintiff with amblyopia, an uncorrectable vision impairment. Albertson's hired him as a truckdriver after a doctor "erroneously certified that he met the [Department of Transportation's] basic vision standard."[55] After his vision was later correctly assessed during a physical, Albertson's fired Kirkingburg despite his application for a waiver of the vision requirement, and refused to rehire him after he received that waiver. Writing for a unanimous Court, Justice David Souter explained that the ADA requires that an impairment must "substantially limit" or impose a "significant restriction" on a major life activity to constitute a disability.[56] In this case, the appeals court had weakened that standard by being "willing to settle for a mere difference" in a major life activity.[57] The Court also emphasized that determining disability requires consideration of "the individual's ability to compensate for the impairment"[58] and "the statutory obligation to determine the existence of disabilities on a case-by-case basis."[59]

Toyota v. Williams[60] involved a plaintiff with carpal tunnel syndrome caused by use of pneumatic tools on an engine fabrication assembly line. Williams requested an accommodation when a change in her duties led to nerve pain. Fired after missing a substantial amount of work, Williams sued. Her own request that she be allowed to perform her previous duties contradicted her claim that her condition substantially limited performing manual tasks, and the district court found that her impairment was not a disability. The appeals court reversed, holding that Williams' condition need only limit her from a class of manual activities "'affecting the ability to perform tasks at work.'"[61] As it had done in *Sutton*, the Court examined the ADA's text and considered important Congress's finding that "some 43,000,000 Americans have one or more physical or mental disabilities."[62] The Court said: "If Congress intended everyone with a physical impairment that precluded the performance of some isolated, unimportant, or particularly difficult manual tasks to qualify as disabled, the number of disabled Americans would surely have been much higher."[63]

II. The ADA Restoration Act

The ADARA states that these decisions were "contrary to explicit congressional intent expressed in the [ADA] committee reports."[64] Disability advocates similarly argue that these decisions "narrowed the definition of who qualifies as an 'individual with a disability'" under the ADA.[65] They offer the ADA Restoration Act, as its title suggests, as a restoration of the ADA's definition and coverage. It is appropriate, therefore, to examine the ADARA's language and likely impact to determine whether the ADARA is consistent with the ADA's basic principles, language, and coverage. This article will address three of the ways in which the ADARA, as introduced in July 2007, would change the original Act.

A. Mitigating Measures

Disability advocates focus their most consistent criticism on the Supreme Court's holding that mitigating measures must be considered in determining whether an impairment is a disability. When he introduced the ADARA, Senator Harkin said that this holding "ignored Congressional intent,"[66] creating "an absurd and unintended catch 22-type situation."[67] On the one hand, an unmitigated impairment might be a disability, but also render someone unqualified for a particular job. On the other hand, a mitigated impairment might make someone qualified for the job, but no longer disabled. The "Dear Colleague" letter Senator Harkin circulated with Senator Arlen Specter (R-PA), the legislation's principal co-sponsor, similarly criticized the Supreme Court for creating an "unintended catch-22," and said: "That is why we have introduced [the ADARA]."[68] True to its sponsors' word, the ADARA would prohibit consideration of mitigating measures that "the individual may or may not be using" in determining whether an individual has a disability.[69]

The mitigating measures issue has similarly been the primary target of articles, letters, press releases, congressional testimony, and other statements by scholars[70] and advocacy groups such as the American Association of People with Disabilities,[71] National Council on Independent Living,[72] American Association of Retired Persons,[73] American Civil Liberties Union,[74] Leadership Conference on Civil Rights,[75] United Spinal Association,[76] and American Diabetes Association.[77] This issue also dominated the hearings on the ADARA in the Senate Health, Education, Labor, and Pensions Committee (HELP)[78] and the House Judiciary Committee[79] and Education and Labor Committee.[80] News reports about the *Sutton* decision,[81] and even attorneys advising employers on ADA issues, have similarly highlighted mitigating measures as the primary issue.[82]

Properly understanding and addressing the mitigating measures issue requires two important distinctions. The first is that the ADA is silent on the mitigating measures issue. Reading the statute belies claims that "Congress explicitly stated that it did not intend mitigating measures to be considered in determining whether a person has a disability."[83] The ADA's so-called legislative history does contain statements, as in the Senate committee report, that "whether a person has a disability should be assessed without regard to the availability of mitigating measures."[84] But a statement in a committee report is not a statement by Congress. As the Seventh Circuit recently put it, "Congress did not enact its members' ***beliefs***; it enacted a text."[85] Neither did it enact its committees' legislative reports. While courts can use legislative history material to clarify ambiguous language that Congress did enact into law, that material cannot be used as a proxy for statutory language that Congress never enacted at all. The Supreme Court has repeatedly held that "the authoritative statement is the statutory text, not the legislative history or any other extrinsic material. Extrinsic materials have a role in statutory interpretation only to the extent they shed a reliable light on the enacting Legislature's understanding of otherwise ambiguous terms."[86] The ADA contains no terms whatsoever, ambiguous or otherwise, regarding mitigating measures.

In *Sutton*, the Supreme Court had to "say what the law is"[87] regarding mitigating measures by construing a statute that says nothing about the issue. The Court based its conclusion on what Congress did say in other provisions of the ADA. As such, since its conclusion regarding mitigating is consistent with those provisions, the Court's conclusion in *Sutton* was a reasonable construction of the statute. Criticizing the Supreme Court for refusing to go beyond construing the existing ADA to in effect create a statutory provision Congress did not enact is misplaced. Prohibiting consideration of mitigating measures, then, does not restore the ADA on this issue, because the ADA does not address it.

The second important distinction is between the definition of disability and criteria, such as mitigating measures, for applying that definition. The Supreme Court clearly made this distinction in *Sutton*, citing the ADA's definition of disability[88] and the EEOC regulations defining its components,[89] but identifying the issue in the case as "whether disability is to be determined with or without reference to corrective measures."[90] Mitigating measures are a criterion for determining whether an impairment is a disability, but does not implicate the definition of disability.

Disability advocates fail to make this distinction, mistakenly asserting that the Court changed the definition itself. As the American Diabetes Association put it, for example, "[a]t the heart of the problem lies the definition of disability."[91] Others have gone even further, claiming that these Supreme Court decisions have excluded altogether certain impairments or conditions from ADA coverage. The Consortium for Citizens with Disabilities, for example, asserts that courts have "dramatically changed the meaning of 'disability' under the ADA" in a way that "exclud[es] individuals with serious health conditions like epilepsy, diabetes, cancer, HIV, muscular dystrophy, mental health conditions, and multiple sclerosis."[92] At the Senate HELP Committee hearing on the ADARA, this group distributed a list of 14 impairments under the heading "People NOT Covered Under the ADA." At that same hearing, advocates displayed a chart asserting that each of these impairments is "not a disability under the ADA today." The ACLU also claims that people with eight listed "and many other impairments" do not qualify for ADA protection.[93]

These categorical claims are incorrect. No court has ever held that these or any other impairments are either always or never disabilities.[94] No impairment has ever been excluded from coverage under the ADA. To the contrary, Congress has consistently rejected such per se generalizations in federal disability statutes for more than three decades. Under the Rehabilitation Act, under the ADA before *Sutton*, and under the ADA today, courts decide whether individuals with these impairments are disabled based on the impact, not the identity, of their impairment. The definition of disability has not changed.

If the Supreme Court in *Sutton* changed the definition of disability in a way that simply excludes certain impairments from ADA coverage, there should be a noticeable shift since that decision in the outcomes of discrimination cases involving those impairments. The Equal Employment Opportunity Commission, which administers the ADA's Title I employment provisions, compiles such statistics, reporting the percentage of disability discrimination claims that the agency concludes have merit.[95] In the aggregate, the proportion of disability claims with such reasonable cause outcomes rose by 30% from the

1997-'99 fiscal years before *Sutton* to the 2000-'07 fiscal years since that decision.[96]

The EEOC also provides claim resolution figures for cases involving individual impairments, including the specific conditions that disability advocates claim are now categorically excluded from ADA coverage. The EEOC could not conclude that a discrimination claim has merit under the ADA if the impairment underlying that claim is not covered by the ADA. Merit factor resolutions of cases involving some impairments did decline from the 3 years before *Sutton* to the years since that decision: asthma cases by 8%, epilepsy cases by 7%, and hearing impairment cases by 2%.[97] But merit factor resolutions in cases involving other impairments rose by much larger margins: diabetes cases by 12%, bipolar disorder cases by 13%, cancer cases by 14%, multiple sclerosis cases by more than 27%, and mental retardation cases by more than 59%.[98] Each of these is an impairment that disability advocates claim is now entirely excluded from ADA coverage, and each is obviously still covered by that statute.

Because the Supreme Court did not change the definition of disability, these and other impairments will continue to constitute disabilities if they substantially limit an individual's major life activities. As a result, if Congress chose (which it certainly may) to prohibit consideration of mitigating measures as a criterion for determining whether an impairment is a disability, it would be changing, rather than restoring, the ADA.

B. Definition of Disability

The discussion above outlined that the mitigating measures issue is the primary concern raised by disability advocates, that Congress did not address that issue in the ADA, and that it is distinct from the definition of disability. The ADARA would nonetheless go beyond prohibiting consideration of mitigating measures and change the ADA's definition of disability itself. Under the ADARA, a disability would be simply "a physical or mental impairment"[99] unlimited by factors such as duration, severity, or limitation on an individual's function, and without regard to whether "any manifestation of the impairment is episodic"[100] or the impairment is "in remission or latent."[101]

By retaining and codifying the extremely broad definition of impairments currently found in the ADA's legislative history and implementing regulations, the ADARA would thus define every condition that affects the body or mind as a disability. This would change not only the definition of disability but the very concept of disability on which that definition is based, and for the first time place the ADA at odds with other federal disability statutes. And it renders inexplicable the ADARA's provision prohibiting consideration of mitigating measures. There would be no need to prohibit consideration of mitigating measures if all impairments, including unmitigated ones, are automatically disabilities.

In addition to its language, the ADARA's likely results show that it would expand, rather than restore, the ADA's coverage and impact. Congress stated in the ADA that "some 43,000,000 Americans"[102] have disabilities, or approximately 17% of the population at the time it was enacted.[103] Estimating the coverage of much broader definitions need not proceed arbitrarily. In a significant 1986 report, available to Congress when it passed the ADA, the NCD described a "health conditions approach" to defining disability which would include "all conditions… which impair the health or interfere with the normal functional abilities of an individual."[104] This tacitly functional definition is broader than the ADA in two respects. The "normal functional abilities" category is broader than the ADA's "major life activities." And the "interfere with" degree of functional impact is broader than the ADA's "substantially limit." The NCD estimated that this broad health conditions approach would cover more than 160 million Americans,[105] or approximately two-thirds of the American population in 1986.[106]

The California Fair Employment and Housing Act uses a definition of disability that is even broader than the NCD's health conditions approach. It defines a disability as a mental[107] or physical[108] condition that merely "limits a major life activity." By diluting an impairment's degree of functional impact even further, this definition appears to be minimally functional in form, but is virtually per se in substance. If applied nationally, this virtual per se definition would classify as disabled at least as many, and probably more, Americans as the NCD health conditions approach. As a result, a definition of disability which requires only a demonstrable impact or limitation on an individual's function, as opposed to the ADA's substantial limitation, would cover more than two-thirds of the population.

The ADARA is at the far end of this spectrum, deleting entirely all limiting factors and using an explicitly per se definition of disability.[109] Any impairment, no matter what its duration, intensity, functional impact, or symptoms would be a disability under the ADARA. As a result, the percentage of Americans covered by the ADARA would be substantially greater than under the NCD health conditions approach, perhaps 90% or higher, which today would constitute at least 275 million Americans. Significantly, when Senator Harkin introduced the ADARA, he stated that today there are "50 million Americans with disabilities,"[110] or approximately 16.5% of the current population.[111] While the disabled population has thus remained stable since the ADA's passage, the ADARA's per se definition would likely cover more than six times as many people as Congress intended the ADA to cover. Changing the definition of disability is not only unnecessary to address the mitigating measures issue, but changing it from a functional to a per se definition fundamentally changes the ADA to cover far more Americans than Congress ever contemplated, let alone intended.

C. Rules of Construction

Various statutes, including the ADA, contain rules of construction intended to guide courts in construing and applying statutory provisions. The Supreme Court has said that "in interpreting a statute a court should always turn to one, cardinal rule before all others. We have stated time and again that courts must presume that a legislature says in a statute what it means and means in a statute what it says there."[112] In other words, legislating, which the Constitution assigns to Congress alone,[113] involves determining what statutes mean as well as what they say. To that end, the ADA's rules of construction focus on particular words or phrases, clarifying their definition and guiding their application in discrete situations covered by individual statutory provisions.[114]

The ADARA contains so-called rules of construction of an entirely different kind which the Congressional Research Service has been unable to find in any other federal statute. Rather than the internally-focused process of explaining the meaning and application of the language Congress enacted, the ADARA's externally-focused rules of construction would invite, perhaps even mandate, the judicial and executive branches to continue changing the ADA's meaning and application. The ADARA, for example, would require that the ADA's provisions "be broadly construed"[115] and that deference be given to regulations and guidelines issued by executive branch agencies, "including provisions implementing and interpreting the definition of disability."[116]

In other words, the ADARA would treat Congress as beginning, but not ending, the legislative process. Ironically, disability advocates say that the ADARA is necessary because the courts changed the meaning of the ADA's provisions from what Congress intended. The ADARA's rules of construction, however, would essentially require exactly what its backers condemn. These provisions are not properly called rules of construction at all. Statutory construction involves determining what the legislature meant by what the legislature said. Changing the meaning of a statute's words changes that statute as surely as changing the words themselves. Because neither the judicial nor the executive branch has such power, Congress in the ADARA would be abdicating its constitutional authority by passing its lawmaking baton to the other branches.

III. Enhancing Disability Protection Without Abandoning Principle

The ADARA's language and likely results demonstrate that it would significantly expand, rather than restore, the ADA's coverage and impact. A per se definition of disability—whether explicit, virtual, or incremental—would abandon the principles of individuality and functionality that today form the basis of federal disability policy, and go far beyond addressing the mitigating measures issue that disability advocates have identified. And the ADARA's rules of construction would invite the judicial and executive branches to change, rather than interpret and apply, the statute.

Congress can, however, enhance the ADA's protections without abandoning its basic principles by addressing the mitigating measures issue without changing the definition of disability. This balanced approach both requires something from and provides something for the disability and business sides of the equation. On the disability side, this approach requires remaining focused on the truly disabled, but provides a more generous application of the criteria for doing so, thereby minimizing the "catch-22" situations brought about by the consideration of mitigating measures. On the business side, this approach requires a broader view of the impairments that may qualify as disabilities but provides the same underlying definition of disability. Defining disabilities as a subset of impairments keeps the ADA focused on the truly disabled and keeps its directives toward business reasonable. It was this balanced approach, incorporating various interests, that led to the consensus behind the ADA in 1990 and it can provide the basis for consensus in enhancing the ADA's protection today.

In September 1991, *Time* magazine said that I was "the key Republican in the deal" that, one year earlier, had produced the ADA.[117] I believe the ADA remains a model of bipartisan legislative compromise that continues to help millions of Americans. The final product was more balanced, focused, and consistent with other disability statutes than the original. It accommodated the interests of business while promoting the interests of the disabled. Congress can take the same approach, with the same result, with legislation that enhances the ADA's protections without abandoning its principles.

Endnotes

1 Senator Tom Harkin (D-IA) introduced S.1881 and Rep. Steny Hoyer (D-MD) introduced H.R.3195 on July 26, 2007, the 17th anniversary of the date that President George H.W. Bush signed the ADA into law.

2 S.1881, §2(b)(2).

3 Cong. Rec., July 26, 2007, at S10151 (statement of Sen. Harkin).

4 *Id.*

5 Charles B. Craver, *The Judicial Disabling of the Employment Discrimination Provisions of the Americans with Disabilities Act*, 18 Lab. Law. 417,417 (2003).

6 Nancy Lee Jones, *The Americans with Disabilities Act (ADA): The Definition of Disability*, CRS Report for Congress (Apr. 23, 2008), at 1.

7 Introduced on May 9, 1989, S.933 would eventually attract 63 cosponsors and pass the Senate by a vote of 76-8 on Sept. 7, 1989. After working out differences with the House version, the Senate passed the ADA conference report by a vote of 91-6 on July 13, 1990.

8 Cong. Rec., July 13, 1990, at S9529 (statement of Sen. Harkin).

9 Cong. Rec., Sept. 7, 1989, at 19804 (statement of Sen. Hatch).

10 Cong. Rec., July 11, 1990, at 17365 (statement of Sen. Hatch).

11 42 U.S.C. §12102(2)(A) ("a physical or mental impairment that substantially limits one or more… major life activities").

12 *Id.* §12102(2)(B) ("a record of such an impairment").

13 *Id.* §12102(2)(C) ("being regarded as having such an impairment").

14 Title I.

15 Title II.

16 Title III.

17 Title IV.

18 42 U.S.C. §12112(b)(5)(A).

19 Rehabilitation Act of 1973, Public Law 93-112, Section 2(1). The ADA uses "individuals with disabilities" rather than "handicapped individuals" because it is more "in line with the sensibilities of most Americans with disabilities." The Americans with Disabilities Act of 1989, Senate Report 101-116, Committee on Labor and Human Resources (1989), at 21 (hereinafter S.Rep.101-116). Nonetheless, "[n]o change in definition or substance is intended to be attributed to this change in phraseology." *Id. See also* Americans with Disabilities Act of 1990, House Report 101-485, Part 2, Committee on Education and Labor (1990), at 55 (hereinafter H.Rep.101-485,Pt.2) (Congress intended that the ADA's definition be "comparable to the definition used in… the Rehabilitation Act of 1973."). *See also* Chai Feldblum, *The Americans with Disabilities Act Definition of Disability*, 7 Lab. Law. 11, 11 (1991) (the ADA's definition of "person with a disability" is "the same definition which Congress adopted to define a 'person with a handicap' for purposes of Title V of the Rehabilitation Act of 1973.").

20 E.E. Black, Ltd. V. Marshall, 497 F. Supp. 1088, 1099 (D. Hawaii 1980).

21 29 U.S.C. §705(20)(B)(i).

22 49 U.S.C. §41705(a)(1).

23 42 U.S.C. §3602(h)(1). This was "the first time that 'disability' was added to the list of traditionally protected groups in a general federal anti-discrimination statute." Arlene Mayerson, *The Americans with Disabilities Act—An Historic Overview*, 7 Lab. Law. 1, 3 (1991).

24 The Individuals with Disabilities Education Act of 1975 implements the same principles in the specialized context of education. In addition to its title, the statute's first congressional finding describes the "national policy of ensuring equality of opportunity, full participation, independent living, and economic self-sufficiency for *individuals with disabilities*." 20 U.S.C. §1400(c)(1) (emphasis added). Like the Rehabilitation Act before it, this statute requires that an individual not only have an impairment but also, "by reason thereof, needs special education and related services." 20 U.S.C. §1401(3)(A)(ii).

25 42 U.S.C. §12101(b).

26 S.Rep. 101-116, at 2 (emphasis added).

27 *See also* H.Rep. 101-485, Part 2, at 22-23.

28 Americans with Disabilities Act of 1990, House Report 101-485, Part 3, Committee on Judiciary (1990), at 23 (hereinafter H.Rep.101-485, Pt.3);

29 Americans with Disabilities Act of 1990, House Report 101-485, Part 4, Committee on Energy and Commerce (1990), at 23 (hereinafter H.Rep.101-485, Pt.4).

30 42 U.S.C. §12102(2)(A). *See also* 29 C.F.R. §1630.2(g)(1).

31 H.Rep.101-485, Pt.2, at 30.

32 *Id*. at 40.

33 42 U.S.C. §12101(a)(7).

34 Nancy Lee Jones, *supra* note 6, at 10.

35 One popular online dictionary defines the term *disability* as "lack of adequate… physical or mental ability" or "a physical or mental impairment that interferes with or prevents normal achievement." http://dictionary.reference.com/browse/disability.

36 42 U.S.C. §12102(2)(A).

37 S. Rep.101-116, at 22; H.Rep.101-485, Pt.2, at 52; H.Rep.101-485, Pt.3, at 28.

38 S.Rep.101-116, at 22; H.Rep.101-485, Pt.2, at 51.

39 29 C.F.R. §1630.2(h)(1).

40 42 U.S.C. §12101(a)(7).

41 National Council on Disability, Equality of Opportunity: The Making of the Americans with Disabilities Act 223 (1997).

42 *Id.*

43 S.Rep.101-116, at 20.

44 Nancy Lee Jones, *supra* note 6, at 2.

45 527 U.S. 471 (1999).

46 *Id*. at 482.

47 *Id*. at 483.

48 *Sutton*, 527 U.S. at 485.

49 527 U.S. 516 (1999).

50 *Id*. at 520.

51 *Id.* at 521.

52 *Id.*

53 *Id*. at 524. Murphy "has put forward to evidence that he is regarded as unable to perform any mechanic job that does ot call for driving a commercial motor vehicle and thus does not require DOT certification." *Id*. Murphy, in fact, "secured another job as a mechanic shortly after leaving UPS." *Id*.

54 527 U.S. 555 (1999).

55 *Id*.

56 *See* 29 C.F.R. §1630.2(j)(ii).

57 *Albertson's*, 527 U.S. at 565.

58 *Id*.

59 *Id*. at 566.

60 534 U.S. 184 (2002).

61 *Id*. at 192 (quoting Williams v. Toyota Motor Manufacturing, 224 F.3d 840,843 (6th Cir. 2000)).

62 42 U.S.C. §12101(a)(1).

63 *Toyota*, 534 U.S. at 197.

64 S.1881, §2(a)(7).

65 Cong. Rec., July 26, 2007, at S10151 (statement of Sen. Harkin).

66 *Id.*

67 Cong. Rec., July 30, 2007, at S10211 (statement of Sen. Harkin).

68 Letter of November 13, 2007. Senator Harkin has repeatedly criticized the "absurd and unintended Catch-22" which he says the Supreme Court's decisions created. *See, e.g.,* Cong. Rec., July 26, 2007, at S10151; Cong. Rec., July 30, 2007, at S10211. This message has been echoed by commentators and reporters. *See, e.g.,* Jim Radermacher, *Congress Must Restore Rights of People with Disabilities*, Chi. Sun Times, Feb. 17, 2008, at A28 ("These decisions have created an absurd and unintended Catch-22"); Lynne Landsberg, *A Helping Hand*, Baltimore Sun, Apr. 2, 2008, at 15A ("Our courts have created an inescapable Catch-22"); Andrew J. Imparato, *Supreme Court Deals a Blow to Disabled Americans*, Milwaukee J. Sentinel, July 11, 1999, at 3 ("This puts people in a Catch-22 situation that Congress never intended").

69 S.1881, §4(2)(B)(II).

70 *See, e.g.,* Bonnie Poitras Tucker, *The Supreme Court's Definition of Disability Under the ADA: A Return to the Dark Ages*, 52 Alabama L. Rev. 321,342 (2000); Sharona Hoffman, *Corrective Justice and Title I of the ADA*, 52 Am. U.L. Rev. 1213,1232 (2003); Marta Russell, *Backlash, the Political Economy, and Structural Exclusion*, 21 Berkeley J. Emp. & Lab. Law 335,354 (2000); Julie McDonnell, *Sutton v. United Airlines: Unfairly Narrowing the Scope of the Americans with Disabilities Act*, 39 Brandeis L.J. 471,483 (2000); Molly Joyce, *Note: Has the Americans with Disabilities Act Fallen on Deaf Ears? A Post-Sutton Analysis of Mitigating Measures in the Seventh Circuit*, 77 Chi.-Kent L. Rev. 1389,1403 n.116 (2002); Elizabeth Theran, *'Free to Be Arbitrary and…Capricious': Weight-Based Discrimination and the Logic of American Antidiscrimination Law*, 11 Cornell J. Law & Pub. Pol'y 113,185 n.346 (2003); Deirdre M. Smith, *The Paradox of Personality: Mental illness, Employment Discrimination, and the Americans with Disabilities Act*, 17 Geo. Mason U. Civ. Rts. L.J. 79,121 n.233 (2006).

71 Cheryl Sensenbrenner, who chairs the board of the American Association of People with Disabilities, testified to the House Judiciary Subcommittee on the Constitution that decisions such as *Sutton* are "at complete odds with clear Congressional intent" on issues such as mitigating measures and create a "'double whammy' for people with disabilities." http://judiciary.house.gov/Hearings.aspx?ID=182.

72 Letter to Rep. Jarrold Nadler, Oct. 4, 2007 (the *Sutton* decision on mitigating measures "creates a Catch-22 in which employees can be discriminated against on the basis of their disability but unable to enforce their rights because they cannot meet the high threshold the courts have set to prove they are disabled.").

73 Letter to Rep. Jarrold Nadler, Oct. 11, 2007 ("Recent interpretations of the ADA have diminished its protections for those who are deemed to have 'corrected' their disability through medication or other assistive devices.").

74 Letter to Senators Edward Kennedy & Michael Enzi ("The court decisions… have created an unintended Catch-22 where individuals taking medication or using other mitigation measures to manage their condition may no longer qualify as 'disabled' under the ADA."). The ACLU's legislative counsel has said that the courts "have created a Catch-22 wherein individuals taking medication or using other measures to manage their disability may no longer qualify as 'disabled' under the ADA and are denied protection

from employment discrimination." Press Release, *ACLU Commends Senate Committee for Support of ADA Restoration Act*, November 15, 2007, *available at* http://www.aclu.org/disability/32795prs20071115.html.

75 Letter to Senators Edward Kennedy & Michael Enzi, Nov. 14, 2007 (Congress never intended to exclude individuals because the effects of their conditions "could be mitigated through treatment").

76 Statement, ADA Restoration Act (undated), *available at* http://www.unitedspinal.org/advocacy/employment-discrimination/ (courts "have created an absurd 'Catch-22'whereby a person is 'too disabled' to do the job—but not 'disabled enough' to be protected by the ADA.").

77 Statement of Oct. 4, 2007 (the ADARA "prevents the courts from disqualifying an individual from [ADA] protection because of the effects of 'mitigating measures'").

78 Witness statements for the hearing held on November 15, 2007, can be accessed at http://help.senate.gov/Hearings/2007_11_15_b/2007_11_15_b.html.

79 Witness statements for the hearing held on October 4, 2007, can be accessed at http://judiciary.house.gov/Hearings.aspx?ID=182.

80 Witness statements for the hearing held on January 29, 2008, can be accessed at http://edlabor.house.gov/hearings/fc-2008-01-29.shtml.

81 *See, e.g.,* Joan Biskupic, *Five Cases at Supreme Court Could Affect Disabilities Law*, Wash. Post, Feb. 21, 1999, at A3 ("To the sisters and others with serious handicaps that can be improved by taking medicine or using some device, it is a Catch-22"); Michael Remez, *Supreme Court Limits Conditions Covered by ADA*, Hart. Courant, June 23, 1999, at A1 ("To disability rights advocates, the court's logic amounts to a Catch-22").

82 *See, e.g.,* Lawrence Postol, *To Be, or Not to Be: The ADA Catch-22*, *available at* http://www.dcbar.org/for_lawyers/resources/publications/washington_lawyer/july_august_2000/ada.cfm ("the case law has evolved such that there is a Catch-22 to the ADA. To qualify for ADA protection, a worker needs a condition that is severe—significantly limiting a major life activity, but not so severe that the worker cannot perform the essential functions of his or her job with reasonable accommodations.").

83 Statement by Paralyzed Veterans of America before the House Committee on the Judiciary, Subcommittee on Constitution, Civil Rights and Civil Liberties, October 4, 2007.

84 S.Rep. 101-116, at 23. *See also* H.Rep. 101-485, pt.3, at 28 ("The impairment should be assessed without considering whether mitigating measures… would result in a less-than-substantial limitation.").

85 Jones v. Harris Assoc., 2008 U.S. App. LEXIS 10804 (May 19, 2008) (emphasis in original).

86 Exxon Mobil Corp. v. Allapattah Servs., 545 U.S. 546, 568 (2005). *See also* BedRoc Ltd., LLC v. U.S., 541 U.S. 176, 187 n.8 (2004) (referring to the Court's "longstanding precedents that permit resort to legislative history only when necessary to interpret ambiguous statutory text").

87 Marbury v. Madison, 5 U.S. 137,177 (1803).

88 527 U.S. at 478.

89 *Id*. at 479-80.

90 *Id*. at 481.

91 Statement of the American Diabetes Association, *supra* note 77. *See also* Letter from National Health Council to Reps. Steny Hoyer and F. James Sensenbrenner, Oct. 9, 2007 ("the courts have narrowed the definition of disability"); Statement of Epilepsy Foundation, Oct. 4, 2007 ("the courts have narrowed the definition of disability"); Letter from Leadership Conference on Civil Rights, *supra* note 75 ("court decisions narrowing the definition of 'disability'"); Statement of National Disability Rights Network, Nov. 15, 2007 ("the courts have narrowed the definition of disability").

92 Consortium for Citizens with Disabilities, *Statement in Support of the Americans with Disabilities Act Restoration Act of 2007* (undated).

93 Letter, *supra* note 74.

94 Following the hearing on the ADARA before the Senate Health, Education, Labor, and Pensions Committee on November 15, 2007, I asked each of the witnesses, which included several disability advocates, several questions in writing, including the following: "Has any court, state or federal, ever ruled that, under the Americans with Disabilities Act, any particular impairment is, by definition, a disability without regard to its effect on a major life activity? If so, please provide citations to such rulings." By the date this article was submitted to *Engage*, I had received no response from any of the witnesses.

95 These are cases in which the EEOC determines "that a charge had merit in cases where it could not be successfully conciliated." Sharona Hoffman, *supra* note 70, at 1248.

96 http://www.eeoc.gov/stats/ada-charges.html.

97 http://www.eeoc.gov/stats/ada-merit.html.

98 *Id*.

99 S.1881, §4(2)(A)(i).

100 *Id*. at §4(2)(B)(III).

101 *Id*. at §4(2)(B)(IV).

102 42 U.S.C. §12101(a)(1).

103 The Census Bureau estimates the population of the United States on July 1, 1990, at 249.5 million. http://www.census.gov/popest/archives/1990s/popclockest.txt.

104 http://www.ncd.gov/newsroom/publications/1986/toward.htm.

105 *Id*.

106 *Id*. The Census Bureau estimates that the population of the United States on July 1, 1986, at 240.1 million. http://www.census.gov/popest/archives/1990s/popclockest.txt.

107 California Government Code §12926(i).

108 *Id*. at §12926(k).

109 An incremental per se approach would begin by defining specific impairments, rather than all impairment, as disabilities. Like the ADARA, this approach would be at odds with the ADA's principles of individuality and functionality, as well as with other federal disability statutes. Treating citizens with particular impairments as members of interest groups has the additional political liability of inviting those whose impairments are not on the list to continue lobbying Congress for inclusion.

110 Cong. Rec., July 26, 2007, at S10151. *See also* Susan Gluck Mezey, *Judicial Interpretation of Title III of the Americans with Disabilities Act*, 15 J. Disability Pol'y Stud. 147,147 (2004) ("By 2000, the number of people with disabilities in the United States had risen to almost 50 million").

111 The Census Bureau estimates the population of the United States on November 13, 2007, at 303.3 million. http://www.census.gov/population/www/popclockus.html.

112 Conn. Nat'l. Bank v. Germain, 503 U.S. 249, 253-54 (1992).

113 U.S. Const., art. I, §1 ("All legislative powers herein granted shall be vested in a Congress of the United States").

114 *See, e.g.,* 42 U.S.C. §12112(b); *id*. at §12114(b); *id*. at §12143(f); *id*. at §12182(b); *id*. at §12188(b); *id*. at §12201; *id*. at §12210(b).

115 *Id.* at S.1881, §8(e).

116 *Id.* at §8(g).

117 Laurence I. Barrett, *Presidential Candidates "Always Attack, Never Defend*," Time, Sept. 23, 1991.

Super-Tuesday for Equal Rights

*By Ward Connerly & Jennifer Gratz**

The Court expects that 25 years from now, the use of racial preferences will no longer be necessary to further the interest approved today.

Justice Sandra Day O'Connor
Grutter opinion (June 23, 2003)

When the *Gratz* and *Grutter* opinions were released in 2003, many believed that eliminating race preferences was a lost cause, at least for a while. But the organizations of which we are a part—the American Civil Rights Institute (ACRI) and the American Civil Rights Coalition (ACRC)—set out to abbreviate Justice O'Connor's predicted twenty-five years.

Just days after the *Gratz* and *Grutter* decisions, we announced our intentions to assist the people of Michigan place a civil rights initiative on the 2004 ballot. The language of the California Civil Rights Initiative (Proposition 209) was used as a model; the operative clause of the initiatives reads:

> The state shall not discriminate against, or grant preferential treatment to, any individual or group on the basis of race, sex, color, ethnicity, or national origin in the operation of public employment, public education, or public contracting.

We knew this would be no small task. Michigan, after all, was considered "ground zero" for the issue of race-based preferences, and because of the lawsuits against the University of Michigan, the opposition was already organized and determined to stop the people from voting. The American Civil Liberties Union (ACLU), the National Association for the Advancement of Colored People (NAACP), a group named By Any Means Necessary (BAMN), and many others joined together and filed numerous lawsuits to stop the initiative from gathering signatures and gaining support. After the Michigan Civil Rights Initiative failed to make the 2004 ballot opponents claimed victory. But, quietly, our small cadre of supporters redirected our efforts to the 2006 ballot.

The process for qualifying an initiative for the ballot sounds simple: gather enough signatures in the prescribed amount of time and your issue gains access to the ballot. However, nothing is simple when it comes to race. The Michigan effort turned in the most signatures in the state's history—and even that margin was not comfortable enough. Opponents challenged the signatures in every court fathomable—from state court to federal court to the court of public opinion. Political posturing aside, the courts ultimately ruled that the people had the right to vote. And, on November 7, 2006, the people of Michigan reignited the national movement to end race based preferences.

The Michigan victory was the first step to curbing Justice O'Connor's twenty-five year sentence of race preferences. Immediately after, when asked her thoughts about the Michigan Civil Rights Initiative, Justice O'Connor simply stated, " It is entirely within the right and privilege of voters..." And so, it is only fitting that more states have the opportunity to exercise their right to prohibit the use of racial preferences.

When we launched "Super-Tuesday for Equal Rights," we knew that our opponents would consider this their last stand and would pull out all the stops. For example:

• In past efforts, and in the current battle-ground states, supporters of race preferences tried to blur the line between affirmative action programs that grant preferential treatment based on race and general affirmative action programs with misleading rhetoric. Only programs—whether named "affirmative action" or not—which are outright discriminatory or grant preferential treatment are eliminated by a civil rights initiative.

• Opponents to color-blind government often fail to recognize that "affirmative action" and "race preferences" are not necessarily one and the same. In fact, in 1961, when President John F. Kennedy first introduced the nation to the concept of "affirmative action," he did not advocate preferential treatment. Executive Order 10925 stated that "affirmative action" must be taken to "ensure that applicants are employed, and that employees are treated during employment, without regard to their race, creed, color, or national origin." Kennedy's message was simple: treat all people equally, *without regard* to race. The types of programs Kennedy's version of "affirmative action" allowed for are preference-free. For instance, one can believe in the need to help those who are socio-economically disadvantaged, one form of affirmative action, without advocating or using preferences based on race.

• Supporters of race preferences confuse "equal opportunity" with equal outcome, and use rhetoric that confuses the public as well. Affirmative action programs intended to guarantee equal opportunity can (and should) exist without preferential treatment. Americans tend to believe that anyone, regardless of race or sex, should be allowed to compete, and once that occurs, that the government should get out of the way and let the chips fall where they may.

• Supporters of preferences say that they, like everyone, are against quotas, but that they believe in and celebrate "diversity." However, the term diversity is nothing more than a code-word for quota in most cases under our current preferential regime.

• Supporters of preferential treatment claim that somehow breast cancer screening centers, neo-natal programs, or domestic violence shelters also qualify as preferential treatment. This tactic preys on voter fear, on the hope that the public will believe that if preference programs are eliminated health programs which are

....

** Ward Connerly is President of the American Civil Rights Institute, a former member of the Board of Regents of the University of California, and a 2005 recipient of the Bradley Prize for his defense of the American ideals of freedom, liberty and equality. Jennifer Gratz is Director of Research at the American Civil Rights Institute, was the lead plaintiff in the landmark US Supreme Court case, Gratz v. Bollinger, and the 2007 recipient of the American Conservative Union's Reagan Award for her role as Executive Director of the Michigan Civil Rights Initiative.*

often specific to one gender would be in jeopardy. In fact, none of these health-based programs have been eliminated in any of the states where a civil rights initiative has already passed and been in practice—in some cases, for a decade or longer.

• Opponents allege that it is fraudulent to say that ending race preferences is consistent with the 1964 Civil Rights Act, and that ending race preferences does not spell Doomsday for *all* affirmative action programs. The media seems to accept these allegations—strengthening the depth of their message. Of course, the very essence of civil rights is to treat everyone equally. No amount of mental gymnastics can twist logic enough to convince rational thinkers that when one is preferred based on race another is not discriminated against by the same criterion.

Nonetheless, we did not expect civility and the law to be cast aside in this debate. In Missouri, for example, activist politicians changed the language of the initiative, leading to a lengthy court battle. The Missouri constitution allows petition sponsors up to eighteen months to collect signatures from voters, but because of the legal challenge the Missouri Civil Rights Initiative had less than four months to circulate its petition.

In the four months that the Missouri Civil Right Initiative had to collect signatures, petition circulators were arrested for petitioning on public property, such as in front of public libraries, on public sidewalks, and just outside offices of the Department of Motor Vehicles. In addition, the unions, the Association of Community Organizations for Reform Now (ACORN), the Coalition to Defend Affirmative Action, Integration & Immigrant Rights and Fight for Equality By Any Means Necessary (BAMN), and the Soros-funded Ballot Initiative Strategy Center joined together to organize a "blocking" campaign. Groups that typically consisted of five to ten people would surround a petition circulator and intimidate him or her and potential signers. Who wants to sign a petition when someone is telling them that their picture will appear on a blog with the caption "racist" attached to it? Finally, in the closing days there were rumors that opposing groups were illegally paying for petition signatures and then discarding those signatures.

In Nebraska, while not yet as heated as Missouri, it has come to our attention that opponents are organizing a campaign to have individuals sign names other than their own on the Nebraska Civil Rights Initiative petitions. This, of course, is illegal and makes checking signature validity more than difficult.

In Colorado, opponents are creating an initiative of their own. This version hijacks the language of the Colorado Civil Rights Initiative, but adds an exception. The initiative would ban preferential treatment based on race except when the Supreme Court has already allowed preferences. This is meant to do nothing more than confuse voters. It is rumored that a similar tactic is being considered in Arizona.

Yet, when given the chance, the voters have demonstrated overwhelmingly their opposition to race preferences. Not all states allow the people to make these decisions on their own. Only seventeen states allow for the people to initiate constitutional amendments, and four states (California, Washington, Florida, and Michigan) already have language prohibiting discrimination and preferential treatment, either through a citizen-approved initiative or an executive order—as in Florida. Super-Tuesday for Equal Rights was our attempt to quickly bring five more states under the fold. As of now, it appears that three of those states (Arizona, Colorado, and Nebraska) will have the opportunity to join those states free of preferences in 2008 and the other two (Missouri and Oklahoma) will have to regroup, as was done in Michigan, for the 2010 election.

Ultimately, however, the nation is poised to end the era of race preferences long before Justice O'Connor's twenty-five-year sentence has expired in 2028. One can only hope that with a critical mass of voters choosing to end race preferences, politicians will finally have the fortitude to join the people and recognize that people should be judged based on their individual merit and not by the color of their skin.

Corporations, Securities & Antitrust

Sovereign Wealth, Private Equity, and Hedge Funds . . . Oh My

*By Jeffrey H. Ballabon**

Americans have long exhibited a suspicion of concentrated pools of capital controlled by small groups of people. During convulsive economic times, with little understanding as to the causes and great fear as to the effects of the turmoil, we have tended to the diversion of scapegoating paranoia. Alan Brinkley's 1982 book on Depression-era populists Huey Long and Father Coughlin captures this mood well in describing how the two demagogues railed against "large, faceless institutions; wealthy, insulated men; vast networks of national and international influence: all exercising power and controlling wealth that more properly belongs in the hands of ordinary citizens."

The last time we experienced disruptions in our financial system on a scale like we have seen of late—in the wake of The Great Depression—the regulatory response was massive, and the legal edifice that was erected, in the form of our still regnant banking and securities laws, was designed in particular to divorce investors' pecuniary interests from their ability to control American industry in which they are invested.

The enacted restrictions and prohibitions were not undertaken lightly—nor were their hazards to economic efficiency misunderstood. Washington intended to prevent financial institution control of industrial companies. Beyond mere mistrust of Wall Street, this was an attempt by regulators to enact a broad-ranging regime that would guide (or restrict, depending on your point of view) the growth of the financial services industry and our capital markets for the foreseeable future.

In its 1934 report on stock exchange practices, the Senate Committee on Banking and Currency argued that investment companies had become "the instrumentality of financiers and industrialists to facilitate acquisition of concentrated control of the wealth and industries of the country." The report urged Congress to "prevent the diversion of these trusts from their normal channels of diversified investment to the abnormal avenues of control of industry."

At the time, the nascent mutual fund industry was a great scapegoat and the perception of risk likely was overstated. Lawmakers wanted to protect against the eventuality that the mutual funds would abuse the resources of their portfolio companies. But in hindsight, the resulting regulatory regime seems like a solution that was in search of a problem, and the distancing of shareholders from company operations has led to some of the worst excesses of corporate abuse.

Today, distrust of concentrated pools of capital continues, aimed at a new crop of suspects: sovereign wealth, private equity, and hedge funds. Now, as then, populist rhetoric is rising and with it the clamor for government action. Adding urgency is the view of many politicians in Washington that private equity and hedge fund managers are ripe, low-hanging fruit with enough juice to fund numerous social programs.

Sovereign Wealth Funds

The U.S. is the biggest recipient of foreign direct investment. The International Trade Commission reports that in 2006 the U.S. received over $465 billion from foreign investors—amounting to roughly 13.5% of U.S.'s gross domestic product. U.S. capital markets offer a stable and predictable legal system, relatively low taxes, and access to the most coveted consumer market on the globe. There is a global trend developing of sovereign governments forming massive pools of capital specifically to invest in the U.S.

Conceivably, foreign powers could have sinister reasons for wanting to invest in the U.S.: their funds could be used to destabilize financial markets, protect their own domestic industries or even to expropriate security-sensitive technologies. Yet such risks have not materialized in any appreciable way. Political and industrial espionage are as old as nation-states, but there is scant evidence, if any, that sovereign wealth funds have served as Trojan horses for nefarious activities.

And suspicion has real costs. If we turn away sovereign wealth fund capital, innovation will be stifled, productivity reduced, economic growth undermined and employment depressed. Between 2003 and 2007, over 3,300 new projects were announced or opened on account of foreign investment, yielding $184 billion in investment and about 447,000 new jobs in the U.S. We need these numbers to grow, not recede.

Moreover, numerous mechanisms, including banking and export controls, mitigate the risk of foreign ownership of sensitive assets. In addition, the Department of Commerce's Invest in America initiative and the Treasury Department's working group on sovereign wealth funds are working with foreign governments to establish voluntary protocols regarding transparency, stability and security.

Recently, the Treasury Department proposed regulations that would appear to expand the scope of review of foreign investments, including sovereign wealth funds. These regulations confirm that investments below 10% of a U.S. business may be subject to review and approval by the Committee on Foreign Investment in the United States (CFIUS). CFIUS is an inter-agency committee chaired by the Secretary of Treasury which reviews foreign investment transactions with the aim of safeguarding national security. Under these proposed regulations the threshold for CFIUS review includes situations where an investment provides the foreigner with "the power... to determine, direct, or decide important matters affecting the entity." Previously, a safe harbor for less than 10% investments was thought to apply.

....

** Jeffrey H. Ballabon, Esq., a former counsel to the Senate Commerce Committee, is President of the Ballabon Group LLC, a government relations and crisis management firm in New York and Washington, DC. He gratefully acknowledges the assistance and expertise of his private investment fund clients in the preparation of this article.*

Thus, the move toward limiting foreign capital already has begun, and there is a risk that the trend will expand to endanger much-needed sources of funding in the wake of a devastating credit bubble. So long as sovereign wealth funds exhibit a rational investment strategy designed to maximize profit, we should be wary of any further regulatory restriction, lest we sabotage ourselves.

Private Equity Funds

Private equity firms are a unique and pervasive creature of the capital markets. Structured as long-term, illiquid investment vehicles, these firms typically take controlling positions in private operating companies that have been, among other things, emancipated from a larger public company, transitioned from private ownership, or built by rolling up smaller enterprises into a single firm.

Funding for private equity firms often comes from larger pools of capital, such as pension funds and charitable endowments seeking alternative investments to help generate outsized returns to fund their long-dated liabilities. Historically, the key to private equity firms' success was not only the long-dated nature of their capital but cheap and easy financing in the form of leverage extended by investment banks and then syndicated to other financial institutions. Until recently, with as little as 10-20% down, private equity firms were able to purchase and control enormous enterprises.

With such power comes scrutiny. Lambasted as "locusts," "vultures," and other less flattering epithets—particularly by domestic and foreign free market critics—these firms are under increasing pressure. The high debt loads which fund their purchases are seen by some as a potentially destabilizing force in the capital markets and beyond as economic contraction looms large. It is likely, however, that negative impulses in reaction to private equity are both overstated and misguided, barring the menacingly self-fulfilling prophecy of populist regulatory intervention.

Given their relatively longer investment horizons, private equity funds could prove uniquely positioned to withstand the current credit storm—provided it is not overly prolonged. Private equity firms typically aim to "harvest" (i.e., sell) their investments 3-5 years after acquisition. In the hands of patient capital, private equity portfolio companies can weather the credit crisis with owners who recognize the difference between illiquidity (the short-term inability to freely finance, buy, or sell at full value) and insolvency (a fundamental inability to fund operations on an on-going basis).

As rational actors, private equity firms can be expected to shield fundamentally sound but illiquid businesses and assets until market conditions improve. Similarly, they should be expected to restructure and/or shed businesses and assets that prove to be uneconomical. Moreover, there is no doubt that they will find value in the remains of failed businesses.

And even a wave of private equity led defaults is unlikely to stir up serious systemic risk. A study by the McKinsey Global Institute concludes that private equity borrowing continues to form only a small part of the overall corporate debt market, 11% of overall corporate borrowing in the U.S. and Europe in 2006. According to Diana Farrel, Director of McKinsey Global Institute, if we assume a spike of private equity defaults of 15% from the historic highs of 10%, estimated implied losses would equal only 7% of the 2006 syndicated lending issuance in the U.S. and 3% in Europe. As McKinsey's study points out, private equity-owned companies are worth just 5% of the value of companies listed on U.S. stock markets and 3% of those in Europe.

Unfortunately for private equity funds and their hedge fund cousins (discussed below), the political climate is ominous. The Treasury Department's recently released blueprint for future market regulation envisions much greater scrutiny. Under Treasury's long-term "optimal" plan, the Federal Reserve would act as a safety and soundness regulator with the power to extract "detailed financial information" from any firm viewed as engaging in investments with potentially system-wide effects. The proposal even envisions a central bank that can more or less impose any "corrective actions to address financial stability problems."

More than anything else, private equity firms are allocators of capital. They can be trusted to serve their own profit motives, and as such maximize the value of their holdings. There is reason to doubt that regulators operating under pressure-driven political mandates (even with the advantage of economy-wide information) would do so much better as to justify their intrusion into lawful investment activities.

Hedge Funds

What is a hedge fund? Given their diverse investment strategies, varied investment horizons, assorted sizes and areas of expertise, this is a surprisingly difficult question to answer. The few commonalities of hedge funds are their structure and their payment schemes. In domestic form these funds are typically structured as private partnerships; offshore, as private Cayman corporations. Investors range from wealthy individuals and charitable endowments to pension funds and sovereign states. Hedge fund managers are typically paid a fixed fee (1%-2% of assets under management) and a cut of the profits (typically 20%) every year. One certainty about hedge funds: it is good work if you can get it.

Hedge funds also are the pariahs of the capital markets. Criticized for making too much money and the occasional fantastic flame-out, there is little sympathy for hedge fund managers. Often depicted by the media as "murky" or "secretive" unregulated pools of capital, suspicion of these investment vehicles abounds.

But hedge funds serve important market functions. By introducing specialization in trading strategies and financial analysis, hedge funds help ensure that markets are comprised of many investors with heterogeneous views of value. When market participants disagree on value, they deploy capital in the direction they favor and, importantly, provide liquidity to markets through trading. They bet long (if they expect prices to rise) or short (if they predict a decline), and in so doing facilitate more efficient and accurate market pricing.

Sensible and dependable pricing has extraordinary value well beyond the capital markets, and hedge funds are given far too little credit for this ancillary benefit that emanates from their trading practices. Without effective price feedback loops, all parts of the modern economy, micro and macro, are at risk of misallocating their resources. One would think that busted bubbles (whether of the Internet or credit variety) would teach us the importance of efficient market pricing and caution against limiting players that facilitate price discovery.

In addition, the sheer abundance and diversity of hedge funds provide a counterbalance to the concentrated power of massive global banking institutions. Securities and Exchange Commission (SEC) Chairman William O. Douglas, a key figure in 1930s financial legislation, articulated an overarching goal of fragmenting economic power under the view that "tremendous power" lays in the hands of firms and people who have the ability to dominate financial markets. Hedge funds are a market-based fragmenting of capital (both cash and human) with the salutary effect of dispersing power of banks.

Despite their notoriety, hedge funds are the runts of the capital markets. Neither a dominant force on a relative or absolute basis, these small, nimble players can nonetheless achieve extraordinary results. As hedge funds grow larger and more bank-like, however, there is no doubt that calls for their regulation will increase; but hedge funds already are subject to significant regulation and forced transparency.

As soon as hedge funds amass 5% stakes in listed companies, they are typically forced to disclose publicly their positions (as well as the prices they paid and the timing of each purchase) by way of filings with the SEC. Once they reach 10% ownership positions, every transaction must be publicly reported. Firms that manage in excess of $100 million are required to report publicly the bulk of their listed company holdings on a delayed quarterly basis. Holdings in derivative instruments such as customized options or exotic equity swaps are not publicly disclosed as a matter of requirement, but these financial instruments are traded on the so-called over-the-counter market (i.e., face to face with bank counterparties). As a result, a hedge fund's bank counterparties know what it holds, how much it holds, what it is worth, when it buys and when it sells and there is complete transparency for all such transactions from the bank's end.

Thus, hedge funds hardly are the popularly depicted ravenous vampires of the capital markets, casting neither shadow nor reflection. If regulators truly do view a collective failure of these firms as an emerging source of systemic risk or action, they should acknowledge the benefits they provide and enhance transparency by providing more carrots than sticks as they contemplate regulation.

Finally, transparency to regulators does not necessarily mean greater transparency to other market participants. Private investment partnerships may object less to providing insight to regulators, so long as they do not have to share their ideas (and potential profits) with their competition in the capital markets.

CONCLUSION

We stand again on the brink of significant government intervention in the capital markets. Sovereign wealth, private equity, and hedge funds, although popular political targets, are investment vehicles that provide real economic benefits including much-needed risk capital to fuel growth for the U.S. economy. While they carry with them some risk factors that extend beyond their immediate investors, the primary burden of any losses they may incur are likely to fall on the shoulders of their backers and not on the U.S. citizenry writ large. In contemplating more aggressive intrusion, lest we saddle these vital investment pools with detrimental rules and restrictions, we should consider the following: (1) Limitations on sovereign wealth, private equity, and hedge funds come at an immediate cost we can ill afford; (2) Absent evidence of widespread or significant political scheming, scrutiny of sovereign funds should be, at this stage, limited to a threshold "rational investor" test; and (3) Policymakers should proceed cautiously, skeptical of populist and ideologically anti-capitalist political motivations and pressure.

The SEC's Inglorious Role in Limiting Small Business's Access to Capital

*By Rutheford B. Campbell, Jr.**

One of the most curious and misdirected regulatory approaches of the Securities and Exchange Commission (SEC) is the Commission's relentless refusal to permit small corporations to solicit broadly for external capital.[1] The Commission has over time been ably assisted in this unfortunate approach by state blue sky laws and state securities regulators.[2] As a result, small businesses, which are vital to our national economy and otherwise face enormous structural impediments when they compete for external capital, are further disadvantaged by burdensome, inefficient, and anti-competitive governmental regulatory schemes.

To some extent, it has always been a perfect storm for small businesses in this regard. At the federal level, the SEC has never understood small businesses, the way they raise capital and the obstacles they face in the capital markets.[3] There are also matters of public choice and fashion at work here. A public choice analysis suggests that small entrepreneurs have been unable to overcome the collective action problems they encounter when they compete for efficient rules from the Commission.[4] Relatedly, as a matter of fashion, high profile issues—matters, for example, involving large publicly traded companies and the regulation of public trading markets—dominate the Commission's attention, leaving little agency time for consideration of the problems of small issuers. But whatever the explanation, the Commission has never had the ability, inclination or interest to fix the problems it has created for small issuers.

The other component of the perfect storm is state blue sky laws. States and their securities regulators have the capacity to eviscerate nearly any federal regulation that is sympathetic to the capital formulation needs of small companies and over the years have shown a hostility to legitimate capital formation activities by small companies.[5]

Small entrepreneurs, who already suffer major structural disadvantages in capital formation, are therefore further and significantly disadvantaged by an ineffective and generally disinterested Commission and by misguided and hegemonic state regulators.

The purpose of this article is to make the case for Commission action freeing small companies from regulatory rules that unfairly limit their legitimate capital formation activities. The focus of the article is Regulation D,[6] which is the most likely path small issuers take in order to meet the requirements of the Securities Act of 1933. Regulation D, however, requires issuers in marketing their securities to refrain from any "general solicitation or general advertising."[7] State blue sky laws also effectively prohibit any general solicitation by small businesses attempting to rely on Regulation D.

The Commission can—and should—eliminate both the federal and state prohibition against general solicitations in Regulation D offerings. Permitting small issuers to solicit broadly in a Regulation D offering would improve small businesses access to external capital without any loss of investor protection.

A. The Important But Disadvantaged Place of Small Businesses in our Economy

Data demonstrate the importance of small businesses to our national economy. There are about 5 million small businesses in the United States (businesses with less than 20 employees).[8] This amounts to almost 90% of all business units.[9] These small firms employ approximately 20 million workers, which amounts to 19% of the nation's entire workforce.[10] If one considers firms with less than 100 employees, those firms employ nearly 40 million workers,[11] amounting to 37% of all jobs nationally.[12]

A somewhat more qualitative evaluation of small businesses may lead one to conclude that even these impressive raw numbers understate the real value of small firms to our nation. Additional data, supplemented by educated estimates, suggest that small businesses may be disproportionately innovative,[13] provide entrepreneurial opportunities for historically disadvantaged groups,[14] and create new jobs disproportionate to their relative size.[15]

Data and educated estimates, therefore, confirm the apparent—that small businesses are hugely important to our nation.

Data also confirm what is apparent to everyone, which is that small businesses need access to external sources of capital.[16] Starting a new business and maintaining that business, especially as the business becomes successful and begins to expand, inevitably requires external capital.

Small businesses' search for external capital is hampered by significant economic and structural impediments. One impediment is that small firms usually need small amounts of capital. This means that expenses in a small offering—legal, accounting and offering expenses[17] —will be very high relative to the size of the offering, and it is relative, not absolute, offering costs that foreclose small firms from the capital markets.[18] To use extreme examples to make this point, offering expenses of $90,000 will likely foreclose an offering of $100,000 by a company, but the same $90,000 offering costs will not foreclose a $20 million offering by a company.

The other significant structural impediment that small businesses face is the absence of financial intermediation services. Underwriting services are not available for small offerings. The modest size of offerings by small businesses will not support underwriting fees sufficient to compensate underwriters for their efforts in investigating, learning, and selling the securities of a small issuer. The unavailability of those intermediation services are a significant disadvantage to the small company attempting to access external sources of capital.[19]

This means that small companies—companies whose significance to the economy may be under described by reference to the approximately 20% of all jobs that they provide—are disadvantaged by exogenous factors that in some

** Rutheford B. Campbell, Jr. is the Everett H. Metcalf, Jr. Professor of Law at the University of Kentucky.*

cases exclude them from the capital markets and in all cases drive up their costs of capital beyond the efficient cost levels encountered by larger firms.

B. The Overall Theory of Federal Securities Laws: Disclosure Philosophy, Mandated Disclosures, and Exemptions from Mandated Disclosures

Consider the philosophy and purpose of the Securities Act of 1933 in light of the exogenous structural impediments small businesses face when they attempt to raise external capital.

The 1933 Act is based on a disclosure philosophy.[20] The cornerstone of the 1933 Act is Section 5,[21] which mandates disclosure of closely prescribed investment information by companies selling their securities into the capital markets. Since efficient trading depends on fully informed parties, a rule requiring issuers to disclose investment information to purchasers appears, at least preliminarily, to be sound.

The problem, however, is that mandated disclosure may cause a significant drag on capital formation. This can be illuminated by imagining excessively burdensome disclosure requirements, which drive up the issuer's transaction costs[22] and in turn diminish the value of the trade between the issuer and investors or, in the worst case, actually exclude the issuer from offering its securities.[23] Alternatively, one may describe the problem as mandating inefficient levels of disclosure. Because the level of disclosure required in a particular transaction is dictated by the government rather than the parties themselves, it may result, for example, in more information (and thus more expense) for the parties than they would have agreed upon, were they free themselves to set the level of disclosure.[24] The top down rule, in other words, destroys part of the value of the trade and thus diminishes the incentive for value creating transactions.

In an attempt to ameliorate the economic problems caused by a ubiquitous application of a mandated disclosure rule and to balance the competing interests of capital formation and investor protection, Congress and the SEC have carved out a number of exceptions to the registration and prospectus delivery requirements of Section 5. Generally, these exceptions may be seen as involving situations in which the parties to the sale of securities are in a position cheaply to acquire for themselves investment information necessary for efficient trading. So, for example, if parties have geographic proximity to one another[25] or have access to information as a result of position or economic bargaining power,[26] exemptions from the registration and prospectus delivery requirements remove governmentally prescribed and mandated disclosure rules and permit the parties themselves to fashion their own levels of disclosure.

While the 1933 Act from the beginning implicitly recognized the need to balance investor protection with capital formation, Congress later amended the 1933 Act explicitly to require that Commission rules strike an appropriate balance in that regard. Section 2(b) of the 1933 Act now mandates that the Commission, when in its rulemaking it is required to consider the "public interest," is to consider the effect of its action not only on "the protection of investors" but also on "efficiency, competition and capital formation."[27]

Notwithstanding such implicit and now explicit mandates from Congress, the Commission over long decades has refused to accord appropriate consideration to the capital formation needs of small issuers. One of its important failures in that regard has been the impediments it has constructed to the efficient search for external capital by small businesses, and the most pernicious of those impediments may well be the prohibition on general solicitations in Regulation D.[28]

C. Regulation D

Regulation D[29] is a regulatory exemption from the registration requirement of Section 5 and, facially, at least, is consistent with the Commission's obligation to balance investor protection and capital formation.

Rules 504, 505 and 506 of Regulation D provide exemptions from registration, and predicate the availability of the exemptions on more investor protection requirements as deals get larger. The investor protection devices in the Regulation are disclosure[30] and purchaser qualification requirements (*e.g.,* purchaser sophistication).[31] Rule 504 provides an exemption for offering up to $1 million and imposes no disclosure or purchaser qualification requirements. Rule 505 provides an exemption for offerings up to $5 million and imposes a disclosure obligation but no purchaser qualification requirement. Rule 506 provides an exemption without regard to the size of the offering and requires disclosure and purchaser qualification.

This stair-stepped approach—Brequiring additional investor protection as the size of the transaction increases—is a sound philosophy, but the prohibition in Rules 504, 505, and 506 against any general solicitation for investors undercuts the claim that Regulation D strikes a sensible balance between investor protection and capital formation. The prohibition against general solicitation significantly and adversely affects the ability of small issuers to find external capital but offers no material protection to investors.

D. Prohibition Against "General Solicitation"

The prohibition in Regulation D against any "general solicitation" has two components—"general" and "solicitation."

The term "solicitation" should be understood to have the same meaning as "offer" under the 1933 Act. It is a broad term that applies to any action undertaken by an issuer for the purpose of facilitating a sale of its securities. Under such a purpose or intent test, activities by an issuer that are intended to condition the market for a sale would amount to an "offer" or "solicitation". Thus, even those activities that fall far short of a formal or common law offer would amount to a "solicitation" under Regulation D.[32]

The definition of "general" has always been something of a mystery. At first blush, one might think of "general" as meaning a large number. So, a large number of solicitations (offers) may amount to a "general" solicitation. The Commission, however, has never put a quantitative limit on "general."

Pursuing a more indirect interpretative path—relying on Commission releases, no action letters and scholarly interpretations—one may conclude the following regarding when solicitations or offers reach the level of "general": (1) offers are "general" if they are likely to reach an undetermined number of offerees;[33] (2) offers limited to sophisticated or accredited offers may still be "general;"[34] (3) indiscriminate offers—those not preceded by some screening or vetting—are more likely to be considered "general" than offers in which offerees are screened or vetted;[35] (4) notwithstanding the absence of any specific number test, numbers are important, and the more offerees the more likely the solicitation is to be "general;"[36] and (5) a pre-existing relationship between the offeree and the issuer or issuer's agent reduces the likelihood that the solicitation (offer) is "general."[37]

The purpose of listing these five factors, however, is not to illuminate the line between actions that are "general" and non-general. Instead, the purpose is to show that the concept of "general" is broad both in scope and marginal ambiguity and, as a result, is effective in precluding issuers that rely on Regulation D from an efficient search for external capital.

It is clear that an issuer who attempts to identify potential investors through the use of any medium of wide circulation, including newspapers, radio, TV or the internet, is involved in a "general" solicitation and thus precluded from using Regulation D.[38] It is also clear that even more limited methods of identifying potential investors may involve levels of risk the cause reliance on Regulation D to be economically irrational. Assume, for example, that a small issuer wishes to use Regulation D as a way to raise $2 million in equity. In order to identify potential investors, the issuer proposes to send a letter to 150 persons offering the opportunity to invest in the offering. If that letter creates a 0.3 probability of amounting to a "general" solicitation[39] and thus destroying the availability of Regulation D and creating a potential $2 million liability for the issuer, a rational issuer may be unwilling to accept that amount of residual risk. Thus, the broad marginal ambiguity of Regulation D may make the exemptions practically unavailable to issuers, even in instances in which conduct may have a relatively low probability of amounting to a "general" solicitation.

Precluding issuers that rely on Regulation D from an efficient search for investors involves costs—both to the issuer and to society—for which there are no comparable benefits. Indeed, it is impossible to find any material benefit that is generated by limiting the issuer's ability to offer its securities broadly, so long as appropriate investor protection devices are effectively in place *at the point of sale*.

Consider, for example, the investor protection devices of Regulation D, which are disclosure and investor qualification requirements (sophistication or accreditation). The effectiveness of neither of these protections would in any way be compromised by allowing companies to solicit broadly for investors, so long as those investor protection devices were effectively in place at the time of sale.

If the Commission were to eliminate the prohibition against general solicitations in Regulation D, small issuers searching for capital would be free to solicit broadly for investors, using, if they so chose, radio, TV, newspapers, periodicals, internet, etc. Issuers could also use less expansive investor identification techniques, such as sending solicitation letters to 150 potential investors or unlimited calls to friends and business associates. In all such cases, however, investor protection requirements would be imposed at the relevant point, which is at the time of sale. Thus if the particular Regulation D offering requires both disclosure and investor qualification, the issuer would have to ensure that any broadly solicited offerees who became purchasers of the offering were at the point of sale qualified and fully informed.[40] The effectiveness of those investor protection provisions would be uncompromised by the broad solicitation.

The simple and effective prescription, therefore, is for the Commission in Regulation D to select the investor protection devices that are appropriate, always balancing (as they are obligated to do) investor protection and capital formation. These protection devices should be imposed prior to sale, leaving issuers relying on Regulation D free—subject only to antifraud rules—to solicit investors broadly.

E. The Role of the States and NSMIA

While the elimination of the prohibition against general solicitation would be a relatively simple administrative matter for the Commission,[41] achieving the final desired result, which is to free small issuers to solicit broadly for capital, implicates another formidable obstacle, and that is state securities laws.

At the present time, issuers offering their securities under Rule 504 or Rule 505 of Regulation D are subject to state securities registration requirements, since the National Securities Markets Improvement Act of 1996 (NSMIA) did not preempt state registration requirements for offerings under Rule 504 or 505.[42] As a result, issuers relying on those exemptions under Regulation D are likely meet state registration requirements by qualifying for either the state's small offering exemption from registration[43] or its Uniform Limited Offering Exemption (ULOE).[44] The small offering exemption is a statutory exemption and is typically limited to a very few offerees.[45] The ULOE is a state regulatory exemption predicated on the offering's meeting the requirements of federal Rule 505 or Rule 506 and additional requirements under ULOE designed to enhance investor protection.[46]

It is highly unlikely that either of these state exemptions would be available for a Regulation D offering that permitted a general solicitation. As described above, the state small offering exemption has strict quantitative limitations on the number of permissible offerees,[47] which would prohibit any general solicitation. As concerns the availability of the ULOE, state regulators would certainly resist any state coordination with a Regulation D offering that permitted general solicitations. State securities regulators have a history of resisting general solicitations for exempt offerings[48] and a demonstrated willingness aggressively and effectively to protect their own administrative turf.[49]

The Commission, however, has two paths by which it could prevent states from neutralizing a federal rule that permitted a general solicitation for investors in a Regulation D offering. First, the Commission could by its own regulation expand NSMIA's preemption. Under NSMIA, Congress delegated to the Commission authority to expand the federal preemption over state securities regulation to offerings made to "qualified purchasers," as defined by the Commission.[50] Both the 1933 Act itself and the history of NSMIA strongly suggest that the Commission would be well within its delegated authority to define a "qualified purchaser" as including one who purchases in an offering under a revised version of Regulation D that permitted general solicitations.[51]

The second path open to the Commission is to lead a legislative initiative to expand NSMIA's preemptive scope. A complete federal preemption of state registration requirements is certainly the preferred prescription for the longstanding and significant pernicious effects that state blue sky laws have caused in efficient capital formation, especially capital formation by small issuers.[52]

Whichever option it chooses, neutralizing state hegemony over federal policy is essential if the SEC is ever to construct a Regulation D that permits a broad search for investors by small companies. It is certain that state regulators would fight that move by the SEC, but it is time for the Commission to exercise its own hegemonic advantage for the benefit of small entrepreneurs and the economy.

CONCLUSION

Small businesses are essential to our national economy, and efficient access to external capital is essential to small businesses. Structural obstacles—small capital needs and the absence of financial intermediation—put small businesses at a significant disadvantage, when they compete for external sources of capital. Federal and state securities rules that prohibit a broad solicitation for external capital exacerbate this problem.

The Commission has for too long ignored the pernicious effects of its own regulations on the legitimate capital formation needs of small issuers and has been inappropriately deferential to the misdirected actions of well meaning but overly zealous state securities regulators. Small businesses and the rest of us have been the losers in this.

The Commission should take steps to ensure that Regulation D is available for small issuers that solicit broadly for their external capital.

Endnotes

1 This problem has existed for decades. *See* Rutheford B Campbell, Jr., *The Plight of Small Issuers Under the Securities Act of 1933: Practical Foreclosure from the Capital Market*, 1977 Duke L.J. 1139 (hereinafter Campbell, *The Plight Under the 1933 Act*); Rutheford B Campbell, Jr. *The Plight of Small Issuers (and Others) Under Regulation D: Those Nagging Problems That Need Attention*, 74 Ky. L.J. 127 (1985-86) (hereinafter Campbell, *The Plight Under Regulation D*).

2 *See* Rutheford B Campbell, Jr. *An Open Attack on the Nonsense of Blue Sky Regulation*, 10 J. Corp. L. 553 (1985) (hereinafter, Campbell, *An Open Attack*).

3 Once again, this is a long standing problem. For example, as originally promulgated, Rule 144, 17 C. F. R. ' 230.144 (1975) was practically unavailable for resales by investors in small companies, because they were unable to meet its requirements. Nonetheless, the Commission's release adopting Rule 144 blatantly attempted to force all resales of restricted securities into Rule 144 transactions. *See* Campbell, *The Plight Under the 1933 Act*, *supra* note 1, at 1150-54.

4 *See* Rutheford B Campbell, Jr., *The Impact of NSMIA on Small Issuers*, 53 Bus. Law. 575, 583-85 (1998) (hereinafter Campbell, *The Impact of NSMIA*).

5 *See* Campbell, *An Open Attack, supra* note 2; Rutheford B Campbell, Jr., *The Insidious Remnants of State Rules Respecting Capital Formation*, 78 Wash. U. L.Q. 407 (2000) (hereinafter Campbell, *The Insidious Remnants*).

6 17 C.F.R. '' 230.501-.508 (2008).

7 The prohibition against general solicitation, 17 C. F. R. ' 230.502(c) (2008), is incorporated by reference into the requirements for an exemption under Rule 504, 17 C. F. R. ' 230.504(b)(1) (2008), under Rule 505, 17 C. F. R. ' 230.505(b)(1) (2008), and under Rule 506, 17 C. F. R. ' 230.506(b)(1) (2008).

While other requirements of Regulation D make that exemption less than perfectly efficient for small issuers, it is the prohibition on general solicitations that is most harmful to the legitimate efforts of small companies attempting to access external capital efficiently. *See* Campbell, *The Plight Under Regulation D*, *supra* note 1, at 136-43; Daugherty, *Rethinking the Ban on General Solicitation*, 38 Emory L.J. 67 (1989).

8 U.S. Small Business Administration, The State of Small Business: A Report of the President 2000 (hereinafter The State of Small Business 2000), at 61, Table A.4.

9 *Id.* at 61, Table A.4 (reporting that 89.4% of firms had less than 20 employees).

10 *Id.* (reporting that 18.8% of all workers were employed by firms with less than 20 employees).

11 *Id.* (reporting that firms with less than 100 employees employed 39,653,019 workers).

12 *Id.* (reporting that firms with less than 100 workers employed 36.7% of all workers).

13 U.S. Small Business Administration, The State of Small Business: A Report of the President 1994 (hereinafter, The State of Small Business 1994), at 15 (estimating that small businesses and their employees generate proportionately more innovations than larger businesses); U.S. Small Business Administration, The State of Small Business: A Report of the President 1998 (hereinafter The State of Small Business 1998), at 3 (then President Bill Clinton referred to small businesses as "continual sources of new ideas... and their experimental efforts are an essential part of the organic and ever-changing American economy.").

14 The State of Small Business 2000, *supra* note 8, at 102, Table A.14 (showing increases in minority and female ownership of businesses); *id.* at 17 (A[s]mall business continued to be an important means by which women, minorities, and immigrants entered the American economic mainstream and managed to increase their share in the economy.).

15 *Id.* at 84, Table A.9 (reporting that 50.2% of all new jobs were created by small firms) .

16 *See* Rutheford B Campbell, Jr., *Regulation A: Small Businesses' Search for "A Moderate Capital"*, 31 Del. J. Corp. L. 77, 86-88 (2006) (hereinafter Campbell, *"A Moderate Capital"*) (discussing and providing data that demonstrate small businesses' resort to external sources of capital); The State of Small Business 1998, *supra* note 13, at 3 (reporting that approximately 85% of firms with 10 to 19 employees relied on external capital).

17 James D. Cox, Robert W. Hillman & Donald C. Langevoort, Securities Regulation Cases and Materials 149 (4th ed. 2004) (*hereinafter* Cox, Hillman & Langevoort, Securities Regulation (4)) ("reporting a 'recent estimate' for the expenses of an IPO as '$150,000-$300,000 in fees to counsel; $100,000-$150,000 for the audit; $10,000-$20,000 for underwriter counsel expenses... and $50,000-$100,000 for printing costs'").

18 *See* Campbell, *"A Moderate Capital"*, *supra* note 16, at 88-90.

19 *See* Ronald J. Gilson & Reinier H. Kraakman, *The Mechanisms of Market Efficiency*, 70 Va. L. Rev. 549, 613-21 (1984) (describing the value of financial intermediation to firms selling securities); and Donald C. Langevoort, *Angels on the Internet: The Elusive Promise of "Technological Disintermediation" for Unregistered Offerings of Securities*, 2 J. Small & Emerging Bus. L. 1, 16-18 (1998) (citing social science-based research "that also speaks to the role of intermediation in the capital-raising process." *Id.* at 16.

20 Louis Loss & Joel Seligman, I Securities Regulation 171-93 (3d ed. 1989) (discussing the battle over philosophies for federal securities laws).

21 15 U.S.C. ' 77e (2008).

22 *See* Cox, Hillman & Langevoort, Securities Regulation (4), *supra* note 17.

23 In economic terms, if transactions costs exceed the value created by the trade, rational parties will not trade. Thus, it is highly unlikely that a company will make a $100,000 offering if it costs $90,000 to make the mandated disclosures.

24 *See, e.g.,* Richard A. Posner, Economic Analysis of Law 481 (7th ed. 2007) ("Capital markets are competitive, and competitive markets generate information about products sold.").

25 The intrastate exemption provided by Section 3(a)(11) of the 1933 Act, 15 U.S.C. ‘ 77c(11) (2008), and Rule 147 promulgated thereunder, 17 C.F.R. ‘ 230.147 (2008), predicate the availability of the exemption on the issuer's and the investors' being located in the same state.

26 *See, e.g.*, SEC v. Ralston Purina Co., 346 U.S. 119 (1953) (listing as important to the availability of the exemption in Section 4(2) factors such as investors' ability to "fend for themselves" and their "access to the same kind of information that registration would disclose").

27 15 U.S.C. ‘ 77b(2) (2008).

28 Cohn, *The Impact of Securities Laws on Developing Companies: Would the Wright Brothers Have Gotten Off the Ground?*, 3 J. Small & Emerging Bus. L. 315, 355 (1999) ("No limitation characterizes the phobia of securities regulators more than the prohibitions against general advertising and solicitation."). *See* Daugherty, *supra* note 7.

29 17 C.F.R. ‘' 230.501-.508 (2008).

30 Rule 502(c), 17 C.F.R. ‘ 230.502(c) (2008), which prohibits any "general solicitation or general advertising," is incorporated by reference into Rule 504, 17 C.F.R. ‘ 230.504(b) (2008), Rule 505, 17 C.F.R. ‘ 230.505(b)(1) (2008), and Rule 506, 17 C.F.R. ‘ 230.506(b)(1) (2008).

31 Rule 506(b)(ii), 17 C.F.R. ‘ 230.506(b)(2)(ii) (2008), requires each purchase to be either an "accredited investor" or sophisticated.

32 *See* Campbell, *The Plight Under Regulation D*, *supra* note 1, at 137-40 (describing SEC no-action letters and release that provide define "offer" or "solicitation").

33 *See, e.g.*, Aspen Grove, SEC No-Action Letter (Dec. 2, 1982) (*available on* Lexis Fedsec library, No act file) (offer in a thoroughbred horse trade publication).

34 *See, e.g.,* The Texas Investor Newsletter, SEC No-Action Letter (Jan. 23, 1984) (*available on* Lexis, Fedsec library, No act file) (offer to 2,000 accredited investors is "general").

35 *See, e.g.,* Aspen Grove, SEC No-Action Letter, *supra* note 33 (distribution of offering materials at a horse sale without vetting or screening of distributees).

36 *See* Campbell, *The Plight Under Regulation D*, *supra* note 1, at 141.

37 *See* the discussion in James D. Cox, Robert W. Hillman & Donald C. Langevoort, Securities Regulation Cases and Materials 302-310 (5th ed. 2008) (hereinafter Cox, Hillman & Langevoort, Securities Regulation (5)) (discussing the importance of pre-existing relationship, *id.* at 302, solicitations by brokers on behalf of issuers, *id.* at 304, and match-making services or investment databases, *id.* at 306).

38 *See, e.g.,* Sec. Release No. 7233, ex. 20 (Oct. 6, 1995) (offering "on the internet would not be consistent with the prohibition against general solicitation").

39 *See* Campbell, *The Plight Under Regulation D*, *supra* note 1, at 142 (opining that offerings to 100 pre-screened investors should not be considered "general").

40 Under the present version of Regulation D, when disclosure is required, it must be accomplished "a reasonable time prior to sale," 17 C.F.R. ‘ 230.502(b)(1) (2008), and the investor qualification requirements of Rule 506 are imposed respecting each "purchaser." 17 C.F.R. ‘ 230.506(b)(2)(ii) (2008).

41 One technical matter for the Commission would be whether Rule 506 could remain a Section 4(2) exemption without a prohibition on general solicitations, since common law interpretations of Section 4(2), 15 U.S.C. ‘ 77d(2) (2008), seem to define "public offering" by reference to offers rather than sales. *See* Cox, Hillman, Lanevoort, Securities Regulation (5), *supra* note 37, at 283 ("As... virtually every... case on Section 4(2) make[s] clear, the critical inquiry concerning the private offering exemption is the need of the offerees B not just the purchasers...."). The Commission could solve this problem by re-enacting Rule 506 as an exemption authorized by its delegated authority in Section 28, 15 U.S.C. ‘ 77bb (2008).

42 15 U.S.C. ‘ 77r (2008) (preempting state registration provisions respecting registration for "covered securities", which do not include securities offered under Section 3(b) of the 1933 Act). NSMIA did, however, preempt state securities laws in the case of Rule 506 offerings. 15 U.S.C. ‘ 77r(b)(4)(D) (2008). If the Commission were to re-enact Rule 506 as a Section 28 exemption as a way to permit a general solicitation, *see supra* note 41, securities issued under that version of Rule 506 would not within today's version of NSMIA be "covered securities" and thus not be eligible for preemptive status.

43 *See, e.g.*, Unif. Sec. Act ‘ 402(b)(9), 7B U.L.A. 601-602 (1958).

44 *See* Uniform Limited Offering Exemption, NASAA Rep. (CCH) & 6201 (Oct. 1990).

45 *See, e.g.*, Unif. Sec. Act ‘ 402(b)(9), 7B U.L.A. 601-602 (1958) (ten offerees in a twelve month period).

46 *See* Uniform Limited Offering Exemption, NASAA Rep. (CCH) & 6201 (Oct. 1990).

47 *See, e.g.*, Unif. Sec. Act ‘ 402(b)(9), 7B U.L.A. 601-602 (1958) (ten offerees in a twelve month period).

48 *See, e.g.*, Mark A. Sargent, *A Future for Blue Sky Law,* 62 U. Cin. L. Rev. 471, 477-79 (1993); and Marc J. Steinberg, *The Emergence of State Securities Laws: Partly Sunny Skies for Investors*, 62 U. Cin. L. Rev. 395, 408-11 (1993), which describe the resistence of state securities administrators to changes in Regulation A.

49 State regulators operating through NASAA were, for example, an important force in reducing the pre-emptive reach of NISMA. *See* Campbell, *The Insidious Remnants*, *supra* note 5, at 413; Campbell, *The Impact of NSMIA*, *supra* note 4, at 585-86.

50 15 U.S.C. ‘ 77r(b)(3) (2008).

51 *See* Campbell, *The Impact of NSMIA*, *supra* note 4, at 586-87 (the 1933 Act the Commission "is able to expand significantly the scope of preemption, limited only by the principles of the protection of investors and the promotion of competition and capital formation." *Id.* at 586).

52 With regard to anticipating resistence that the Commission may face if it were to lead a legislative initiative to expand NSMIA, one should recall that as originally proposed the legislation that became NSMIA preempted all state control over registration, except for offerings under the intrastate exemption. State securities regulators and their organization, the North American Securities Administrators Association, were effective in significantly weakening the final version of NSMIA, constricting the scope of preemption and essentially eviscerating the benefit of the legislation for small businesses. *See* Campbell, *The Impact of NSMIA*, *supra* note 4, at 581-89.

Mercier v. Inter-Tel and the Reformulation of the *Blasius* Standard

*By Andrew Carriker**

On August 14, 2007, the Delaware Court of Chancery decided *Mercier v. Inter-Tel, Inc.*[1] One of the more interesting conclusions the court delivered was that the *Blasius* standard should be reformulated "as a genuine standard of review that is useful for the determination of cases, rather than as an after-the-fact label placed on a result."[2] "Such a reformulation," the court said, "would be consistent with prior decisions recognizing the substantial overlap between and redundancy of the *Blasius* and *Unocal* standards, and would have the added benefit of creating a less prolix list of standards of review."[3]

The *Blasius* standard is applied to the action of a board of directors taken "for the primary purpose of thwarting the exercise of a shareholder vote."[4] Even if the board's action is taken in subjective good faith, the board must show that it had a "compelling justification" to take the challenged action.[5] Members of Delaware's judiciary, including Vice Chancellor Strine, who authored the opinion in *Mercier*, have questioned whether a *Blasius* standard need exist at all, when existing standards, particularly the *Unocal* standard, seem to be sufficient for cases that would otherwise be reviewed under *Blasius*.[6] In *Mercier*, the Delaware Court of Chancery attempts to place the *Blasius* standard in the context of the *Unocal* standard.

Facts in *Mercier*

Inter-Tel had been courted by potential buyers since 2005, and since that time a special committee of independent directors had been formed to consider the various proposals the company was receiving. By fall 2006, Inter-Tel's former chief executive officer, Steven G. Mihaylo, the owner of 19% of the company's stock and a private equity partner, had withdrawn an offer to purchase all of the company's stock at $23.25 per share, after Inter-Tel's stockholders voted against a resolution calling on the company's board of directors to sell the company in an auction. In spring 2007, however, Inter-Tel announced that its board had approved a merger agreement with Mitel Networks Corporation at $25.60 per share. The merger agreement contained a no-shop provision that was subject to a "fiduciary out" permitting the board of Inter-Tel to consider an unsolicited alternative proposal that was reasonably likely to lead to a superior proposal.

On June 4, 2007, Mihaylo proposed a transaction in which the company would use a combination of cash on hand and new debt to acquire up to 60% of its own shares at a price of $28 per share. Mihaylo believed that, based on the corporation's expected earnings, Inter-Tel's remaining shares would trade at almost $30 per share. The court refers to this proposed transaction as the "Recap Proposal." Four days later, Mihaylo disclosed his intention to seek control of the board at the next annual meeting if the company's merger with Mitel were to be defeated.

In the weeks preceding the special meeting, major stockholders showed a preference for the Recap Proposal, even though Inter-Tel's special committee tried to explain to stockholders that the Recap Proposal was flawed and that it recommended instead the merger with Mitel. After Institutional Shareholder Services ("ISS") recommended that stockholders vote against the merger and Mitel refused to increase its offer, the special committee realized that if the special meeting were held, the merger, and an adjournment of the meeting to seek the votes required to approve the merger, would almost certainly be defeated.

The special committee then began to consider postponing the meeting. According to the court, the special committee reviewed a number of factors that might occur if the meeting were delayed, including whether the electorate would more closely reflect the actual ownership of the company's shares as of the time of the vote and whether arbitrageurs would buy additional shares that they would be likely to vote in favor of the merger. In addition, the special committee considered whether stockholders would be more inclined to approve the merger because Inter-Tel's earnings were down in its most recent fiscal quarter (and thus Inter-Tel was becoming a less desirable investment to potential buyers) and credit markets were tightening.

On June 29, 2007, before the meeting scheduled for that day had commenced, the special committee postponed the meeting. On July 6, 2007, Inter-Tel announced preliminary results for its second fiscal quarter and disclosed that those results had fallen short of the projections contained in the company's merger proxy statement. That same day, the company also stated that it expected its results for the full fiscal year 2007 to be well below previous estimates.

ISS later reversed its position and recommended that stockholders approve the merger, and shortly thereafter Mihaylo withdrew his Recap Proposal. At the meeting on August 2, 2007, over 87% of Inter-Tel's outstanding shares were voted, of which almost 72% voted to approve the merger. Of the shares not controlled by Mihaylo, more than 90% of the shares voted were in favor of the merger.

Standard of Review

The plaintiff stockholder sought review of the special committee's postponement of the meeting under the *Blasius* standard.[7] For its part, the special committee sought review of its actions under the business judgment rule.[8] The special committee relied particularly on *In re the MONY Group Inc. S'holder Litig.*, a case in which the court declined to apply the *Blasius* standard and instead reviewed the defendant board's actions under the business judgment rule.[9]

The court instead applied "a reasonableness standard consistent with the *Unocal* standard."[10] Under this reasonableness standard, the court required that the board of directors of Inter-Tel (a) identify a "legitimate corporate objective," one that was

** Andrew Carriker is an Associate in the Wilmington (DE) office of Skadden, Arps, Slate, Meagher & Flom LLP. He would like to thank Brian G. Lenhard for reviewing a draft of this article. The views expressed here are his own, and not necessarily those of his firm.*

"proper and not selfish," served by its decision to postpone the meeting and set a new record date and (b) show that the board's actions were "reasonable in relation to their legitimate objective" and "did not preclude the stockholders from exercising their right to vote or coerce them into voting a particular way."[11]

Why did the court hesitate to apply the *Blasius* standard? The court discussed four principal objections to use of the *Blasius* standard. First, the trigger for application of the *Blasius* standard, action by a board of directors that has as its primary purpose the disenfranchisement of stockholders, is itself a conclusion and not a method by which to arrive at a judicial determination.[12]

Second, the requirement that the board of directors show a "compelling justification" for its action is too stringent a standard. Even the Delaware Supreme Court has observed that the *Blasius* standard is so strict that it is "applied rarely."[13]

Third, the relationship of the *Blasius* standard to the *Unocal* standard is unclear.[14] Cases involving corporate elections often arise in the context of mergers and acquisitions, and, as a result, judges sometimes must consider whether to apply both the *Blasius* standard and the *Unocal* standard (or either one or the other standard). The court made noticeable reference to *MM Companies, Inc. v. Liquid Audio, Inc*. and *Chesapeake Corp. v. Shore*, two cases in which the respective courts suggested that an analysis under *Unocal* subsumed analysis under *Blasius*.[15] The court explained that the *Unocal* standard at the same time (a) implicitly requires directors to show that they acted for proper reasons (or, in other words, that they did not act for inequitable purposes) and so addresses concerns about the directors' fiduciary duties and (b) compels directors to identify their "legitimate objectives" and to explain their actions as necessary to advance those objectives.[16]

Fourth, while *Blasius* itself concerned the election of directors, certain passages in the opinion suggest that the court in *Blasius* believed its test ought to be applied in any case in which stockholders were allegedly disenfranchised of their voting rights.[17] The court cautioned that the reasoning in *Blasius* carried less force when the matter to be considered by stockholders had little or no bearing on whether directors would continue in office.[18]

But the court did apply the *Blasius* standard—it simply applied the standard in a modified form and in the context of *Unocal*. The court, for example, stated its conclusions with respect to the second prong of the reasonableness standard in the "compelling justification" language of *Blasius*.[19] By looking to the *Unocal* standard for the analytic tool with which to review the special committee's actions, however, the court reformulated the *Blasius* standard as a standard of review that was subsumed within a *Unocal* analysis and that was not an independent standard of judicial review. To the extent that *Mercier* reformulated the *Blasius* standard, perhaps the court may be said to have engaged in "doctrinal pruning" of the relationship between *Blasius* and *Unocal*.[20]

The court saw its analysis as consistent with *Unocal* and the "directional teaching" of such cases as *Liquid Audio*, *Chesapeake*, and *MONY*.[21] Vice Chancellor Strine further observed, "I do not believe that this test should be used as to director conduct not affecting either an election of directors or a vote touching on matters of corporate control. This test is a potent one that should not be used in garden variety situations, when more traditional tools are available to police self-dealing or improperly motivated director action."[22] In addition, Vice Chancellor Strine specifically rejected the idea that the reasonableness standard used in *Mercier* signalled any tolerance for the concept of "substantive coercion."[23]

Applying this reasonableness standard to the facts of the case, the court found that the special committee acted out of a good faith concern that the merger with Mitel was in the best interests of Inter-Tel's stockholders and that, if the special meeting were held as scheduled and the merger failed to gain approval, the advantages of the merger would be irretrievably lost.[24] After discussing the effects of changing the meeting's record date—primarily how moving the record date could allow arbitrageurs to buy additional shares at a price below the merger consideration—the court found that changing the record date "did not unfairly tilt the odds" against Mihaylo or any other stockholder who opposed the merger.[25] What determined the outcome of the meeting was that ISS and Inter-Tel's stockholders who held stock on both record dates came to view the merger "as the value-maximizing option."[26] Postponement of the meeting and the setting of a new record date, furthermore, neither precluded Inter-Tel's stockholders from freely choosing to reject the merger nor coerced those stockholders into approving the merger.[27] The court concluded its analysis by emphasizing that the "compelling justification" test originating in *Blasius* ought to be replaced in this case with a "legitimate objective" test. But, in deference to the authority of *Liquid Audio* and other cases that "seem to give continuing life to the compelling justification usage," the court made an explicit determination that the special committee "demonstrated a compelling justification for its action, even if that standard applies."[28]

Portnoy v. Cryo-Cell

Whether *Mercier's* reformulation of the *Blasius* standard becomes widely adopted by the judiciary in Delaware remains to be seen. Vice Chancellor Strine has, however, already had the opportunity to comment on the reformulated standard. In *Portnoy v. Cryo-Cell International, Inc*., stockholder David Portnoy proposed a rival slate of directors to that proposed by the management of Cryo-Cell International, Inc.[29] In the course of trying to have the management's slate of directors elected at the company's annual meeting of stockholders, the chief executive officer of Cryo-Cell, Mercedes Walton, entered into an agreement with stockholder Andrew Filipowski to add Filipowski to the board in exchange for his support of the management's slate of directors. Walton further agreed with Filipowski that, should management's slate be elected, Walton would ensure that the board be expanded and that a designee of Filipowski be added to the board. At the annual meeting, Walton extended the meeting until she was certain that the management's slate of directors had secured election, although she did not explain the delays to stockholders in attendance.

The court considered Portnoy's first claim, that Walton and the other incumbent directors of Cryo-Cell violated their fiduciary duties when Walton agreed with Filipowski to provide a seat on the company's board in exchange for Filipowski's support of management's slate of directors, to be a charge of "vote buying" and applied the analysis of *Schreiber v. Carney*.[30] The court noted that the "method for addressing behavior influencing the conduct of a corporate election" that was used in *Mercier* resembled the *Schreiber* test. In both tests, the initial question was whether the board acted with proper motivation. With regard to the second step in each test, the court suggested that the standard in *Mercier* (whether the board's actions were reasonable in relation to their legitimate objective) was more useful than the test of entire fairness in Schreiber.[31]

Portnoy's final complaint was that Walton acted inequitably in her conduct of the annual meeting.[32] In finding that Walton improperly delayed during the meeting without being honest about why she was stalling, the court cited *Mercier* for the test that requires a showing that Walton's actions were "motivated by a good faith concern for the stockholders' best interests, and not by a desire to entrench [herself]."[33] The court distinguished the case at hand, which involved "an actual election of directors, in which the insiders delay because they believe the stockholders are making a mistake in choosing new leadership," from the circumstances presented in *Mercier*.[34] Rather than reverse the election altogether, however, the court ordered that the company promptly hold a special meeting at which a new election would be held and presided over by a special master.[35]

CONCLUSION

The court in *Mercier* presents a somewhat circumscribed holding.[36] Its discussion of the relationship between the *Blasius* standard and the *Unocal* standard, however, represents a significant contribution to the Delaware courts' struggle (as Vice Chancellor Strine characterizes it in *Chesapeake*) to place the *Blasius* standard within the framework of a *Unocal* analysis, when *Unocal* would otherwise govern.[37] Even as the court acknowledges that it is bound by precedent to speak in the language of *Blasius*, the court's reformulation of *Blasius* reflects its dissatisfaction with a broad application of that standard and its preference for the *Unocal* standard of review.

Endnotes

1 929 A.2d 786 (Del. Ch. 2007).

2 *Mercier*, 929 A.2d at 788.

3 *Id.*

4 Blasius Indus., Inc. v. Atlas Corp., 564 A.2d 651, 660, 662 (Del. Ch. 1988).

5 *Blasius*, 564 A.2d at 661.

6 William T. Allen, et al., *Function over Form: a Reassessment of Standards of Review in Delaware Corporate Law*, 56 Bus. Law. 1287, 1311-1316 (2001). *See also, e.g.*, Chesapeake Corp. v. Shore, 771 A.2d 293, 323 (Del. Ch. 2000) ("Given this interrelationship [*sc.* between *Blasius* and *Unocal*] and the continued vitality of *Schnell v. Chris-Craft*, one might reasonably question to what extent the *Blasius* 'compelling justification' standard of review is necessary as a lens independent of or to be used within the *Unocal* frame.").

The *Unocal* standard of review derives from *Unocal Corp. v. Mesa Petroleum Co.*, 493 A.2d 946 (Del. 1985), as interpreted by *Unitrin, Inc. v. American General Corp.*, 651 A.2d 1361 (Del. 1995), and requires that, if the board of directors of a target company takes defensive measures in the face of an unsolicited bid for control, the board must have reasonably perceived that the proposed takeover was a threat to the corporation's effectiveness and policy, and the board's defensive measures must have been reasonable in response to the threat, any defensive measure that precludes or coerces stockholder choice being unreasonable.

7 *Mercier*, 929 A.2d at 804-05.

8 *Id.* at 805. The business judgment rule is "a presumption that in making a business decision the directors of a corporation acted on an informed basis, in good faith and in the honest belief that the action taken was in the best interests of the company." Smith v. Van Gorkom, 488 A.2d 858, 872 (Del. 1985) (quoting Aronson v. Lewis, 473 A.2d 805, 812 (Del. 1984)).

9 853 A.2d 661 (Del. Ch. 2004). In *MONY*, the board of directors had postponed a meeting at which stockholders were to vote on a proposed merger. The board had postponed the meeting so that stockholders could be provided additional information and so that the electorate would more closely reflect the actual ownership of the company's stock. The directors were disinterested in the transaction (they would lose their seats on the board if the merger were completed), and the postponement had no preclusive or coercive effect and did not disenfranchise stockholders of their voting rights.

10 *Mercier*, 929 A.2d at 810.

11 *Id.* at 810-11.

12 *Id.* at 806.

13 *Id.* at 806 (citing Williams v. Geier, 671 A.2d 1368, 1376 (Del. 1996)).

14 *Id.* at 806-07.

15 *See* MM Cos., Inc. v. Liquid Auto, Inc., 813 A.2d 1118, 1131 (Del. 2003) ("This case presents a paragon of when the compelling justification standard of *Blasius* must be applied within *Unocal*'s requirement that any defensive measure be proportionate or reasonable to the threat posed."); *Chesapeake*, 771 A.2d at 323 ("If *Unocal* is applied by the court with a gimlet eye out for inequitably motivated electoral manipulations or for subjectively well-intentioned board action that has preclusive or coercive effects, the need for an additional standard of review is lessened.").

16 *Mercier*, 929 A.2d at 806-08.

17 *Id.* at 808-09; *Blasius*, 564 A.2d at 660 ("[A] decision by the board to act for the primary purpose of preventing the effectiveness of a shareholder vote inevitably involves the question who, as between the principal and the agent, has authority with respect to a matter of internal corporate governance. That is, of course, true in a very specific way in this case which deals with the question who should constitute the board of directors of the corporation, but it will be true in every instance in which an incumbent board seeks to thwart a shareholder majority.").

18 *Mercier*, 929 A.2d at 808.

19 *Id.* at 788 ("Recognizing, however, that the Supreme Court's recent decision in *Liquid Audio* continued to employ the 'compelling justification' language of *Blasius* within the context of an appropriate *Unocal* review of director conduct that affects a corporate election touching on corporate control, I also find that directors fearing that stockholders are about to make an unwise decision that poses the threat that the stockholders will irrevocably lose a unique opportunity to receive a premium for their shares have a compelling justification—the protection of their stockholders' financial best interests—for a short postponement in the merger voting process to allow more time for deliberation.") and 813 ("I will apply the analysis I have outlined, which I believe is consistent with the direction the teaching in *Liquid Audio* logically leads, but will also expressly set forth whether I conclude that the Inter-Tel Special Committee has demonstrated a compelling justification for its actions.").

20 Allen et al., *supra* note 6, at 1311. Allen et al. further suggest that the *Blasius* standard be eliminated altogether "as a 'stand-alone' review doctrine." *Id.* at 1312.

21 *Mercier*, 929 A.2d at 788, 810.

22 *Id.* at 811.

23 *Id.* at 811 (citing Paramount Commc'ns, Inc. v. Time Inc., 571 A.2d 1140, 1153 n.17 (Del. 1989)). The court in *Paramount* refers to "substantive coercion" as one category of threat that "some commentators" have suggested is

posed by a hostile offer. "Substantive coercion" is "the risk that shareholders will mistakenly accept an underpriced offer because they disbelieve management's representations of intrinsic value." *Paramount*, 571 A.2d at 1153 n.17 (citing Ronald J. Gilson & Reinier Kraakman, *Delaware's Intermediate Standard for Defensive Tactics: Is There Substance to Proportionality Review?*, 44 Bus. Law., 247, 267 (1989)).

24 *Mercier*, 929 A.2d at 813. The court noted that the special committee did not act "perfectly," however. *Id*. at 814. Elsewhere in the opinion, for example, Vice Chancellor Strine stated that he was "troubled" by the special committee's failure to disclose that, had the meeting occurred as scheduled, the merger would almost certainly have been rejected and that one factor in setting a new record date was that it would enable arbitrageurs to make additional purchases of Inter-Tel's stock that could be voted at the special meeting. *Id*. at 819-20.

25 *Id.* at 816.

26 *Id.* at 817.

27 *Id.* at 817-18.

28 *Id.* at 818-19. The court continued, "In the corporate context, compelling circumstances are presented when independent directors believe that: (1) stockholders are about to reject a third-party merger proposal that the independent directors believe is in their best interests; (2) information useful to the stockholders' decision-making process has not been considered adequately or not yet been publicly disclosed; and (3) if the stockholders vote no, the acquiror will walk away without making a higher bid and that the opportunity to receive the bid will be irretrievably lost."

29 940 A.2d 43 (Del.Ch. 2008).

30 *Id*. at 66 (citing Schreiber v. Carney, 447 A.2d 17, 25-26 (Del Ch. 1982)). Under *Schreiber*, if the plaintiff can show that the object or purpose of the action was to defraud or in some way disenfranchise other stockholders, the arrangement is illegal per se. Even if the arrangement is not found to have been motivated by fraudulent, disenfranchising, or otherwise inequitable intent, the arrangement is a voidable transaction subject to a test for intrinsic fairness.

31 *Cryo-Cell*, 940 A.2d at 70 n.167.

32 *Id.* at 76.

33 *Id.* at 77 (citing *Mercier*, 929 A.2d at 807).

34 *Id.* at 78 n. 188 (citations omitted):

> In a recent decision, this court held that directors had not breached their fiduciary duty by postponing a vote on an arms-length merger before a meeting was convened. The directors believed that they would lose the vote if it was held that day but had reason to believe that stockholder sentiment was changing, especially in view of changes in the economy's credit markets. A delay ensued during which it was clear to both sides that they needed to continue to press their case, pro and con the merger. As noted in that decision, when directors advocate an affirmative vote on a transaction, they are supposed to do so because they believe in good faith that the transaction will benefit the stockholders. That context is importantly distinct from an actual election of directors, in which the insiders delay because they believe the stockholders are making a mistake in choosing new leadership. In the former case, directors who face no risk of removal are asking for more time to make their case that a non-self-dealing transaction should receive approval. In the latter case, the directors are trying to insulate themselves from ouster, by forcing the insurgents to continue the fight beyond when the election was supposed to be held.

35 *Id.* at 82.

36 *Mercier*, 929 A.2d at 787:

> I conclude that well-motivated, independent directors may reschedule an imminent special meeting at which the stockholders are to consider an all cash, all shares offer from a third-party acquiror when the directors: (1) believe that the merger is in the best interests of the stockholders; (2) know that if the meeting proceeds the stockholders will vote down the merger; (3) reasonably fear that in the wake of the merger's rejection, the acquiror will walk away from the deal and the corporation's stock price will plummet; (4) want more time to communicate with and provide information to the stockholders before the stockholders vote on the merger and risk the irrevocable loss of the pending offer; and (5) reschedule the meeting within a reasonable time period and do not preclude or coerce the stockholders from freely deciding to reject the merger.

37 *Chesapeake*, 771 A.2d at 317.

Criminal Law and Procedure

From *Apprendi* to *Booker* to *Gall* and *Kimbrough*: The Supreme Court Blunders its Way Back to Luck-of-the-Draw Sentencing

*By William G. Otis**

A quarter-century ago, bipartisan majorities in Congress had come to understand that the federal sentencing system was, in today's parlance, "broken." Sentencing was rife with irrational disparity, principally because each judge could sentence as he saw fit—through the prism of his own temperament, experience, or even mood. Judges did not have to follow any uniform sentencing standards, or even proceed under any established theory as to what sentencing was supposed to accomplish. Appellate review of sentencing was virtually non-existent.

To fix the problem, Congress adopted the Sentencing Reform Act of 1984. The Act created a system of mandatory sentencing guidelines, developed largely from existing sentencing patterns, and appellate review to enforce them.

Although the guidelines were often criticized as "sentencing by the numbers," those very numbers succeeded in making the system more transparent, predictable, and accountable than the scattershot, subjective, and sphinx-like "system" they replaced. Mandatory guidelines also succeeded by the most important measure that can be applied to sentencing—to wit, they accompanied, even if they cannot be said exclusively to have produced, a consistent and long-term decrease in the crime rate. The deterrent and incapacitating effects of serious prison time that even a sympathetic judge would, under mandatory guidelines, find it difficult to avoid, did indeed, so it appeared, have their effect.

But sentencing reform carved from the hide of unfettered judicial power was not to last. In a series of opinions starting with *Apprendi v. New Jersey*[1] and ending in *Gall v. United States*[22] and *Kimbrough v. United States*,[3] the Supreme Court killed off determinate sentencing. These decisions rendered the guidelines "advisory only,"[4] and made clear that appellate review of district court sentencing decisions was to be deferential, if not, for practical purposes, empty.

The brief and promising life of determinate sentencing had come to an end. Luck-of-the-draw sentencing was back. So, too, is the invitation for "rehabilitation"-based, defendant-friendly sentencing, all dressed in the soothing if not particularly law-oriented terminology of "judicial discretion." Criminals facing jail—and especially those who speak for them—are likely to welcome this development, knowing from years of experience that "judicial discretion" is code for "lower than guidelines sentences." Whether the rest of us should be equally welcoming it is a different matter.

....

**William G. Otis was a charter member of the Attorney General's Advisory Sub-committee on the Sentencing Guidelines. He was a federal prosecutor in the Eastern District of Virginia under administrations of both parties, and most recently served as Counselor to the head of the Drug Enforcement Administration.*

In The Beginning...

The 1983 Senate Report accompanying the Sentencing Reform Act aptly stated the problem. It observed:[5]

> In the federal system today, criminal sentencing is based largely on an outmoded rehabilitation model. The judge is supposed to set the maximum term of imprisonment and the parole commission is to determine when to release the prisoner because he is 'rehabilitated.' Yet almost everyone involved in the criminal justice system now doubts that rehabilitation can be induced reliably in a prison setting.... Since the sentencing laws have not been revised to take this into account, each judge is left to apply his own notions of the purposes of sentencing. As a result, every day federal judges mete out an unjustifiably wide range of sentences to offenders with similar histories, convicted of similar crimes, committed under similar circumstances. One offender may receive a sentence of probation, while another—convicted of the very same crime and possessing a comparable criminal history—may be sentenced to a lengthy term of imprisonment. Even two such offenders who are sentenced to terms of imprisonment for similar offenses may receive widely differing prison release dates; one may be sentenced to a relatively short term and be released after serving most of the sentence, while the other may be sentenced to a relatively long term but be denied parole indefinitely.

Congress noted that both the prevalence and the degree of unwarranted disparity—meaning disparity unrelated to relevant offense or offender characteristics—was little short of scandalous:[6]

> [Disparity] occurs in sentences handed down by judges in the same district and by judges from different districts and circuits in the federal system. One judge may impose a relatively long prison term to rehabilitate or incapacitate the offender. Another judge, under similar circumstances, may sentence the defendant to a shorter prison term simply to punish him, or the judge may opt for the imposition of a term of probation in order to rehabilitate him. For example, in 1974, the average federal sentence for bank robbery was eleven years, but in the Northern District of Illinois it was only [half that]... Further probative evidence may be derived from [a] 1974 study in which fifty federal district court judges from the Second Circuit were given twenty identical files drawn from actual cases and were asked to indicate what sentence they would impose on each defendant. The variations in the judges' proposed sentences in each case were astounding.

To remedy the problem of luck-of-the-draw disparity, Congress embraced an entirely new concept: sentencing was henceforth to be governed by the rule of law.

Congress thus established the Sentencing Commission to draw up mandatory sentencing guidelines. Judges, while still having considerable discretion to tailor sentences to the individual circumstances of each case—and, in truly exceptional cases, to sentence outside the guidelines entirely—would ordinarily be required to sentence within the guidelines range.

Despite the fact that three of the seven voting members of the Sentencing Commission were, under the Act, to be federal judges, not all their colleagues were enthusiastic about the reining-in of what had been virtually unfettered sentencing authority. Some went so far as to find the guidelines unconstitutional, an intrusion on the separation of powers. (The Supreme Court, in an 8-1 decision, would later reject every significant separation of powers objection.[7])

Congress was well aware that a considerable portion of the federal judiciary, not to mention the criminal defense bar, believed that, if guidelines were inevitable, at least they should be voluntary rather than mandatory. Guidelines opponents noted that voluntary systems had been adopted by several states. Congress addressed the question explicitly and concluded that only a mandatory system could work. Voluntary or "advisory" guidelines simply could not be counted on to establish the overall uniformity, transparency, and accountability that had been so sorely lacking, and merely "suggested" sentences could scarcely be a basis for appellate enforcement. The Senate Report noted, for example:[8]

> The Committee rejected an amendment by Senator Mathias which would have expanded significantly the circumstances under which judges could depart from the sentencing guidelines in a particular case. The Mathias amendment would have permitted deviations from the guidelines whenever a judge determined that the characteristics of the offender or the circumstances of the offense warranted deviation, whether or not the Sentencing Commission had considered such offense and offender characteristics in the development of the sentencing guidelines.
>
> The Committee resisted this attempt to make the sentencing guidelines more voluntary than mandatory, because of the poor record of states [noted] in the National Academy of Science report which have experimented with 'voluntary 'guidelines. In his testimony before the committee on the comprehensive crime control act of 1983 (s. 829), [one] district attorney... noted that the voluntary guidelines in Massachusetts were completely ineffective in reducing sentencing disparities and imposing a rational order on criminal sentencing in the state, because judges generally did not follow them.

Mandatory federal sentencing guidelines became effective on November 1, 1987. It took a few years for the sentences they required to begin to take hold. Once fully in place, the guidelines (along with statutory minimum sentencing) did indeed produce, as critics pointed out, a significant increase in the prison population. What the critics mentioned less frequently was that, with criminals incarcerated instead of out on the street, there was a concomitant significant decrease in the crime rate. This was true for both violent and property crime. Between 1991 and 2005, the property crime rate dropped by more than half, from roughly 354 victimizations per 1000 households to 154.[9] Violent crime saw a similar trend, dropping almost every year from roughly forty-nine victimizations per household in 1993 to twenty-one in 2005—a decrease of close to 60%.[10] While no serious person maintains that mandatory federal sentencing guidelines deserve all the credit for this startling improvement in the crime picture, no one can plausibly deny that they played a significant role.

Success Proves Too Much To Abide

The destruction of determinate sentencing started quietly enough, with *Apprendi v. New Jersey*.[11] There, the defendant fired several shots into the home of a black family that had recently moved in nearby. In a statement to the police shortly afterwards (later retracted), Apprendi admitted that he committed the crime because the victim's family was African-American and he "did not want them in the neighborhood." He was promptly charged in a twenty-three-count indictment. Nothing in the indictment referred to New Jersey's hate crimes statute, however, and there was no count alleging that Apprendi acted with a racial purpose.

Apprendi entered an agreement in which he pleaded guilty to three counts and the state dismissed the others. In doing so, the state reserved the right to request that the court impose an "enhanced sentence" on one of the counts (Count 18) charging possession of a firearm for an unlawful purpose—a count which by its terms carried a maximum sentence of no more than ten years. Apprendi reserved the right to challenge any unindicted "hate crimes enhancement" on constitutional grounds.

The court accepted the plea agreement, and the prosecutor moved for an enhanced sentence exceeding ten years under the uncharged hate crimes statute. The court convened a hearing on the question of Apprendi's purpose in possessing and firing the gun at the victim's house. The sentencing judge concluded, by a preponderance of the evidence, that Apprendi's behavior was motivated by racial bias, and sentenced him to twelve years' imprisonment on Count 18.

A divided New Jersey Supreme Court rejected Apprendi's argument that the two-year enhancement violated his right under the Due Process Clause to a jury determination, beyond a reasonable doubt, of the facts upon which it was based. The U.S. Supreme Court reversed in an opinion by Justice Stevens, joined by Justices Scalia, Souter, Thomas, and Ginsburg. The Court held that any fact that increases the penalty for a crime beyond the prescribed statutory maximum, other than the fact of a prior conviction, must be admitted by the defendant or submitted to a jury and proved beyond a reasonable doubt.

One may take as well-reasoned the Court's holding, subscribed as it was by the Court's most liberal and conservative members, while still noting that it was broader than needed to vindicate the principle at its base. The Court could have held simply that a defendant cannot be sentenced under the provisions of a statute he was never indicted for violating. (Indeed, Justice Thomas said almost exactly that in his concurring opinion, quoting the long-honored rule that "[t]he indictment must allege whatever is in law essential to the punishment sought to be inflicted."[12]) By casting its holding less precisely in terms of what is allowed under the "statutory maximum," the *Apprendi* majority paved the way for a critical breach in the guidelines.

The breach opened four years later, in *Blakely v. Washington*.[13] In that case, the defendant, "evidently a difficult man to live with,"[14] bound his wife in construction tape and abducted her at knifepoint from their home in Grant County, Washington. He drove her to Montana, where he was arrested.

Blakely was charged by the State of Washington with first degree kidnapping. Pursuant to a plea agreement, that count was replaced with a charge of second degree kidnapping, a Class B felony which carries a statutory maximum of ten years imprisonment. At the same time, the Washington Legislature had adopted a guidelines system roughly similar to that of the United States. Under Washington's guidelines, the "standard range" for second-degree kidnapping was forty-nine to fifty-three months. The Washington Sentencing Reform Act provided that a standard range sentence may be enhanced if the sentencing judge finds "substantial and compelling reasons" for doing so.[15]

The prosecutor recommended a sentence within the standard range, but the court did not agree. After a hearing in which the unpleasant details of the kidnapping were adduced, the court concluded that Blakely acted with "deliberate cruelty," a statutorily enumerated ground for an exceptional sentence (or upward departure, as it would be called in federal law). The court thus imposed a ninety-month sentence, thirty-seven months above the maximum of the standard range—but thirty months below the general statutory maximum for a Washington Class B felony.

Blakely made unsuccessful appeals to the state, but met with better luck in the U.S. Supreme Court. With Justice Scalia writing for the same five-justice majority that decided *Apprendi*, the Court concluded that Washington had violated Blakely's right to have a jury determine every fact upon which his sentence was based.

Although little noticed in the aftermath of the *Blakely* opinion, it hinged—as one would expect, in light of the way *Apprendi* had cast its holding—on the Court's interpretation of what the term "statutory maximum" means. The Court thought it was the Washington *guidelines* maximum range—that having been embraced, after all, by a statute, to wit, the Washington Sentencing Reform Act. Washington, by contrast, argued that the statutory maximum was the maximum designated by the statute Blakely pleaded guilty to violating—i.e., the Class B felony statute of second degree kidnapping, which carries a maximum of ten years.

The Court said that the "'statutory maximum' for *Apprendi* purposes is the maximum a judge may impose solely on the basis of the facts reflected in the jury verdict or admitted by the defendant [in his guilty plea]."[16] Thus, the *Blakely* Court continued, the statutory maximum sentence "is no more 10 years here than it was 20 years in *Apprendi*..." What the Court overlooked, however, was that while Apprendi had never been charged with a hate crime, with its twenty year maximum, Blakely *had* been charged with, and had admitted in the plea proceedings, the Class B felony kidnapping of which he was convicted, with its ten year maximum.

What had been slightly imprecise language in *Apprendi* was thus transformed into the time bomb that would detonate under federal sentencing law. Once "statutory maximum" was understood to mean not the maximum designated by statute defining the offense of conviction but the "standard guidelines maximum," the end was in sight. Justice O'Connor's *Blakely* dissent saw the handwriting on the wall: "What I have feared most has now come to pass: Over 20 years of sentencing reform are all but lost, and tens of thousands of criminal judgments are in jeopardy."[17]

Ironically, seven months later, Justice O'Connor and her fellow dissenters would cast the decisive votes to seal the fate she correctly foresaw.

Booker and its Remedy

In *United States v. Booker*, the five-justice *Apprendi* and *Blakely* majorities held the day and declared the federal sentencing guidelines unconstitutional as inconsistent with the Sixth Amendment.[18] While the most serious—indeed the terminal—damage done by *Booker* lay in the remedy it commanded, it is worth a moment's pause to observe that the analytical underpinning of *Booker*'s constitutional holding was even more dubious than the *Blakely* analysis upon which it purported to rest.

In *Blakely*, the Court arguably had at least some basis for believing that the "statutory maximum" which could not be exceeded without a jury determination beyond a reasonable doubt was the standard *guidelines* maximum. That was because in Washington State the legislature's Sentencing Reform Act *itself* specified a standard guidelines sentencing range of forty-nine to fifty-three months for the second degree kidnapping of which Blakely had been convicted. There was, accordingly, at least an argument for regarding fifty-three months as the "statutory" maximum, notwithstanding the longer ten-year maximum provided under the general Class B felony statute Blakely had been convicted of violating. But there is no corresponding statutory designation of a particular sentencing range under federal law: the Sentencing Reform Act of 1984, unlike its counterpart in Washington, does not (and did not) specify particular numerical sentencing ranges. Still less does it specify a particular sentencing range in any given case or class of cases different from the maximum provided by statute defining the offense of conviction. Thus, whatever justification existed in *Blakely* for defining the standard *guidelines* maximum as the *statutory* maximum was significantly weakened, if not absent, in *Booker*.

The true death knell for the federal guidelines lay in *Booker*'s remedy, though. Two remedies were on the table. One was to require the government to prove beyond a reasonable doubt to a jury (or to have the defendant admit in the plea agreement) all the facts the government would seek to have the court consider in determining the sentencing range. The remedial majority in *Booker* (Justice Breyer, with Chief Justice Rehnquist and Justices O'Connor, Kennedy and Ginsburg) rejected that option, largely, it said, because it could not be counted upon to carry forward the "real offense" sentencing regime Congress wanted to establish with the SRA.

Instead, the remedial majority decided that Congress would have preferred to continue "real offense" sentencing via a voluntary or "advisory" system of guidelines. The Court created that system by excising two provisions of the SRA—the provision requiring judges to sentence within the guidelines absent exceptional circumstances, and the provision for *de novo* review in the courts of appeals. It is only a modest oversimplification to say that the new, voluntary regime amounted to "apply-them-when-you-think-best" guidelines, with light-handed appellate review for understandably undefined "reasonableness."

As Justice Scalia noted in partial dissent, "[t]his is rather like deleting the ingredients portion of a recipe and telling the cook to proceed with the preparation portion."[19] It was Justice Stevens, however, who most meticulously exposed the remedial majority's error. Justice Stevens's dissent on that point is worth reading in its entirety, but a few passages give the flavor:[20]

> In order to justify excising [the mandatory and de novo review portions of the SRA], the Court has the burden of showing that Congress would have preferred the remaining system of discretionary Sentencing Guidelines to not just the remedy I would favor, but also to any available alternative, including the alternative of total invalidation, which would give Congress a clean slate on which to write an entirely new law. The Court cannot meet this burden *because Congress has already considered and overwhelmingly rejected the system it enacts today. In doing so, Congress revealed both an unmistakable preference for the certainty of a binding regime and a deep suspicion of judges' ability to reduce disparities in federal sentencing.* A brief examination of the SRA's history reveals the gross impropriety of the remedy the Court has selected.
>
> *****
>
> The text of the law that actually passed Congress… should be more than sufficient to demonstrate Congress' unmistakable commitment to a binding Guidelines system. That text requires the sentencing judge to impose the sentence dictated by the Guidelines ("the court shall impose a sentence of the kind, and within the range" provided in the Guidelines unless there is a circumstance "not adequately taken into consideration by the" Guidelines), and [the de novo appeal provision gives] teeth [to the mandatory provision] by instructing judges that any sentence outside of the Guidelines range without adequate explanation will be overturned on appeal. Congress' chosen regime was carefully designed to produce uniform compliance with the Guidelines. Congress surely would not have taken the pains to create such a regime had it found the Court's system of discretionary guidelines acceptable in any way.

The End of Determinate Sentencing

There was momentary hope that the Supreme Court's creation of advisory guidelines might not lead to a wholesale return to luck-of-the-draw sentencing. In *Rita v. United States*, the Court held that a court of appeals, although not required to do so, may apply a presumption of reasonableness to a sentence within the guidelines (while being at pains to note that it was by no means implying that a sentence *outside* the guidelines could be presumed *unreasonable*).[21]

Rita proved a tepid and fleeting gesture. Less than six months later, in *Gall* and *Kimbrough*, the Court made clear how completely the guidelines had been swept away.

1. In *Kimbrough*, the defendant had been convicted of selling crack cocaine. His guidelines sentencing range was 228 to 270 months. The district court, viewing that sentence as more than necessary and, in particular, as a reflection of little more than an overwrought concern with the dangers of crack cocaine, as opposed to the powdered form of the drug, sentenced Kimbrough to 180 months' imprisonment, four years less than the minimum of the range. The Fourth Circuit reversed, holding that a sentence outside the guidelines range is per se unreasonable when it is based simply on the district judge's disagreement with the disparate treatment of crack and powder cocaine.

2. In *Gall*, the defendant, while a college student, spent seven months in a conspiracy selling ecstasy. He sold roughly 10,000 ecstasy pills, netting himself more than $30,000. Prudently apprehensive that one of his co-conspirators talked too loosely, Gall withdrew from the conspiracy, graduated from college, began a productive life and—so the Court stated—became drug-free. More than three years later, he was indicted for his role in the conspiracy. He self-surrendered and, while free on his own recognizance, started a successful business.

Gall entered a guilty plea admitting that he was responsible for distributing at least 2,500 grams of ecstasy. The government, for its part, agreed inter alia that recent changes in the guidelines that enhanced the punishment for ecstasy would not apply.

Gall's sentencing range was thirty to thirty-seven months' imprisonment. The district court imposed no prison time, however, and sentenced Gall to thirty-six months' probation. The Eighth Circuit reversed, holding that the extent of a departure from the guidelines—in this case considerable—must be "proportional" to the reasons justifying it. The court of appeals thought that the reasons offered by the district court came up short, and remanded for re-sentencing.

3. The Supreme Court reversed in both cases, each time by a vote of seven to two. In *Kimbrough*, the majority, speaking through Justice Ginsburg, held that the guidelines for crack cocaine, like all others after *Booker*, are "advisory only," and that advisory guidelines sentences may be overturned on appeal only for abuse of discretion. While the majority opinion discussed at length the supposed residual importance of a sentencing court's careful and respectful consideration of the guidelines, its language was precatory, and the district court was applauded for its invocation of what was called the SRA's "overarching instruction" to impose a sentence "sufficient, but not greater than necessary," to accomplish Congress's stated sentencing goals.[22]

The majority, like the district court did not define how a sentence of 180 months is determined to be "necessary" (but a sentence of 181 months presumably "unnecessary"). Likewise, the majority made no mention of the first three specific factors listed after the "overarching" principle of the SRA, those being the need for the sentence imposed (1) to reflect the seriousness of the offense, to promote respect for the law, and provide just punishment; (2) to afford adequate deterrence to criminal conduct; and (3) to protect the public from further crimes of the defendant.[23]

Gall was of a piece with *Kimbrough*. The Court, per Justice Stevens, held that the Eighth Circuit erred in requiring the district court to identify "extraordinary" circumstances in order to justify its departure. The majority believed that such a requirement would come too close to creating a presumption of unreasonableness for sentences outside the guidelines, in derogation of *Booker*'s rule that the guidelines are no more than advisory. The majority also criticized the court of appeals for adopting a wooden "mathematical approach" to departure analysis, even though the existence of such an approach in the language of the Eighth Circuit's opinion is difficult to discover. The Court acknowledged that the extent of a departure is relevant to an appellate court's analysis of reasonableness, but emphasized that the reviewing court must give "due deference" to the district court's assessment of the myriad of factors that properly may inform a sentencing decision. In the case before it, the majority scolded the Eighth Circuit for having failed to give sufficient deference to the district court's "reasoned and reasonable" analysis.[24]

4. One need not search through post-*Gall* and post-*Kimbrough* cases in order to understand what is left of determinate sentencing or of the Sentencing Reform Act's central goal of reducing idiosyncratic disparity. *Gall* and *Kimbrough* themselves show all that is needed.

In *Kimbrough*, the Court acknowledged that "uniformity remains an important goal of sentencing"[25]—but not so important that it could not be set aside two sentences later with the observation that "our opinion in *Booker* recognized that some departures from uniformity were a necessary cost of the remedy we adopted."

The phrase, "some departures from uniformity" is a modest assessment of the outright chaos in crack sentencing that is certain to follow in *Kimbrough*'s wake. Some district judges will continue to see crack as the menace it has proven to be and will follow the guidelines. Others will see the guidelines as still in the thrall of "hysteria" about the crack wars of the 1980s and allow breathtaking departures. Under *Rita*, the former cannot be constrained, and under *Kimbrough* neither can the latter. The upshot is less likely to be "some departures from uniformity" than helter-skelter sentencing in wholesale lots, in a major area of federal criminal law—all justified by nothing more than hundreds of individual district judges formulating their own widely divergent versions of the dangers of crack, one chambers at a time. If the Sentencing Reform Act was adopted to put an end to anything, that was it.

The upshot of *Gall* is potentially even more troubling. Most reasonable people would probably agree that there were exceptional circumstances in that case justifying a significant downward departure. But to depart to *no prison time whatever* for a defendant who considerably enriched himself by selling thousands of pills of a dangerous drug, and to depart to that extent when the guidelines called for a sentencing range exceeding three years' incarceration—to take that path, and have a majority of the Supreme Court embrace it as a "reasoned and reasonable" result, sends an unmistakable message. That message was spelled out in the dissent by Justice Alito, who said that the interpretation given *Booker* by the *Gall* majority "means that district judges, after giving the Guidelines a polite nod, may then proceed as if the Sentencing Reform Act had never been enacted."

What We Have Now

Beneath the successful attack on the centerpiece of the Sentencing Reform Act—mandatory guidelines with meaningful appellate enforcement—there has always been an agenda, namely, lighter sentences for criminals. The organized defense bar knew from experience that the way to get to lighter sentencing was to replace mandatory guidelines with the previous regime of "judicial discretion." It knew in particular that when "discretion" is exercised, it is virtually always in only one direction—in the convicted defendant's favor.

The Sentencing Commission's statistics show how right the defense bar was. The single most telling indicator of the imbalance in "judicial discretion" is the incidence and direction of departures allowed. From the time guideline sentencing began up to the present day, the incidence of downward departures has dwarfed the incidence of upward departures by roughly twenty-five to one. It is true that part of this is due to government-sponsored downward departures to reward a defendant's assistance (typically information or testimony about co-conspirators). Also in the mix are a smaller number of government-sponsored downward departures resulting from the "fast track" program for illegal entrants into the United States, mostly in border districts. But even discounting those categories, the number of downward departures vastly outstrips the number of upward departures. Overall, a defendant facing sentencing today has a negligible 1.5% chance of receiving a sentence above the guidelines and a 38% chance of receiving one below. It has come to the point that, in the lexicon of those who deal regularly with sentencing issues, the phrase "downward departure" is regarded as a redundancy.

It thus turns out that "judicial discretion" in the context of the debate about sentencing is a very misleading phrase. If there were anything approaching the even-handedness implied by the phrase, there would be at least roughly equal numbers of upward and downward departures. But that has never been the case. "Judicial discretion" in this area is not discretion at all as commonly understood. It is a one-way street to lower sentences. Indeed, whatever else may be said of them, *Gall* and *Kimbrough* are apt representatives of future sentencing outcomes. In a nutshell, the principal real-world effect of the end of determinate sentencing will be thousands of criminals back on the street before they otherwise would have been.

It would be troubling enough, and dangerous, if that were the end of it, since it is impossible to believe that putting criminals back on the street will have no effect on crime. And it is dishonest to conduct the sentencing debate without acknowledging this fact. But, even with all that, there is yet a greater cost in the end of determinate sentencing, and that is its cost to the rule of law.

Like every other statute, the Sentencing Reform Act was not perfect, and neither were the guidelines it brought into being. But it was a serious and mostly successful effort to bring defined standards into an enormously important area of criminal practice previously left to chance. If the law of evidence had been as arbitrary as the law of sentencing was in the pre-SRA era, it would have been a national scandal. The end of the rule of law in federal sentencing and the return of what *Gall* and

Kimbrough tell us will be effectively unfettered discretion is also a scandal, but the outrage is nowhere to be seen.

If the Court were bent on eviscerating the SRA to this extent, the better approach would have been, as Justice Souter has suggested,[26] to overturn the Act in its entirety and allow Congress to start over, with the mandatory guidelines it knew were the only hope for consistency, together with the jury determination of sentencing facts that *Apprendi* demands.

Instead, we now have something worse, and less honest, than the pre-SRA regime of standardless sentencing. We have standardless sentencing pretending to have standards. The shrewdly opaque message to the public is that we still have sentencing guidelines, only that they are more "flexible" than before. Sentencing Commissioners continue to draw hefty salaries to write guidelines (that can be ignored at will). Probation officers continue to calculate ranges on worksheets (that may count for something or may not). District judges go through the window dressing rehearsed for them in *Gall* and *Kimbrough* (assured by those decisions that if the litany is elaborate enough, it need not be given any weight). A person employing impolite language might call this a charade.

Because the hollowed-out guidelines are still twitching in the land of the un-dead, further depredations to the rule of law, and the proper role of the judicial branch, are sure to follow. Justice Thomas made the point in his dissenting opinion in *Kimbrough*:

> As a result of the [*Booker*] Court's remedial approach, we are now called upon to decide a multiplicity of questions that have no discernibly legal answers….
>
> The outcome [today and those in *Rita* and *Gall*] may be perfectly reasonable as a matter of policy, but they have no basis in law. Congress did not mandate a reasonableness standard of appellate review—that was a standard the remedial majority in *Booker* fashioned out of whole cloth. See 543 U. S., at 307–312 (Scalia, J., dissenting in part). The Court must now give content to that standard, but in so doing it does not and cannot rely on any statutory language or congressional intent. We are asked here to determine whether, under the new advisory Guidelines regime, district courts may impose sentences based in part on their disagreement with a… policy judgment reflected in the Guidelines. But the Court's answer to that question necessarily derives from something other than the statutory language or congressional intent because Congress, by making the Guidelines mandatory, quite clearly intended to bind district courts to the Sentencing Commission's categorical policy judgments. See 18 U. S. C. §3553(b) (2000 ed. and Supp. V) (excised by *Booker*). By rejecting this statutory approach, the *Booker* remedial majority has left the Court with no law to apply and forced it to assume the legislative role of devising a new sentencing scheme.

The road from *Apprendi* to *Booker* to *Gall* and *Kimbrough* is strewn with damage that has been all but ignored—damage to future public safety, to uniformity and honesty in sentencing, and to the proper authority of Congress. In the 1980s, there was a bipartisan consensus strong enough to make federal sentencing conform for the first time to the rule of law. Whether such a consensus exits today is an open question. But the first step toward building one is to understand, as it was understood twenty-five years ago, how urgently it is needed.

Endnotes

1 530 U.S. 466 (2000).

2 128 S. Ct. 586 (2007).

3 128 S. Ct. 558 (2007).

4 *Id.* at 465.

5 S. Rep. 98-225 (1983) at 33.

6 *Id*. at 35 (footnotes omitted).

7 Mistretta v. United States, 488 U.S. 361 (1989).

8 *Id*. at 64.

9 Bureau of Justice Statistics, National Crime Victimization Survey, Property Crime Trends, 1973-2005.

10 *Id*.

11 *Supra*, note 1.

12 1 J. Bishop, Law of Criminal Procedure 50 (2d ed. 1872))

13 542 U.S. 296 (2004).

14 *Id*. at 298.

15 Wash. Rev. Code Ann. § 9A.40.020(1)(2000).

16 *Blakely*, 542 U.S. at 303 (italics omitted).

17 *Id*. at 326 (O'Connor, J., dissenting).

18 543 U.S. 220 (2005).

19 *Id.* at 307 (Scalia, J., dissenting).

20 *Id*. at 292, 294-95 (Stevens, J., dissenting in part) (emphasis added, footnotes deleted).

21 127 S. Ct. 2456 (2007).

22 *Kimbrough*, slip op. at 22, citing 18 USC 3553(a).

23 18 USC 3553(a)(2)(A),(B) and (C).

24 Slip op. at 21.

25 Slip op. at 19.

26 *See Gall*, slip op. at ___ (Souter, J., concurring).

Forensic Science Needs Checks and Balances

*By Roger G. Koppl & Radley Balko**

Justice is the foundation of liberty. Thus, the proper functioning of our criminal justice system is a vital concern for all who value liberty. When figures such as David Hume first laid down the classical liberal principles which form the foundation of our criminal justice system, however, scientific evidence did not yet have the important role in criminal cases that it does today. Since the beginning of the twentieth century, that importance has grown so much that forensics is now a central function of the criminal justice system in the United States and in courts around the world.

But our system of checks and balances has not yet been updated to factor in forensic science—in theory or in practice. As a result, crime labs in most jurisdictions have a virtual monopoly over evidence analysis, and this monopoly structure has created needless and unacceptably high error rates. Today, a jury's verdict often turns on forensic evidence alone. This state of affairs, thus, calls into question whether our criminal justice system, as it stands today, can truly be called "adversarial." In many ways, forensic scientists may be more important to the outcome of a case than defense lawyers and prosecutors—which makes the profession's errors all the more alarming. In this article, we document the poor performance of forensic science and propose a system of checks and balances to fix the broken system.

Errors in Forensic Science

Despite the impression one might glean from popular culture, forensic science and medico-legal investigation are far from error-free. Persistent errors have been documented in a variety of forensics specialties, including forensic pathology, fire investigation, bite mark analysis, fingerprint analysis, and DNA typing. A few examples:

In February of this year, Mississippi exonerated two men convicted of two similar murders just a few miles apart. Levon Brooks was convicted of raping and murdering his girlfriend's three-year-old daughter in 1990. Two years later, Kennedy Brewer would be convicted of a remarkably similar crime, the rape and murder of his girlfriend's three-year-old daughter. Brooks was sentenced to life without parole. Brewer was sentenced to death. Both men were convicted almost exclusively on the testimony of Dr. Steven Hayne and Dr. Michael West (Mitchell).

Hayne is a forensic pathologist who has essentially monopolized Mississippi's autopsy business for twenty years. He has testified to performing between 1,200 and 1,800 autopsies a year, an astonishing figure given that the National Association of Medical Examiners recommends an individual doctor do no more than 325. Dr. West is a "forensic odentologist," or bite-mark analyst, who once claimed that he could trace the bite marks in a half-eaten sandwich at a murder scene back to the defendant. The two have long been criticized by Mississippi defense attorneys and medical malpractice attorneys for jiggering their conclusions to support the theories of prosecutors and plaintiffs' attorneys (Balko, "CSI: Mississippi").

In the Brooks and Brewer cases, Hayne performed the initial autopsy, then called in his longtime collaborator West to do "bite mark analysis." In both cases, West said that marks others would call "indiscriminate scratches and bruises" were really human bite marks. In both cases, West said that he could definitively trace the bite marks back to the defendants. In both cases, the jury believed him, and voted to convict (Mitchell).

In February of this year, officials in Mississippi announced that they had arrested a man named Justin Albert Johnson who confessed to both murders. A DNA match confirmed the confession. Brooks had served eighteen years in prison. Brewer had served fifteen—all of them on Death Row.

Cameron Todd Willingham was executed in Texas 2004. Charged with murdering his three small children by arson, he was convicted with forensic techniques that were current at the time of the fire in 1991, but had been discredited by the time of his execution in 2004 (Mills & Possley, 2004). A "key reference text for the Texas fire marshal's office" (Mills & Possley, 2004), found that many then-standard techniques of arson investigation have since been shown to be inaccurate. The report by the National Fire Protection Association was published on February 10, 1992, within two months of the Willingham fire. Some of these bogus techniques were used against Willingham. The presence of "crazed glass," for example, was thought to indicate that an accelerant had been used. It has since been shown that these intricately patterned cracks can also be caused by dousing hot glass with water, which obviously occurs frequently as firefighters attempt to put out fires (Mills & Possley 2004). The *Chicago Tribune* reports that, "Before Willingham died by lethal injection on Feb. 17, Texas judges and Gov. Rick Perry turned aside a report from a prominent fire scientist questioning the conviction." It would be impossible to say whether or not Willingham was guilty, but it is clear that he was convicted on bad science, and his execution was a perversion of justice.

In Houston, Texas problems forced the city's lab to shut down DNA testing from December 2002 to July 2006, during which time police used a private lab instead (Bromwich 2005, Khanna 2006, Glenn 2006). Before the shutdown, Josiah Sutton, was convicted of rape largely on the lab's DNA evidence, which was later shown to be inaccurate. Imprisoned at the age of sixteen, Sutton served four years before he was released (Koppl 2005). A subsequent audit of the Houston lab revealed many problems, including the risk of cross-contamination from the use of a common evidence screening area for trace, serology, and arson; failure to follow procedures for the calibration of equipment or maintain logs of repair and calibration of equipment; lack of procedure for preparing and preserving case notes; sloppy reports; and improperly and ambiguously labeled reagents (FBI Director, 2002). One particularly alarming line from the report stated, "The audit team was informed that on one occasion the roof leaked such that items of evidence came

** Roger G. Koppl is a Professor at Fairleigh Dickinson University and Director of the Institute for Forensic Science Administration. Radley Balko is a Senior Editor at Reason Magazine.*

in contact with the water" (FBI Director, 2002).

Poor DNA work can lead to false exonerations, as well. In 2001, a man in Pacific, Washington with a prior sex-crime conviction was arrested and charged with the rape of his ten-year old niece. DNA tests excluded him, however, and pointed to someone unknown. About two weeks after the original suspect accepted a deal from the prosecutors and pled guilty to the lesser charge of child molestation, the Washington State Patrol's Tacoma crime lab discovered that the original exculpatory result was tainted by cross-contamination (Teichroeb 2004). Presumably, a proper test would have included the suspect, although at this point, it is impossible to say for sure.

Brandon Mayfield is probably the most prominent American case of a false conviction, due to an incorrect fingerprint match. In 2004, the FBI arrested Mayfield as a material witness in the Madrid train bombing of March 2004. He had been identified as the source of a latent print found on a bag of detonators near the crime scene. After assigning three of its top fingerprint examiners to the case, the FBI declared a "100 percent match" to Mayfield. The Spanish National Police objected, however, and declared a match to man named Ouhnane Daoud. The Spanish authorities' suspicions were confirmed when the FBI withdrew its identification and released Mayfield (Office of the Inspector General 2006).

Florida's Seminole County provides a more recent example of erroneous fingerprint analysis. In March 2007, Tara Williamson, a fingerprint examiner for the Seminole County Sheriff's Office in Florida, wrote a memo accusing her co-worker Donna Birks of misbehavior and incompetence. Her accusations seem to have been correct. By June 2007, investigators from the Florida Department of Law Enforcement (FDLE) discovered six cases in which Birks made a positive identification from prints that should have been considered inconclusive, and a seventh case in which she identified someone who should have been excluded. (Such judgments, of course, assume that the FDLE fingerprint examiners have themselves made correct analyses.) Williamson's memo says that Birks "reported numerous identifications without verification," that she "had a trainee with three-weeks of experience verify latent print identifications," and that on one occasion she sought out a third, retired, examiner to verify an identification after two "examiners in the office were not able to verify the print"—or, in other words, disagreed with her analysis. Williamson reports that these actions violated "basic ethical guidelines" governing fingerprint examination (Williamson 2007). Birks had been promoted to latent print examiner in 1998. It is estimated that she worked on about 1,500 cases over the years. As of June 2007, the FDLE was re-examining 300 of those cases (Stutzman 2007a & 2007b, Williamson 2007).

In October 2007, a Maryland court ruled that the standard "ACE-V methodology" of fingerprint examination is not reliable enough for capital cases (*State of Maryland v. Bryan Rose*). The decision cited evidence that prominently included the Mayfield misidentification. The judge concluded "that ACE-V was the type of procedure" Maryland rules of evidence "intended to banish, that is, a subjective, untested, unverifiable identification procedure that purports to be infallible."

Studies outside the courtroom show similar cause for concern. A 1999 workshop conducted by the American Board of Forensic Odontology, for example, asked bite mark experts to match four bite marks with seven dental models. More than six in ten of participants came back with false positives (Bowers).

Fundamental Principles of Forensic Science Administration

The cases reviewed above are not "isolated incidents." One of the authors here has concluded from his reviews of the evidence on error rates in forensics (including proficiency tests and controlled studies) that "forensic analysis is not sufficiently reliable" (2005a). The advocacy group The Innocence Project reports that seventy-four of the 214 cases in which DNA testing has exonerated a wrongfully convicted defendant involved the introduction of faulty forensic evidence by prosecutors. Many scholars, journalists, activists, and others have also recognized the need to improve forensic science.

The three leading proposals for reform are probably independence, masking, and oversight.

Paul Giannelli is the leading figure in favor of independence. In an important article on forensics, he argues that crime labs "should be transferred from police control to the control of medical examiner [ME] offices" (1997). Admirably, Giannelli notes that, although his proposal "is a substantial step in the right direction," it "is not a panacea".

Risinger et al. (2002) call for "masking," whereby "domain-irrelevant information" would be hidden from forensic scientists. Risinger et al. appeal to a large empirical literature in psychology. The point may be best illustrated, however, by an important study by Dror & Charlton (2006), where the co-authors employed experienced fingerprint examiners to analyze evidence from cases they had decided in the past. The subjects did not know they were looking at their own, earlier cases. In half the cases, they replaced the original case information with information suggesting a conclusion opposite to the original judgment. In half, no such contextual information was supplied. The examiners of their study reversed themselves in six of forty-eight cases. Two of the six reversals were from the twenty-four cases in which no biasing information had been given.

Peter Neufeld and Barry Scheck founded The Innocence Project, which has, to date, participated in over 200 DNA exonerations of persons wrongly convicted. An important figure in any discussion of how to improve forensic science, Neufeld has argued that "[g]overnment oversight and the creation of independent academic centers to validate technologies and techniques, encourage best practices, and enforce appropriately cautious standards for the interpretation of data could dramatically enhance the reliability of forensic science and engender greater public confidence in the outcome" (2005). Neufeld's plea for more scientific research is proper, but beyond scope of this article. His call for oversight is representative of the "repeated calls" for "oversight" noted in a 2003 *Science* editorial (Kennedy 2003).

Students of public choice theory will recognize an important problem with "oversight:" *Quis custodiet ipsos custodes?* Who will guard the guardians themselves? Koppl

(2005) identifies eight remediable features of the current institutional structure of forensic science, each of which reduces reliability. Corresponding to each flaw in the institutional structure is a suggestion for amending current institutions. The proposed suit of reforms is "competitive self regulation." Table 1 summarizes the argument of Koppl (2005).

The key to the proposal is rivalrous redundancy. In the current system, once forensic evidence has been sent to one lab, it is unlikely that another lab will review the same evidence. In this sense, each lab has a monopoly on the analysis of evidence it receives. In pure science, no lab enjoys such a monopoly. Rather, the results of any one lab may be challenged by any other. In forensic science, however, this scenario is unlikely, and this radical difference in network structure may help explain the difference in reliability that seem to exist between the two fields.

Competitive self-regulation creates a salutary rivalry among crime labs. In part, it reduces error rates by making

Table 1: Proposals of Koppl (2005) in Tabular Form

Current System	Resulting Problem	Proposed Institutional Change	Explanation or Comment
Monopoly	Sloppy, biased, and sometimes fraudulent work	Rivalrous redundancy	There should be several competing forensic labs in any jurisdiction. Subject to the constraints of feasibility, some evidence should be chosen at random for duplicate testing at other labs. The same DNA evidence, for example, might be sent to more than one lab for analysis. The forensic worker need not know whether the evidence is examined by another lab. He will know that there could be another lab, and sometimes is.
Dependence	Bias	Independence	Crime labs should be independent of police and prosecutors.
Poor quality control	Persistently poor work	Statistical review	Statistical review would support improved quality control. For example, if a given lab produces an unusually large number of inconclusive findings, its procedures and practices should be examined.
Information sharing	Conscious and unconscious bias	Information hiding	Evidence should be prepared for testing so as to shield the lab doing a test from all extraneous knowledge of the case particulars.
No division of labor between forensic analysis and interpretation	Error from false interpretations of legitimate results	Division of labor between forensic analysis and interpretation	When this measure is combined with the provision of forensic counsel for the defense, errors of interpretation are less likely to go unchallenged.
Lack of forensic counsel	False convictions	Forensic counsel for the indigent	Forensic science decides many criminal cases and yet we do not have a right to forensic counsel similar to our right to legal counsel.
Lack of competition among forensic counselors	Poor quality forensic counsel	Forensic vouchers	A voucher system would give forensic counselors to the indigent an incentive to provide high-quality services to their clients.
Public ownership	Weak financial incentives to provide high-quality work	Privatization	Unlike public labs, private labs would be subject to meaningful fines and civil liability. In the US, the federalist structure of government means federal regulation and oversight are easier when labs are private.

fraud and corruption more difficult. It does so, however, by providing each actor in the system with an external epistemic check currently lacking.

The case of the FBI forensic scientist Jacqueline Blake is a fine illustration. For over two years, Blake systematically failed to run her negative controls when performing PCR/STR DNA analysis. The negative control tests the reagents and equipment used in such analyses, but without including a DNA sample.

It is difficult to guess Blake's true motives. The OIG report on her case, however, seems to suggest that she was a well meaning person who was simply not up to the job. "Some Laboratory employees have speculated that the reason that she failed to process the negative controls was because she lacked confidence in her ability to master PCR/STR testing" (OIG 2004). In this case, as with the Houston Crime Lab, we find a crime lab whose failures went undetected for years because there were no external epistemic controls of the sort we take for granted in pure science.

Events in the Seminole County fingerprint scandal reinforce the point. Recall that fingerprint examiner Tara Williamson wrote a memo accusing Birks of misconduct and incompetence. The investigations initiated in response to that memo revealed that she had made her own errors, including improper verifications of Birks's work (Stutzman 2007b). Williamson has been demoted to dispatcher and is no longer given fingerprint work. It seems clear that she was a well-motivated individual. But good intentions did not prevent her from making errors. Error reduction requires that each forensic lab be subject to an external check, without which even highly motivated actors may unwittingly commit repeated errors.

Redundancy is Cost-Reducing

Redundant testing would seem to be a costly suggestion. Where we now have one crime lab, shall we build three? A closer looks, however, shows that redundancy would *reduce* the taxpayer cost of administering the criminal justice system. One of the authors explains in his forthcoming work (reference below) why rivalrous redundancy would require little or no increase in our basic forensics infrastructure. No grand capital expansion is needed; the central point being the low cost of forensic tests relative to the costs of forensic error.

Using 2002 data, the author (forthcoming) estimated the cost of adding two redundant fingerprint examinations (for a total of three) to each felony case with fingerprint evidence which goes to trial. The average felony sentence in 2002 was about five years. The cost of incarcerating a prisoner was about \$20,000 per year. Even discounting future values to calculate a present value, the costs of incarceration for a false felony conviction were about \$100,000 in 2002. This value is 1,000 times greater than the \$100 cost of two fingerprint tests. The imagined redundancy would eliminate almost all false positive errors in fingerprint examination. Thus, this form of redundancy would save money if the false positive error rate in fingerprint is anything over one in a thousand, or 0.1%.

(The break-even point in the study was 0.115%.) The true rate of false positive errors is likely to be at least 0.8%, and probably more. The author's calculation is thus extremely conservative because it counts only the taxpayer cost of incarcerating the wrongfully convicted, ignoring all other aspects of the social cost of putting the wrong person in jail.

CONCLUSION

In sum, the authors believe that competitive self-regulation would eliminate most errors in our criminal justice system, while reducing taxpayer costs. Paraphrasing, and amending, Madison in *Federalist* 51:

> Competitive self-regulation supplies, by opposite and rival interests, the defect of better motives and greater wisdom. It makes each lab a check on the other, ensuring that the private interest and understanding of every lab may be a sentinel over the public rights.

It is important that we act effectively to ensure not only the reliability of forensic science but also continued public trust in the most vitally scientific element of our criminal justice system. For that to happen, we must tear down the monopoly structure, which has given us needlessly high error rates, and bring forensic science within the fold of our system of checks and balances.

References

Balko, Radley "CSI: Mississippi," *Reason*, October 2007.

Bowers, Michael. 2006. "Problem-Based Analysis of Bitemark Misidentifications: The Role of DNA," Forensic Science International, 159 (supplement 1): S104-s109.

Bromwich, Michael R. 2005. *Second Report of the Independent Investigator for the Houston Police Department Crime Laboratory and Property Room*. Downloaded on 24 May 2007 from http://www.hpdlabinvestigation.org/reports/050531report.pdf

Dror, Itiel and David Charlton. 2006. "Why Experts Make Errors," *Journal of Forensic Identification*. 56(4): 600-616.

FBI Director. 2002. "An Audit of Houston Police Department Crime Laboratory-DNA/Serology Section," December 12-13, 2002.

Giannelli, Paul C. 1997. "The Abuse of Evidence in Criminal Cases: The Need for Independent Crime Laboratories," *Virginia Journal of Social Policy & the Law*, 4: 439-478.

Glenn, Mike. 2006. "Under the microscope: HPD's evidence closely watched this time around," *Houston Chronicle*, 11 July 2006. Downloaded on 24 May 2007 from http://www.chron.com/disp/story.mpl/special/crimelab/4038116.html.

The Innocence Project, "Forensic Science Misconduct," http://www.innocenceproject.org/understand/Forensic-Science-Misconduct.php, accessed on March 10, 2008.

Kennedy, Donald. 2003. "Forensic Science: Oxymoron?" *Science*, 302: 1625.

Khanna, Roma. 2006. "Crime lab's DNA section, shut in '02, will reopen," *Houston Chronicle*, 21 June 2006. Downloaded on 24 May 2007 from http://www.chron.com/disp/story.mpl/special/crimelab/3988198.html.

Koppl, Roger. Forthcoming. "Romancing Forensics: Legal Failure in Forensic Science Administration," in Lopez, Edward, edited, *Government Failure in the Legal System: A Public Choice Review of the Law*, Independent Institute.

Koppl, Roger. 2005. "How to Improve Forensic Science," *European Journal of Law and Economics*, 20(3): 255-286.

Mills, S. and M. Possley. 2004. "Texas Man Executed on Disproved Forensics: Fire that Killed his 3 Children could have been Accidental," *Chicago Tribune*, 9 December 2004.

Mitchell, Jerry. "Innocence Project Letter Blasts Steven Hayne's Work," *Jackson Clarion-Ledger*, February 14, 2008.

Neufeld, Peter. 2005. "The (Near) Irrelevance of *Daubert* to Criminal Justice and Some Suggestions for Reform," *American Journal of Public Health*, 95(supplement 1): S107-S113.

Office of the Inspector General, United States Department of Justice. 2006. *A Review of the FBI's Handling of the Brandon Mayfield Case: Unclassified Executive Summary*. Available at http://0225.0145.01.040/oig/special/s0601/PDF_list.htm.

Office of the Inspector General, United States Department of Justice. 2004. *The FBI Laboratory: A Review of Protocol and Practice Vulnerabilities*, Available at http://www.usdoj.gov/oig/special/0405/final.pdf.

Risinger, Michael, Michael J. Saks, William C. Thompson, and Robert Rosenthal. 2002. "The *Daubert/Kumho* Implications of Observer Effects in Forensic Science: Hidden Problems of Expectation and Suggestion," *California Law Review* 90: 1-56.

State of Maryland v. Bryan Rose, Memorandum Decision, 19 October 2007, Circuit Court for Baltimore County, K06-545.

Stutzman, Rene. 2007a. "Are Innocents Imprisoned? Fingerprint Errors Found," *Orlando Sentinel*, 4 May 2007.

Stutzman, Rene. 2007b. "Fingerprint scandal costs analyst her job," *Orlando Sentinel*, 7 June 2007.

SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology). 2002a. "Friction Ridge Examination Methodology for Latent Print Examiners." Downloaded from http://www.swgfast.org/Friction_Ridge_Examination_Methodology_for_Latent_Print_Examiners_1.01.pdf on 11 June 2007.

SWGFAST (Scientific Working Group on Friction Ridge Analysis, Study and Technology). 2002b. "Quality Assurance Guidelines for Latent Print Examiners." Downloaded from http://www.swgfast.org/Quality_Assurance_Guidelines_for_Latent_Print_Examiners_3.0.pdf on 11 June 2007.

Teichroeb, Ruth. 22 July 2004. "Rare Look Inside State Crime Labs Reveals Recurring DNA Test Problems," *Seattle Post-Intelligencer*.

Williamson, Tara. 12 March 2007. Memorandum to Captain Randy Pittman. Available from the author.

Environmental Law & Property Rights

Buying Green: The Government Steps Up Environmentally Friendly Preferable Purchasing

*By Alec D. Rogers**

With an annual contract budget of over $400 billion, the U.S. government is the largest purchaser of goods and services in the world. Approximately $62 billion in that outlay is spent on information technology equipment and related services. Unlike commercial entities, however, the government's purchasing power is frequently leveraged for purposes other than trying to obtain the item sought at the lowest price. Such purposes include promoting small women-owned and minority-owned businesses, those that are owned by or employ the blind and/or handicapped, and other groups that are considered to be disadvantaged. Similarly, it is often used to punish those who commit other offenses, barring them from the political marketplace for certain legal offenses. In a few cases, it is used to punish otherwise lawful behavior.[1]

One of the most recent trends has been to leverage the procurement process to develop the marketplace for environmentally sound products through a process known as "environmentally preferable purchasing" ("EPP"). Broadly defined, environmentally preferable products are those that "have a lesser or reduced adverse impact on human health and the environment than competing products that serve the same purpose."[2] While the practice is not exactly new at the federal or state level, it has been given additional significant attention in recent years, and the pace is only accelerating as other government programs that provide certification and designation of products as "green" grow and expand as well.

Such programs have existed at the federal level since the Resource Conservation and Recovery Act of 1976 (RCRA), which required federal agencies to develop purchasing programs for products made with recycled waste materials and placed certain mandates on federal grantees to do likewise. The history of EPP, however, demonstrates the gap between policymakers' high level intentions and the actual actions "on the ground" of those doing the purchasing. This article will discuss this history and recent developments aimed at closing the gap to move EPP beyond the aspirational level to finally make it a reality in federal purchasing thirty years after it was instantiated.

I. Energy Policy Act of 1992

Despite the law's general mandates and exhortations, little was accomplished on EPP from 1976 through 2002, most likely because the laws were focused on agency actions and did not provide the sort of concrete, comprehensible direction for individuals necessary for statutes to have any impact. After all, agencies do not purchase goods and services; individual government agents do. Some of these agents are warranted federal procurement officers with extensive training. But much federal purchasing is done with purchase cards by regular federal employees who lack much formal training beyond being told not to use their purchase cards to buy iPods and the like.

The Energy Policy Act of 2002 tried to reinvigorate EPP. It required agencies to purchase energy efficient products and created guidelines to encourage their purchase. In addition it amended RCRA to require the purchase of recycled toner cartridges.

A. Clinton Era Executive Orders

Executive Order 12873 created the Office Federal Environmental Executive and mandated that the government purchase paper with 20% recycled content. It was followed up in 1998 by E.O. 13101, which, among other things, provided greater direction to agency EPP efforts by directing them to reference guidance from the Environmental Protection Agency (EPA). This recognized that many agencies lacked the ability and focus to create their own program, and would need guidance from an expert source.

B. Environmental Designations

Defining what is a "green" item can be very tricky, and is certainly outside the ken of the ordinary procurement official. The federal government has been working on designations that are being used both commercially and by the government to serve as shortcuts to make procuring green technology easier. The first is the EPA's Energy Star program. Energy Star, which is a standard developed in close collaboration with industry to ensure that the designation is neither impossible nor easy to attain, has the advantage of covering numerous product categories. However, it relies heavily on a single environmental metric—energy use—thereby limiting its overall usefulness in incentivizing the development of a comprehensive marketplace for environmentally preferable products.

Another program that is funded through the EPA (and other sources) is the Electronic Product Environmental Assessment Tool ("EPEAT"). EPEAT is more comprehensive than Energy Star. It considers twenty-three mandatory criteria and is flexible in that it allows for different levels of designation (bronze, silver, and gold). However, it currently exists only for desktops, laptops, and monitors. Standards for other categories of equipment are currently underway, however, which will greatly increase EPEAT's usefulness as a tool for EPP in the future.

C. Problems Identified with EPP Execution

Despite these executive orders, a 2001 General Accounting Office (GAO)[3] report found that EPP suffered from two major defects.[4] First, there was little data to verify compliance. Next, there was a low level of awareness among federal employees with contracting responsibilities. Agency procurement systems lacked the necessary mechanisms to determine whether products purchased possessed environmentally preferable traits.

** Alec D. Rogers is former Senior Counsel to the U.S. Senate Committee on Governmental Affairs.*

Reinforcing this lack of data collection is a lack of education among agency officials regarding the legal directives on EPP. "Until the 1990s," GAO reported, "little action was taken to promote such purchases in a government wide or agency wide basis." More discouraging, even after such efforts in that decade to promote EPP, "many procuring officials and other federal purchasers either do no know or implement RCRA requirements for establishing affirmative purchasing programs."

GAO recommended that that Federal Environmental Executive and the Administrator of EPA work with officials at major procuring agencies to develop a process for providing officials with the relevant knowledge about EPP programs and methods for implementing them. Further, the agency said, the Office of Federal Procurement Policy (OFPP) should do more to provide agencies with procurement with specific guidance on fulfilling RCRA's mandates in this regard. Next, GAO recommended that the federal procurement data system be altered to ensure that contractors were complying with RCRA mandates to purchase recycled content products. Finally, GAO recommended that the Office of Federal Financial Management should amend the "Common Rule" to incorporate RCRA's requirements.

II. Early Bush Administration Executive Orders

In August of 2001, President Bush signed executive order 13221—Energy Efficient Standby Power Devices. EO 13221 mandated that only those commercially available, off the shelf, electronic devices that used less than one watt when in "standby" mode be purchased by the US government. If not available, those items with the lowest standby power use should be purchased. This requirement, however, only applied where such products were "life cycle cost effective and practicable" and where the product's "utility and performance was not compromised." GSA, DOD and Energy were to compile a list of items subject to the requirements.

Although 13321 remains "on the books," its utility is fairly minimal. The exceptions pretty clearly undercut the original rule, which was, itself, not very well thought-out. Mandating the use of low power items except in those cases where it did not make sense for one reason or another merely created a situation where acquisition officials were burdened with criteria they were not equipped to apply. 13221 has since been surpassed by more far reaching legal requirements.

A. Energy Policy Act of 2005

Section 104 of the Energy Policy Act of 2005 requires agencies to purchase Energy Star or a Federal Energy Management Program (FEMP) designated products except in those cases where a written determination is made that it is (a) not life cycle cost effective or (b) reasonably available to meet the agency's needs. Although at first blush this might seem to codify existing executive orders, in fact it significantly advances EPP by placing the onus on procurement officials to take action if they are *not* going to purchase an environmentally preferable product. It also provides clear directions to government purchasers by directing them to existing designated products rather than giving them performance metrics, which could be subject to lengthy analysis or, even less helpful, just general direction to purchase "green" products.

Understanding why this is important requires that we contemplate the current state of the federal acquisition workforce. That workforce is universally accepted to be overworked and lacking in sufficient resources. Told that the only way out of an environmentally preferable purchase is to execute a finding, and that to do so would not be cost-effective, or would result in the failure to meet the agency's needs, most overworked procurement officials will simply purchase the environmentally approved item and move on to the next task. The FEMP or Energy Star product is the default purchase barring further analysis, which is not required. It is also important that the Act does not use vague terms, but is specific in what is required. If an Energy Star or EPEAT designated version of what you are looking for is available, you must buy it. Procurement officials are not left to wonder what they need to do to make an "environmentally preferable" purchase.

B. Executive Order 13423 and Subsequent Administrative Action

In early 2007, President Bush signed E.O. 13423, which, among other things, mandates that agencies meet 95 percent of their electronic procurement requirements in those categories with an existing EPEAT standard with EPEAT products among other requirements. This was followed up by forty pages of implementation instructions. Because the procurement mandate is agency-wide, however, rather than focused on individual officials, there will need to be further agency action to ensure that it is implemented through individual performance.

In December, OFPP announced that it was preparing to implement EO 13423 via a policy paper, and issued a draft for comment. The draft provides both substantive guidance and procedures for agencies to implement, and contains a handy reference guide to the myriad provisions existing in law and regulation calling for agency action on EPP. Among these is a mandate that agencies "ensure representation of environmental and energy experts, managers, or technical personnel on integrated procurement teams for all major acquisitions," and consider sustainable design practices, life cycle costs, product take back opportunities, and maximization of energy and resource recovery. Agencies are also required to develop formal affirmative procure programs ("APPs"). APPs will need to, among other things, provide for adequate data reporting on green purchasing, remedying one of the shortfalls identified with EPP in GAO's 2001 report.

CONCLUSION

More mandates, rules, and exhortations will only go so far in getting federal procurement officials and others who actually put their purchase cards on the counter at office supply stores to look for Energy Star, EPEAT designations, and other required indicia of environmentally preferable products. The overburdened federal workforce will need help if it is to ever implement EPP in a meaningful manner. First, it needs to learn of the mandate and requirements. Next, it must be presented with a clear set of operational instructions that allow procurement officials to do their job without significant additional burdens. Finally, there must be mechanisms for meaningful oversight to ensure that these requirements are not being ignored.

One possible tool for oversight may be the development by OFPP and OFEE of an annual scorecard to grade agencies on environmental stewardship. One measure will be an agency's progress on EPP. Similar scorecards for other aspects of the President's Management Agenda have proven useful for communicating to agency heads what administration priorities they will be held accountable for. Scoring performance and making personnel decisions based on it would go a long way to making sure that EPP emerges as something other than just one more box to be checked.

With the expansion of EPEAT, the means to meaningfully implement EPP in a relatively painless manner will soon be at hand. At the end of the day, however, sustained efforts will be needed to communicate to the people "on the ground" the importance of ensuring that the federal government purchases green electronics. This can be done by dedicating the resources to both training procurement officials, and by holding them accountable if they fail to implement the rules. It will be up to the next administration to continue the momentum if EPP in the federal government is to move ahead. EPP's troubled history demonstrates that without sustained support from the top, the progress made since 1976 may slip away as it has in the past.

Endnotes

1 A recent example of this is the prohibition against awarding Department of Homeland Security contracts to those companies that executed a "corporate inversion," which results in a corporate restructuring that moves the headquarters off shore, and reduces the company's US taxes.

2 Federal Acquisition Rule 2.101

3 GAO has since been renamed the "Government Accountability Office."

4 Report of the US General Accounting Office, *Federal Procurement: Better Guidance and Monitoring Needed to Assess Purchases of Environmentally Friendly Products*, GAO-01-430 (June 2001)

Federalism and Separation of Powers

A Floor, Not A Ceiling: Federalism and Remedies for Violations of Constitutional Rights in *Danforth v. Minnesota*

*By Ilya Somin**

Few doubt that states can provide greater protection for individual rights under state constitutions than is available under the Supreme Court's interpretation of the Federal Constitution. More difficult issues arise, however, when state courts seek to provide greater protection than the Court requires for *federal* constitutional rights. Can state courts impose remedies for violations of federal constitutional rights that are more generous than those required by the federal Supreme Court? That is the issue raised by the Court's recent decision in *Danforth v. Minnesota.*[1] In a 7-2 decision joined by an unusual coalition of liberal and conservative justices, the Court decided that state courts could indeed provide victims of constitutional rights violations broader remedies than those mandated by federal Supreme Court decisions. I contend that this outcome is correct, despite the seeming incongruity of allowing state courts to deviate from the Supreme Court's interpretation of the Federal Constitution. The Supreme Court should establish a floor for remedies below which states cannot fall. But there is no reason for it to also mandate a ceiling.

Part I briefly describes the facts and background to *Danforth.* In Part II, I provide a doctrinal justification for the Supreme Court's decision. It makes sense to allow state courts to provide more generous remedies than those mandated by the federal courts in cases where restrictions on the scope of remedies are not imposed by the Constitution itself, but are instead based on policy grounds. State courts can legitimately conclude that these policy grounds are absent or outweighed by other considerations within their state systems, even if they are compelling justifications for restricting the scope of remedies available in federal courts. State courts are in a better position to weigh the relevant tradeoffs in a state legal system than federal courts are. Part III explains the potential policy advantages of allowing interstate diversity in remedies, most importantly inter-jurisdictional competition and an increased ability to provide for diverse citizen preferences and local conditions across different parts of the country. The optimal remedy for a constitutional rights violation in New York may well be different from the optimal remedy for one that occurs in Mississippi.

I. *Danforth* and its Origins

In its 2004 decision in *Crawford v. Washington*, the Supreme Court held that the Confrontation Clause of the Sixth Amendment requires that defendants have the right to "confront" witnesses against them in person at a trial.[2] Three years later, the Court held that states are not required to apply this rule retroactively to pre-*Crawford* convictions.[3] The combination of these two rulings set the stage for *Danforth.*

In 1996, eight years before *Crawford,* a Minnesota Court convicted Stephen Danforth of criminal sexual assault against a minor. The six-year-old victim did not testify at the trial, but the jury "saw and heard a videotaped interview of the child."[4] Danforth challenged his conviction on the grounds that the use of the videotape at his trial violated the Confrontation Clause, as interpreted in *Crawford.* The Minnesota Supreme Court rejected his argument, holding that Minnesota courts were forbidden by federal Supreme Court precedent to "give a Supreme Court decision of federal constitutional criminal procedure broader retroactive application that that given by the Supreme Court."[5] Previous Supreme Court decisions had held that newly announced rules of constitutional criminal procedure do not apply retroactively unless they fall into two narrowly defined categories: rules that forbid state authorities to criminalize the conduct in question and "watershed" rules that "implicate the fundamental fairness of the trial."[6] The Minnesota Supreme Court held that Danforth's case fell outside the scope of both of these categories and concluded that state courts were therefore barred from giving him retroactive relief for this violation of the Sixth Amendment.[7]

In a 7-2 decision, the United States Supreme Court overruled the Minnesota Supreme Court's ruling that state courts are forbidden to grant retroactive relief for violations of constitutional rights in cases where the Federal Supreme Court does not require them to do so.[8] Ironically, the Court's ruling gives state courts greater latitude then they would have been allowed under the Minnesota Supreme Court's approach.[9] The majority held that the case turned on a question of state law remedies, not federal constitutional law.[10] In a forceful dissent, joined by Justice Anthony Kennedy, Chief Justice John Roberts argued that remedies for violations of federal constitutional rights are indeed a matter of federal law, and that the Constitution requires nationwide "'uniformity of decisions throughout the whole United States'" on all federal constitutional issues.[11] At least at first glance, it seems as if Roberts has a point. After all, remedies for violations of constitutional rights are elements of the rights themselves. For example, the Fifth Amendment right to "just compensation" for a taking of private property necessarily includes the right to sue the government for compensation if it takes a citizen's property without paying for it.[12] There is, therefore, some intuitive appeal to the claim that they must be uniform "throughout the whole United States." As Roberts put it, the majority's approach allows "the Federal Constitution... to be applied differently in every one of the several States," thus creating the kind of "disuniformity" that the Constitution was in part established to prevent.[13] However, there are good reasons to permit this kind of disuniformity that are largely ignored by both the dissenters and the majority. Roberts's logic is correct in so far

* *Ilya Somin is Assistant Professor of Law at George Mason University School of Law. This article recently appeared in the Northwestern University Law Review Colloquy.*

as it requires states to provide a minimal level of remedies for violations of federal constitutional rights—a floor. But his logic does not apply with equal force to allowing the Supreme Court to impose a ceiling.

II. Federalism and Policy-Based Limits On Remedies for Rights Violations

Neither the majority nor the dissent in *Danforth* ever seriously considered the fact that limits on the retroactivity of remedies for rights violations do not rest on constitutional mandates but instead rest on policy concerns.[14] When the courts refuse to remedy an admitted constitutional rights violation because of policy considerations that weigh against retroactivity, they are, in effect, subordinating a constitutional concern to a policy one.

Yet this is precisely what happened in *Whorton v. Brockling,* the 2007 Supreme Court case that held that states are not required to apply the *Crawford* Confrontation Clause decision retroactively.[15] In *Whorton*, the Court noted that the procedure followed by the state court in convicting the defendant had violated his Confrontation Clause rights by interpreting the Clause in a way inconsistent with the intent of "the Framers" of the Bill of Rights.[16] However, the Court refused to apply this ruling to pre-*Crawford* cases because the old rule—although based on a flawed interpretation of the Sixth Amendment—did not significantly increase the chances of an inaccurate conviction and therefore did not outweigh the policy considerations weighing against retroactive application of new Supreme Court decisions under *Teague*.[17] Justice O'Connor, the author of *Teague,* has characterized its presumption against retroactivity as an example of how "federal courts exercising their habeas powers may refuse to grant relief on certain claims because of '*prudential concerns*' such as equity and federalism."[18] Other relevant "prudential concerns" weighing against retroactivity include the need to ensure finality in criminal proceedings[19] and the danger of recidivism by offenders released prematurely if their convictions are invalidated.

One might legitimately question whether it is ever permissible for the Court to allow "prudential concerns" to trump constitutional rights. After all, a crucial purpose of enshrining any interest as a constitutional right is precisely to ensure that it overrides ordinary policy considerations, "prudential" or otherwise. The tradeoff between a constitutional right and other objectives that might conflict with it is not for the courts to decide. That decision has already been made by the framers and ratifiers of the Constitution. There are good reasons to believe that the policy judgments of the supermajorities that produce constitutional amendments are likely to be better than those of the Supreme Court.[20] Significantly, the framers and ratifiers did not include a non-retroactivity exception in the Bill of Rights. Perhaps the Court should respect that "prudential" decision rather than subordinating the enforcement of constitutional rights to its own interpretation of prudence.

Nonetheless, I do not pursue this more radical criticism of non-retroactivity here. Justifiably or not, the Court often weakens remedies for constitutional rights when it perceives weighty prudential considerations on the other side. This is particularly true when a new precedent overrides a long-established decision that government officials have relied on in good faith. Most famously, the Court adopted this approach when it ruled in *Brown v. Board of Education II* that southern states need only desegregate their education systems with "all deliberate speed," rather than immediately[21]—despite the fact that continued segregation in what turned out to be a lengthy interim period would lead to an ongoing violation of constitutional rights.

At the same time, the Supreme Court should not have the same kind of power to impose its "prudential" policy preferences on the states as it does when it enforces actual constitutional rights. Federal courts may indeed be in the best position to weigh conflicting policy priorities in federal legal proceedings (assuming that this is a legitimate judicial function at all). State courts, however, are better placed to weigh these issues in the context of state proceedings such as in *Danforth.*

Minnesota courts presumably have greater knowledge about the impact of retroactivity on their own future proceedings than the justices of the Federal Supreme Court. They also have greater incentives to use their knowledge effectively. Should they make a ruling that imposes undue costs on the Minnesota legal system, Minnesota political authorities could curb the state courts' powers by choosing new judges with different views or by passing jurisdiction-limiting legislation. In the twenty-two states with elected judiciaries including Minnesota, judges are subject to electoral checks.[22] In other states, judges are appointed by the governor with participation by state legislatures or "merit commissions."[23] Both methods give judges at least some incentive to consider policy considerations important to their states' judicial systems. By contrast, Minnesota officials and voters have much less influence over the selection of federal judges.

Undoubtedly, both electoral and non-electoral constraints on state judges have significant flaws. For example, widespread political ignorance may greatly reduce the ability of voters to monitor state judges' performance and deny reelection to those who have reached poor decisions.[24] Even so, state judges are clearly more accountable to their states' voters and government officials than federal judges, and therefore have stronger incentives to give due consideration to state-level policy concerns.

The superior knowledge and incentives of state judges relative to federal judges may have little significance in cases where state discretion is limited in order to enforce federal constitutional rights. There, the Constitution does indeed seek to impose "uniformity" of the kind emphasized by Chief Justice Roberts in his dissent. However, the superior position of state judges is very relevant to situations where the supposed justification for federal imposition is simply a matter of "prudential" policy considerations. Here, superior knowledge and incentives counsel in favor of letting state courts set their own rules.

This is especially true with respect to policy arguments against retroactivity that do not apply to state courts invalidating their own state's convictions. For example, Justice O'Connor listed "federalism and comity" among the "prudential concerns" justifying the *Teague* rule.[25] Obviously, these considerations simply do not apply to a state court reviewing the validity of state convictions within its own jurisdiction. By definition, there is no issue of comity in cases like *Danforth*, since comity problems only arise in a situation where one sovereign refuses to respect the decision of another. Nor can there be any "federalism" problem when one state court overrules the decision of another court from its own state.

The *Danforth* majority did recognize that "federalism and comity concerns" do not apply to state courts reviewing their own state's convictions, and also noted that "finality of state convictions is a *state* interest, not a federal one."[26] It even emphasized that there is a "fundamental interest in federalism that allows individual states to define crimes, punishments, rules of evidence, and rules of criminal and civil procedure in a variety of different ways—so long as they do not violate the Federal Constitution."[27] This bedrock principle of federalism, the Court concluded, cannot be overridden by "any general, undefined federal interest in uniformity."[28] However, it failed to draw the more general conclusion that state courts, not federal courts, are in the better position to decide policy issues arising from state judicial rules. Thus, there is a fundamental difference between Supreme Court decisions that enforce federal constitutional rights and those that limit such enforcement on the basis of "prudential" policy considerations.

III. Interstate Variation in Remedies and the Benefits of Federalism

Although I have argued that state courts are generally better placed to evaluate policy concerns about state court remedies than federal courts, it is theoretically possible that Chief Justice Roberts is right to argue that federally imposed "uniformity" in remedies is desirable.[29] Perhaps this is an exception to the general rule that state courts are better judges of state legal rules than federal courts. However, there is good reason to believe that allowing interstate variations in remedies captures some of the standard benefits of federalism. It allows us to reap more of the benefits of interstate diversity, mobility, and competition.

A. Diversity

The ability to satisfy the diverse preferences of populations in different parts of the country is a classic rationale for federalism. Both objective local conditions and citizen preferences may differ from one state to another. It makes sense to allow states to adopt divergent policies in order to take account of such differences.[30]

This point applies to diversity in remedies as much as to other types of policy diversity among states. There are many reasons why the optimal remedy for a constitutional rights violation in one state might be different from the optimal remedy in another. For example, rights violations might be a more common problem in some states than others, which might justify stronger remedies in order to increase deterrence in the state where government officials are more prone to violate the right in question. Similarly, public opinion in State A might value a particular right more than that in State B. A divergence in remedies (with a more generous remedy in State A) could help satisfy the preferences of voters in both states. A uniform federal rule, by contrast, would leave at least one state's voters relatively dissatisfied.

There is also a strong case for interstate variation with respect to the specific question of retroactivity at issue in *Danforth.* If a state has a long, egregious history of violating a particular constitutional right, retroactive application of remedies might be needed in order to root out the systemic consequences of past rights violations. By contrast, this need is likely to be less pressing in a case where the state has rarely violated the right in question. To take one of the most notorious examples in American history, many states—particularly in the South—systematically violated the rights of African-American criminal defendants for decades.[31] In states with this kind of record, retroactive remedies might be more defensible than in states with less history of abuse. Additionally, the costs, as well as the benefits, of retroactivity are also likely to vary between states. In some states, for example, there may be less danger of recidivism and less need to insist on finality of convictions than in others.

These benefits of diversity do not undercut the case for establishing a federally mandated "floor" for constitutional remedies. In the absence of such a floor, states could deny remedies for rights violations entirely, thereby negating the main purpose of creating enforceable constitutional rights in the first place. However, there is no comparable justification for a federally imposed ceiling. If state courts, for their own reasons, decide that they want to provide greater remedies for constitutional rights violations than the Supreme Court requires, they may well have good diversity-based reasons for doing so.

B. Interstate Mobility and Competition

A second crucial rationale for decentralized federalism is the ability of citizens to "vote with their feet" for the state government whose policies they prefer.[32] Citizens dissatisfied with the policies of their state can vote with their feet against them by migrating to a different jurisdiction whose policies they find more congenial. If states are free to adopt diverging policies, there will be more options for potential foot voters. Moreover, competition for taxpaying residents and firms gives states incentives to adopt policies that will attract migrants and convince current residents to stay.[33]

However, foot-voting and competition may not apply as readily to interstate differences in remedies as to other policies. Given the costs of moving, few people or firms are likely to migrate merely because one state has better remedies than another for violations of constitutional rights. This, in turn, reduces the likelihood that states will try to compete with each other on this dimension. Nonetheless, there might be exceptions to this generalization. Residents who are particularly concerned about the danger of a given rights violation may take remedies into account in their moving decisions. In the Jim Crow era, when federal courts were extremely lax in enforcing constitutional protections for African-American criminal defendants, black migrants did indeed take into account the fact that northern criminal justice systems treated them more favorably than southern ones.[34] On the other hand, excessive remedies that overdeter law enforcement might be curtailed by migration on the part of residents seeking to move to areas with lower crime rates.

On balance, foot-voting and interstate competition are less compelling rationales for allowing variations in remedies than diversity. But they have some force, nonetheless.

CONCLUSION

There is good reason for the Supreme Court to establish a floor for remedies for federal constitutional rights violations. On the other hand, there is no comparable justification for it to also establish a ceiling that state courts are not allowed to exceed. To the extent that the Supreme Court's *Danforth* decision tracks this distinction, it should be welcomed.

At this time, the extent to the Court's ruling applies outside the Sixth Amendment context remains unclear. Presumably, the Court's reasoning applies to all cases where state courts provide more generous remedies for violations of federal constitutional rights than the Supreme Court mandates. However, the *Danforth* decision fails to provide a comprehensive explanation of the right-remedy distinction and also fails to explicitly consider the question of how broadly its ruling will apply. However, the Court did hold that "the remedy a state court chooses to provide its citizens for violations of the Federal Constitution is primarily a question of state law" and is therefore not subject to a federal court-imposed ceiling.[35] This suggests that its logic applies to all such remedies, not just to those involving criminal proceedings. Certainly, the justification offered here for the floor-ceiling distinction in *Danforth* applies with equal force to similar cases involving other constitutional rights.

Endnotes

1 128 S.Ct. 1029 (2008).

2 Crawford v. Washington, 541 U.S. 36, 68–69 (2004).

3 Whorton v. Bockting, 127 S.Ct. 1173 (2007).

4 Danforth v. Minnesota, 128 S.Ct. 1029, 1033 (2008).

5 Danforth v. State, 718 N.W.2d 451, 456 (Minn. 2006), *rev'd*, *Danforth*, 128 S.Ct. 1029 (2008)..

6 Teague v. Lane, 489 U.S. 288, 310 (1989) (plurality opinion). Although the so-called "*Teague* rule" was first outlined in a plurality opinion joined by only four justices, it was accepted by the majority of the Court in a later decision that followed soon afterwards. *See* Penry v. Lynaugh, 492 U.S. 302, 313 (1989) (endorsing and applying the *Teague* rule).

7 *Danforth*, 718 N.W.2d at 455–57.

8 *Danforth*, 128 S.Ct. at 1038–47.

9 Several other state supreme courts have held that the *Teague* rule does not constrain state courts acting to remedy constitutional rights violations that occurred during state postconviction legal proceedings. *See, e.g.,* Daniels v. State, 561 N.E.2d 487, 489 (Ind. 1990); State *ex rel.* Taylor v. Whitley, 606 So. 2d 1292, 1296–1297 (La. 1992); State v. Whitfield, 107 S.W.3d 253, 266–268 (Mo. 2003); Colwell v. State, 118 Nev. 807, 816–819, 59 P. 3d 463, 470–471 (Nev. 2002) (per curiam); Cowell v. Leapley, 458 N.W.2d 514, 517–518 (S. D. 1990); *cf.* State *ex rel.* Schmelzer v. Murphy, 548 N.W.2d 45, 49 (Wisc. 1996) (holding that state courts are not required to follow the *Teague* rule, but adopting it of its own volition as a matter of state law).

10 *Danforth*, 128 S.Ct. at 1046–47.

11 *Id*. at 1053 (Roberts, C.J., dissenting) (quoting Martin v. Hunter's Lessee, 14 U.S. (1 Wheat.) 304, 347–48 (1816)).

12 Cf., Armstrong v. United States, 364 U.S. 40 (1960) (upholding such a suit as part of the right to "just compensation").

13 *Danforth,* 128 S.Ct. at 1054.

14 The importance of this distinction to the outcome of *Danforth* is also not considered in Professor Michael Dorf's recent critique of *Danforth,* the only previous published analysis of *Danforth* by a legal scholar that I know of. *See* Michael Dorf, *Did Justice Stevens Pull a Fast One? The Hidden Logic of a Recent Retroactivity Case in the Supreme Court*, Findlaw's Writ, Feb. 25, 2008, *available at* http://writ.news.findlaw.com/dorf/20080225.html (last visited Feb. 27, 2008).

15 127 S.Ct. 1173 (2007).

16 *Id*. at 1182.

17 *Id*. at 1181–84.

18 Withrow v. Williams, 507 U. S. 680, 699 (1993) (O'Connor, J., concurring in part and dissenting in part) (emphasis added).

19 *See Danforth*, 128 S.Ct. 1029, 1040–41 (2008) (noting that the *Teague* rule is partly based on the need to ensure "finality of state convictions").

20 For a more detailed discussion of this point, citing relevant literature, see Ilya Somin, *'Active Liberty' and Judicial Power: What Should Courts Do to Promote Democracy?* 100 Nw. U. L. Rev. 1826, 1855–59 (2006).

21 349 U.S. 294 (1955).

22 Stephen J. Choi, et al., *Professionals Or Politicians: The Uncertain Empirical Case for an Elected Judiciary*, U. Chicago Law & Econ. Working Paper No. 357 (Aug. 2007), at 3, *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1008989.

23 *Id*.

24 *See, e.g.,* Ilya Somin, *Voter Ignorance and the Democratic Ideal*, 12 Critical Rev. 413, 415–38 (1998) (discussing the impact of political ignorance).

25 Withrow v. Williams, 507 U. S. 680, 699 (1993) (O'Connor, J., concurring in part and dissenting in part) (emphasis added).

26 Danforth v. Minnesota, 128 S.Ct. 1029, 1041 (2008) (emphasis in original).

27 *Id*.

28 *Id*.

29 See n. 10–12 and accompanying text.

30 For summaries of the diversity rationale for decentralization, *see, e.g.,* Ilya Somin, *Closing the Pandora's Box of Federalism: The Case for Judicial Restriction of Federal Subsidies to State Governments,* 90 Geo. L.J. 461, 464–65 (2002); and Michael W. McConnell, *Federalism: Evaluating the Founders' Design*, 54 U. Chi. L. Rev. 1484, 1493–95 (1987).

31 *See, e.g.,* Randall Kennedy, Race, Crime, and the Law 76–134 (1997) (detailing this history).

32 *See* Somin, *Political Ignorance and the Countermajoritarian Difficulty: A New Perspective on the "Central Obsession" of Constitutional Theory,* 89 Iowa L. Rev. 1287 1344-50 (2004) (discussing "foot voting" and contrasting it with conventional ballot box voting).

33 For the benefits of interstate competition, see Somin, *supra* note 30, at 465–69; *see also* Thomas R. Dye, American Federalism: Competition Among Governments 1–33 (1990).

34 *See, e.g.,* Daniel M. Johnson & Rex R. Campbell, Black Migration in America: A Social and Demographic History 84-85 (1981) (discussing this motive for black migration to the North in the early twentieth century).

35 *Danforth*, 128 S.Ct. at 1045.

The Fourth Amendment and Federalism: The Supreme Court Tussle Over the Two

*By Marc M. Harrold & Michael John Gorman**

This term, the Supreme Court will decide two cases on the Fourth Amendment. Both come from state supreme courts (Arizona and Virginia) and demonstrate a subtle but sizeable trend in state high courts to find greater individual protections in the Constitution, without first looking to their own charters of freedom. By finding those rights in the federal constitution, state courts are forcing the Supreme Court's hand. It is a needless game of chicken that good federalism can avoid.

Commonwealth of Virginia v. David Lee Moore

Tension between the federal government and the several states is nothing new. Indeed, it is expected in our system of dual sovereignty. What makes *Virginia v. Moore* revolutionary is its needless stretching of federalism's tight strings.

About the Case

In September 2003, a Portsmouth detective overheard radio chatter about a man known as "Chubs" driving around the city. The detective knew that "Chubs" was recently released from federal prison and had a suspended license. He radioed citywide for officers seeing "Chubs" to stop him. Two other detectives stopped "Chubs," later identified as David Moore, who in fact was driving on a suspended license. Although Virginia law mandated issuing a citation to Moore,[1] the detectives chose to arrest Moore. They handcuffed him, placed him in a squad car, and read him the *Miranda* warnings. While in custody, Moore also gave written consent to search his hotel room. There the detectives realized that, through miscommunication, neither officer had searched Moore at the scene of his arrest. So they searched him at the hotel room, finding crack cocaine and cash.

Moore admitted owning the crack. But at trial Moore moved to suppress evidence obtained from the search of his person, arguing that the search violated the Fourth Amendment, among others. The trial court denied Moore's motion and, at a bench trial, found him guilty of possession with intent to distribute.

On appeal, a panel of the Virginia Court of Appeals found that Moore's salvation lay in *Knowles v. Iowa*,[2] where the Supreme Court decreed that issuance of a citation for a traffic offense does not justify a full field search. In doing so, however, the panel may have ignored another Supreme Court case, *Atwater v. City of Lago Vista*,[3] where the Court permitted a field search for someone placed under custodial arrest for a traffic offense. It justified its tight embrace of *Knowles* by explaining that § 19.2-74 distinguished Moore's case from *Atwater*, because an officer cannot perform a custodial arrest for a traffic offense pursuant to § 19.2-75. Only a citation may be issued. And so, Moore's arrest and subsequent search were unconstitutional, making suppression "a logical and necessary extension of the decision in *Knowles*."[4]

The Commonwealth of Virginia moved for an en banc consideration and the court of appeals obliged. The full court reversed, holding that as long as the officers had probable cause to believe an offense occurred in their presence, they could arrest for that offense without offending the Fourth Amendment, despite Virginia law. Just because Virginia Code § 19.2-74 mandated a citation instead of an arrest did not create a *federal* constitutional violation and remedy. The court reinstated Moore's conviction. The Virginia Supreme Court reversed the en banc court of appeals, distinguishing *Atwater*. Because *Atwater* dealt with the validity of an arrest itself, not a subsequent search, the Virginia Supreme Court found its force was weak. Also, the court found that the officer in *Atwater* benefited from a state statute that gave complete discretion to arrest; because of §19.2-74, the officers in *Moore* did not. Because the detectives could not have made an arrest under state law, *Knowles* (and Virginia cases interpreting it) required suppressing the fruits of a full, field-type search. By the Virginia Supreme Court's handiwork, § 19.2-74 now had a federal constitutional patina; the remedy for violating it was suppression.

The U.S. Supreme Court has decided to weigh in on that conclusion.[5]

At the High Court

Besides the parties' briefs, the court benefited from briefs submitted by a modest set of seven amici. Virginia, along with its amici,[6] feared an uncertainty in Fourth Amendment jurisprudence. If each state could shellac constitutional varnish onto its statutes, the constitution actually loses its luster. Police officers, already perplexed by search and seizure standards, would become impotent in the field, having to determine when and how they can search after arresting.

Virginia also cried foul over its supreme court's encroachment on federal constitutional standards. Any groundswell to supply greater protection should come from state law sources (e.g., state constitutions), not from grafting state statutory standards onto federal constitutional analysis. Its own supreme court, the Commonwealth argued, had usurped federal authority.

Standing squarely between the twin pillars of uniformity and proper constitutional order, the Commonwealth and its amici petitioned for a federal constitution without state law accessorizing. Otherwise, as the United States as amicus made clear, "[c]onstitutionalizing the myriad and often technical state restrictions on arrest [would risk] creating a 'bog of litigation.'"[7]

Moore was more optimistic about the legal system's ability to cope. The Virginia Supreme Court's decision, according to Moore, was grounded in the principles of the Fourth

** Marc Harrold is Senior Counsel at the National Center for Justice and the Rule of Law and Visiting Professor at the University of Mississippi School of Law. Michael John Gorman is a member of the state bar of Georgia and a freelance appellate brief writer.*

Amendment. It was consistent with the Supreme Court's search and seizure jurisprudence:

As the Court has written since time immemorial, the cornerstone of the Fourth Amendment is reasonableness. To be reasonable an arrest and search must be justified by a significant governmental interest. No significant governmental interest can exists because Virginia's legislature determined that a custodial arrest was unnecessary under the circumstances.

By that logic, Moore continued, traces of state law are always discernable in the federal constitutional analysis.[8] The Virginia Supreme Court did nothing novel. Far from usurping federal authority, the court followed it.

Oral Argument

Oral argument before the Supreme Court was contentious. Justice Ginsburg began the questioning by asking Virginia's deputy solicitor general whether the search would have been illegal if the detectives had followed state law (by issuing a summons).

"Yes," the deputy replied.

"So," Justice Ginsburg asked, "would you explain the logic to saying that when police violate State law, then the evidence can come in; but when they comply with State law, it can't?"

That exchange colored Virginia's remaining time. Justice Scalia wanted to know who has arrest authority (a federal janitor, for instance) and exactly what the Commonwealth wanted from the Court. The justices seemed interested in how state law should affect federal constitutional doctrine. Should the nature of the offense, defined under state law, color the analysis? Should there be any color in the black-and-white probable cause standard?

The United States was next. Justice Scalia asked the deputy solicitor general whether the federal janitor at the Justice Department could arrest.

"Well, certainly, Justice Scalia, such an individual wouldn't have positive law authority to engage in an arrest," the deputy responded.

"Just as this person here didn't have positive law authority to engage in an arrest." Justice Scalia retorted.

To the United States, that may have been, but that does not alter the proper analysis: reasonableness. In this arena, it is exemplified by the black-and-white probable cause standard. Positive law authority, weighted however, should be one consideration in that standard. Thus, no state color is necessary (or proper) according to the United States.

The justices peppered Moore's attorney, Thomas Goldstein, with practical considerations. Justice Breyer wanted to know why the Fourth Amendment must turn on trivial things when cases like *Whren v. United States* (garbage search) refused to go down that road. Justice Alito and the Chief Justice wanted to know where to draw the line. Was it at state statutes or perhaps at departmental policy? Goldstein did not bite at the distinction. He stayed focused on the singular theme that an illegal arrest, no matter how it is illegal, breeds an unreasonable search.

The justices also broached the question of whether a constitutional remedy was even appropriate. If Virginia did not provide an exclusion remedy for violating its statute, why should the Court? Goldstein answered, because the evidence flowed from unconstitutional conduct. In a broader sense, the Court's jurisprudence implicates state law in some, but not all, cases. But this case surely does, Goldstein said.

Needlessly Stretching the Tight Strings

Indeed, primary control over this case rested with the Virginia Supreme Court, which could have resolved it without resorting to the Fourth Amendment. Under principles of federalism and institutional legitimacy, the court should have looked to its law first. Virginia law, not the Fourth Amendment, should have been the harbor of Moore's salvation. Perhaps in the contours of the Virginia Constitution a remedy lay slumbering. But it remained undiscovered because the court jumped instead to the federal constitution. Why? In the past, that court has held that the state constitution—the charter of its sovereign character—offers no more protection than the Fourth Amendment.[9] In Virginia, one's sole recourse is to the federal constitution, because Virginia courts have so chosen. That is an unfortunate result, antithetical to principles of good federalism.

Federalism demands more of Virginia courts. It demands they reclaim the thrown of state constitutional analysis. It demands they look to the state constitution as the primary source of positive rights, before recourse to the federal constitution. The court should look there even if it is likely to find nothing. If the Virginia Constitution, after having been meaningfully interpreted, provided no relief here, then so be it. Proceed to the Fourth Amendment. In the end, the court should recognize that the result is immaterial so long as the analysis is not.[10]

Had the Virginia court interpreted its constitution, the case would have been over. Instead, the court jacked up the federal constitutional floor across the country. Enter the Supreme Court that now must explore legal propriety and national implications. It is a needless exercise.

Arizona v. Gant

The U.S. Supreme Court's upcoming opinion in *Arizona v. Gant*[11] will re-examine the breadth and nature of one of the Fourth Amendment's most well-established exceptions: search incident to arrest.[12] The ruling has the potential to be a landmark. Besides affecting the Arizona state case, the ruling has the potential to affect one of the most well-established and enduring exceptions to the Fourth Amendment to the U.S. Constitution.[13]

The Facts

After receiving a tip about narcotics activity, two uniformed Tucson police officers went to a house where Rodney Gant answered the door.[14] The officers asked if the owner was home. Gant told them that he would be returning later in the afternoon. The officers left and ran a records check. They learned that Gant had a suspended driver's license and an outstanding warrant for driving on a suspended license.

Later that evening, the officers returned to the house. While they were at the residence, Gant drove up, parked in the driveway, and got out of the car. As he exited, an officer beckoned. Gant walked approximately eight to twelve feet towards the officer, who immediately placed him in handcuffs.

Gant was locked in the backseat of a patrol car, under the supervision of a police officer. The scene was "secure." The two officers then searched the passenger area of Gant's vehicle and found a weapon and a plastic baggie containing cocaine which led to criminal charges.[15]

The Arizona Supreme Court's Misinterpretation of Robinson and Re-insertion of Justification and 'Balancing' Analyses

In *Gant*, the Arizona Supreme Court held that "*Robinson* does not hold that *every* search following an arrest is excepted from the Fourth Amendment's warrant requirement."[16] This is true. First, a search incident to arrest is not "*every*" search that is performed post-arrest, but, instead is a search limited in scope to the suspect's person and open and closed containers within the wingspan or control of the arrestee.[17]

The Arizona Supreme Court then went on to state that "[o]nce those concerns are no longer present, however, the 'justifications [underlying the exception] are absent' and a warrant is required to search." The court cites two cases to bolster its holding; curiously both were decided prior to *Robinson*.[18] Contrary to what the Arizona Supreme Court says, *Robinson* "teaches" us that additional justification is not necessary, stating "[a] custodial arrest of a suspect based on probable cause is a reasonable intrusion under the Fourth Amendment; that intrusion being lawful, a search incident to the arrest *requires no additional justification*. It is the fact of the lawful arrest which establishes the authority to search…."[19]

The Arizona Supreme Court decided on a factual basis whether the "justification[s] [underlying the exception]"[20] were present, even though the U.S. Supreme Court has expressly held that no additional "justification" is needed. It is the status of the individual as one whom has been lawfully arrested (custodial arrest) that allows the police to perform the search.[21]

The Arizona Supreme Court also ignores another aspect of the *Robinson* holding: that the search incident to arrest doctrine is not only an *exception* to the Fourth Amendment, but is also a *per se* reasonable search under the Fourth Amendment.[22] As such, the type of balancing[23] the court engages in is unnecessary and prohibited because the U.S. Supreme Court has already determined that a search incident to arrest of a suspect who has been subject to lawful custodial arrest is *per se* reasonable.

Gant's Potential to Impact the Search Incident to Arrest Doctrine as Applied to Digital Devices

The progression of a great deal of Fourth Amendment caselaw can be seen in light of the emerging technologies that make up the fact patterns at issue. Automobiles,[24] helicopters,[25] buses,[26] public phone booths,[27] pen registers,[28] and portable thermal imaging devices[29] are all examples of how Fourth Amendment jurisprudence has evolved with how Americans live, travel, and communicate.

Currently, the law permits the police, pursuant to a lawful custodial arrest and without further justification (in other words based solely on the custodial arrest) to search the person of the arrestee and the area, including containers, (open or closed) within the reach or under the arrestee's control.[30] While it may seem intuitive that this search would be limited to a search for evidence, instruments of escape or items that could pose a danger to the arresting officers, *Robinson* rejects the need to articulate the particular reason police perform a search incident to arrest. The facts of *Robinson* are illustrative. In *Robinson*, the police officer, after arresting an individual for a revoked driver's license, searched a crumpled cigarette package and discovered heroin. There was no specific or additional justification[31] to search the crumpled cigarette package; it was searched because it was on the person of an arrestee. Nothing more. Nothing less.

As such, if digital devices (iPhones, iPods, Xboxes, laptops, the list is long and ever-expanding) are simply containers (as many courts have found them to be) then they can be searched incident to arrest if they are on the person of, or within the wingspan or under the control of, an individual subjected to lawful custodial arrest without additional justification, consistent with the U.S. Supreme Court's holding in *Robinson*.

While the *Gant* case does not specifically involve the search of a digital device, the ruling could have a major impact on searches of digital devices incident to arrest. If the U.S. Supreme Court limits the scope of *Robinson* by re-inserting a justification requirement the effect will be two-fold: specifically, the search of digital devices incident to arrest will be curtailed dramatically and, more generally, it will mark a dramatic shift in the application of one of the Court's most well-established Fourth Amendment doctrines.

Endnotes

1 Va. Code Ann. § 19.2-74 (Michie 2007).

2 525 U.S. 113 (1998).

3 532 U.S. 318 (2004).

4 Moore v. Commonwealth, 609 S.E.2d 74, 79 (Va. 2005).

5 Virginia v. Moore, 128 S. Ct. 28 (2007).

6 The Commonwealth was joined by a gaggle of its sister states who spent considerable time stressing the "profound consequences" that follow from the Virginia Supreme Court's decision. Among these are: a constitutional remedy that varies state-to-state; potential for conflicting judgments among the states and between state courts and the Supreme Court of the United States; and, a general misapprehension of search and seizure standards.

7 Brief of United States as Amicus Curiae, Virginia v. Moore, No. 06-1082, at 5 (filed Nov. 2007) (citing Wyoming v. Houghton, 526 U.S. 295, 305 (1999)).

8 For example, the Supreme Court has held that "[s]tate law determines the validity of arrests without warrant." Johnson v. United States, 333 U.S. 10, 15 n.5 (1948*); see also* Michigan v. DeFillippo, 443 U.S. 31 (1979).

9 El-Amin v. Commonwealth, 607 S.E.2d 115, 116 n.3 (Va. 2005), interpreting Va. Const. art. I, § 10 (Virginia's search and seizure analog).

10 To see which states have recognized that truism of proper federalism, see Michael J. Gorman, *Survey: State Search and Seizure Analogs*, 77 Miss. L.J. 417 (2008).

11 162 P.3d 640 (Ariz. 2007).

12 This article specifically discusses the justification question addressed by the Arizona Supreme Court in *United States v. Gant*. The court also discusses (among others) *Chimel v. California*, 395 U.S. 752 (1969) and *New York v. Belton*, 453 U.S. 454 (1981). Full discussion of these cases, and their respective doctrines, is outside the scope of this short article. For example, in *Thornton v. United States*, the Supreme Court ruled that the rule of *Belton* is not limited to situations where the officer makes contact with the occupant

while the occupant is inside the vehicle—"*Belton* governs even when an officer does not make contact until the person arrested has left the vehicle." Thornton v. United States, 541 U.S. 615, 617 (2004).

13 As indicated, herein, U.S. Supreme Court precedent has established that search incident to lawful arrest is not only an exception to the Fourth Amendment but is also a per se reasonable search under the Fourth Amendment.

14 Arizona v. Gant, 143 P.3d 379 (2006).

15 *Id.*

16 *Id.*

17 "[W]hen a policeman has made a lawful custodial arrest of the occupant of an automobile, he may, as a contemporaneous incident of that arrest, search the passenger compartment of that vehicle." New York v. Belton, 453 U.S. 454, 460 (1981). Further, "[i]t follows from this conclusion that the police may also examine the contents of any containers found within the passenger compartment, for if the passenger compartment is within reach of the arrestee, so also will containers in it be within his reach." *Id*.

18 United States v. Robinson, 414 U.S. 218 (1973). The Arizona Supreme Court curiously cites to *Preston v. United States*, 376 U.S. 364 (1964) and *Chambers v. Maroney*, 399 U.S. 42 (1970), both decided after *Robinson*. The court cites to these cases for propositions that were rejected in *Robinson* in a circular-logic fashion.

19 *Robinson*, 414 U.S. at 235.

20 Arizona v. Gant, 143 P.3d 379 (2006) (internal citations omitted).

21 In theory, the Arizona Supreme Court could have asserted that individuals are afforded more protection in a state criminal matter due to Arizona State Constitution (if such protections exist); this was not at issue in this case.

22 *Robinson*, 414 U.S. at 235 ("…and we hold that in the case of a lawful custodial arrest a full search of the person is not only an exception to the warrant requirement of the Fourth Amendment, but is also a 'reasonable' search under [the Fourth Amendment].")

23 The traditional balance between the government's need to intrude upon the protected interest of an individual (protected interest determined by the two-pronged test (objective/subjective) to establish a reasonable expectation of privacy.)

24 Carroll v. United States, 267 U.S. 132 (1925).

25 Florida v. Riley, 488 U.S. 445 (1989).

26 Florida v. Bostick, 501 U.S. 429 (1991).

27 Katz v. United States, 389 U.S. 347 (1967).

28 Smith v. Maryland, 442 U.S. 735 (1979).

29 Kyllo v. United States, 533 U.S. 27 (2001).

30 *See generally* New York v. Belton, 453 U.S. 454 (1981). If the U.S. Supreme Court, in *Gant*, limits the scope of its *Belton* ruling, like a limitation of the scope of the *Robinson* ruling, this also could have a dramatic impact in the context of the search incident to arrest doctrine and digital devices. Again, a full discussion of *Belton*, and other search incident to arrest cases, is outside the scope of this article.

31 It does not appear that it was searched specifically for evidence, instruments of escape or weapons.

Financial Services and E-Commerce

Reflections on Credit Card Practices and Legislative Reform Proposals

*By Todd J. Zywicki**

The growth in consumer use of credit cards over the past three decades has transformed the American economy, placing in consumers' hands one of the most powerful financial innovations since the dawn of money itself. Credit cards have transformed the ways in which we shop, travel, and live. They have enabled the rise of the e-commerce economy, delivering goods and services to consumers' doorsteps and permitting consumers to shop when and where they like, unconstrained by traditional limits on competition and consumer choice. They have enabled consumers to travel the world without the inconvenience of travelers' checks. And they have transformed the way in which we live, from such small improvements such as relieving us the inconvenience of checks and frequent visits to ATM machines to large improvements such as providing security against crime. Credit cards can be used as a transactional medium, a source of credit, or even as a short-term source of cash. Credit cards provide consumers with additional benefits, from cash back on purchases, frequent flier miles, car rental insurance, dispute resolution services with merchants, and 24-hour customer service. It has been aptly observed that that with a credit card you can buy a car; without a credit card you cannot even rent one. Many of these benefits, of course, have been most salient for lower-income, young, and other similar populations, and unsurprisingly, growth in credit card use has been rapid among those populations.

But the myriad uses of credit cards and the increasing heterogeneity of credit card owners has spawned increasing complexity in credit card terms and concerns about confusion that may reduce consumer welfare. American consumers encounter complexity every day in the goods and services they purchase, such as cars, computers, and medical services, just to name a few. And the complexity of credit card terms is modest when compared to that of the Internal Revenue Code, as are the penalties (financial and otherwise) for failure to understand its terms. The relevant issue for regulation, therefore, is whether the complexity is warranted in light of its benefits.

In considering whether further legislation or regulation of credit card terms or disclosures is appropriate, two questions should be considered. First, what is the problem to be corrected through regulation? And second, will the benefits of the regulation justify the costs, including the unintended consequences of the regulation?

Based on what is known about consumer use of credit cards and credit card practices, it is doubtful that an analysis of these simple questions can justify further governmental intervention in the credit card industry. In fact, the increasing dynamism of the credit card industry suggests that regulators would be better served by revisiting, modernizing, or reconsidering certain extant regulations, rather than piling on additional regulation.

This is not to imply that certain credit card issuers or practices do not seem unfair or improper. But there are ample tools for courts and regulators to attack deceptive and fraudulent practices on a case-by-case basis when they arise. Unlike case-by-case common law adjudication, however, legislation or regulation addresses itself to *categorical* rulemaking, thus before categorical intervention is warranted it is necessary to examine whether categorical problems have arisen.

I. What is the Problem To Be Corrected through Regulation?

Advocates of greater regulation have alleged three problems that are purported to justify additional regulation of the credit card market: (1) Consumer overindebtedness caused by access to credit cards, (2) Unjustifiably "high" interest rates on credit cards, and (3) A growing use of so-called "hidden" fees. Reviewing the empirical evidence available on these issues, however, there is no sound evidence that any of them present a meaningful problem for which greater regulation is appropriate.

A. Consumer Overindebtedness

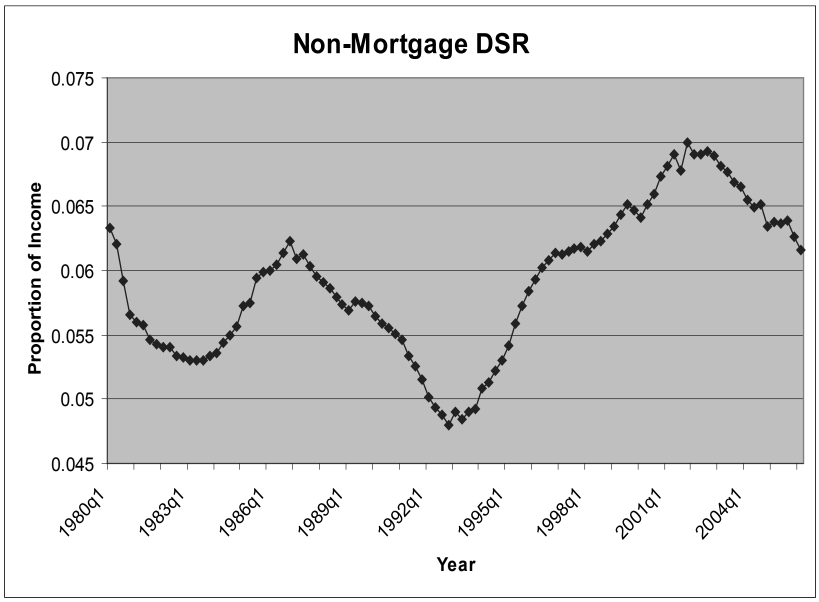
There is no doubt that consumer use of credit cards has increased over time, as has credit card debt. But available evidence reveals that this increase in credit card debt has not in fact resulted in an increased financial distress for American households. Instead, this increased use of credit cards has been a *substitution* from other types of consumer credit to an increased use of credit cards.[1] For instance, when consumers in earlier generations purchased furniture, new appliances, or consumer goods, they typically purchased those items "on time" by opening an installment loan and repaying the loan in monthly payments or through a layaway plan. A consumer who needed unrestricted funds to pay for a vacation or finance a car repair would typically get a loan from a personal finance company or a pawn shop. Today, many of these purchases and short-term loans would be financed by a credit card, which provides ready access to a line of credit when needed, without being required to provide a purchase-money security interest, dealing with the up-front expense and delay of a personal finance loan, or pawning goods.[2] Credit cards are far more flexible and typically less expensive than these alternative forms of consumer credit, thereby explaining their rapid growth in consumer popularity over time. Federal Reserve economist Thomas Durkin observes that credit cards "have largely replaced the installment-purchase plans that were important to the sales volume at many retail stores in earlier decades," especially for the purchase of appliances, furniture, and other durable goods.[3] Former Federal

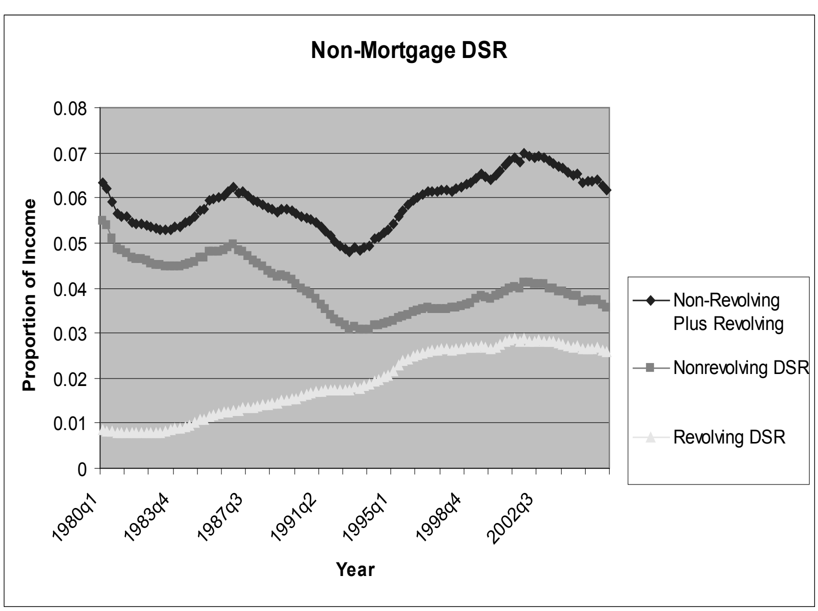
** Todd. J. Zywicki is Professor of Law at George Mason University. This article is adapted from testimony before the House of Representatives Financial Services Committee Subcommittee on Financial Institutions and Consumer Credit Hearing on "Credit Card Practices: Current Consumer and Regulatory Issues." Portions of this testimony also draw on a forthcoming book by the author on consumer credit and consumer bankruptcy.*

Reserve Chairman Alan Greenspan similarly observed, "[T]he rise in credit card debt in the latter half of the 1990s is mirrored by a fall in unsecured personal loans."[4]

In fact, the evidence suggests that the growth in credit cards as a source of consumer credit is explained almost completely by this substitution effect. Thus, even as credit card use has risen rapidly over time, it does not appear that this has contributed to any increase in consumer financial distress.[5]

Since 1980, the Federal Reserve has calculated on a quarterly basis the "debt service ratio," which measures the proportion of a household's income dedicated each month to payment of its debts.





As this figure illustrates (top), the overall debt service ratio for non-mortgage debt (consumer revolving plus non-revolving debt) has fluctuated in a fairly narrow band during the period 1980 to 2006. In fact, the non-mortgage debt service ratio was actually slightly higher at the beginning of the data series in 1980 (0.0633) than at the end in the first quarter of 2006 (0.0616) with local peaks and troughs throughout.

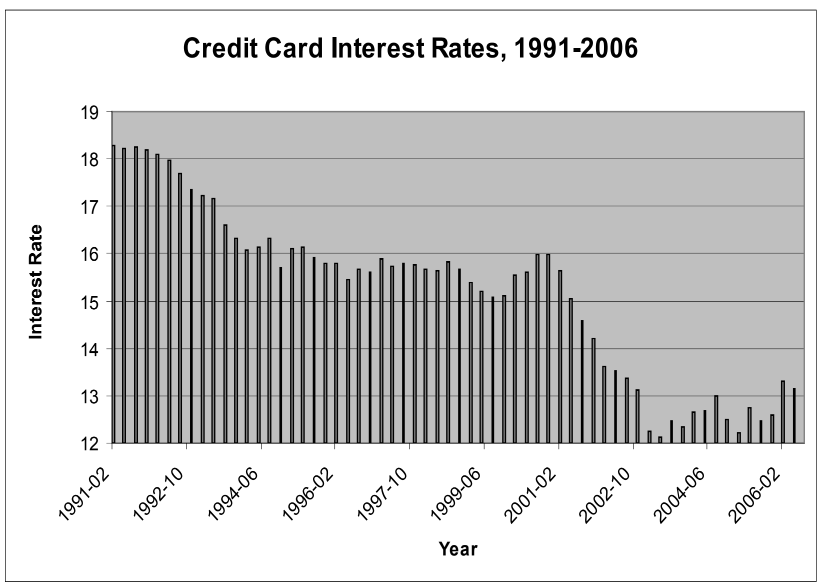
Further isolating non-mortgage consumer debt into revolving and nonrevolving components illustrates the substitution effect (see bottom graph).

As can be readily observed, from 1980 there has been a gradual downward trend in the debt service burden of non-revolving installment credit, such as car loans, retail store credit (such as for appliances or other consumer goods) and unsecured loans from personal finance companies, that mirrors the upward trend for the credit card debt service burden over this same period, leaving the overall consumer credit debt service ratio unchanged. Moreover, according to the Survey of Consumer Finances, the percentage of households in financial distress (as measured by a total debt service ratio, including mortgage credit, of greater than 40%) has fluctuated within a narrow band since 1989.[6]

This substitution effect of credit card for other types of consumer credit has been most pronounced for lower-income debtors, primarily because this group historically has faced the most limited credit options; thus, credit cards are likely to seem especially attractive to them. As a report of the Chicago Federal Reserve Bank concluded, "The increase in the credit card debt burden for the lowest income group appears to be offset by a drop in the installment debt burden. This suggests that there has not been a substantial increase in high-interest debt for low-income households, but these households have merely substituted one type of high-interest debt for another."[7] As with the overall population, the percentage of lowest-quintile households in financial distress has been largely constant since 1989, and in fact, the percentage of lowest-income households in financial distress is actually at its lowest level since 1989.

In fact, it is likely that this data actually tends to *overestimate* the contribution of revolving debt to the debt service ratio, because of peculiarities in the way in which the debt service ratio is measured. First, there has been a dramatic increase in household wealth holdings over the past decade or so, first because of the roaring stock market of the late 1990s, and then the rapid appreciation in housing values into the 2000s. Because consumers rationally borrow against and consume some percentage their accumulated wealth, during periods of rapidly increasing household wealth (such as during the 1990s) consumers would be expected to increase their consumption and consumer debt in order to liquidate some of this accumulated wealth. The ratio of consumer credit to household net worth has been about 4% of household wealth for at least the past fifty years; thus, as consumer wealth rises, consumers will tend to increase their debt holdings even though their measured income does not increase.[8]

Second, the data used here to measure revolving credit likely tends to overestimate the true amount of revolving credit because of a rise in transactional use over time, an overestimation that tends to grow over time. Revolving credit is measured by the credit card balance *outstanding* at the end of a given month, regardless of whether it is actually revolved or paid off at the end of the billing cycle. As a result, the data also report as part of outstanding revolving credit balances on transactional accounts that will be paid at the close of the billing cycle, but happen to be outstanding at the time of reporting. Because some of this transactional debt is still outstanding at the end of the month, it is recorded as an outstanding debt balance and thus an increase in transactional credit card use will artificially increase the measured amount of revolving credit and overstate revolving credit as a percentage of income.



Transactional or "convenience" use of credit cards as a purchasing rather credit medium has been rising over time, both in terms of number of credit card transactions as well as dollar values. During the past fifteen years, convenience use grew by approximately 15% per year, whereas the amount borrowed on credit cards as revolving credit grew only about 6.5% per year.[9] In part, the increase in transactional use of credit cards has been driven by the spread of rewards cards, such as cash-back programs or frequent flyer miles.

The mismeasurement of transactional credit card use as credit card borrowing tends to overstate credit card debt by approximately ten percent, a figure that has doubled in the past decade as a result of the rapid rise of credit card convenience use.[10] The percentage of credit card transactions that are paid off at the end of each month relative to those that end up revolving has risen over time, indicating a growth in convenience use. In addition, the median monthly charge amount for convenience users has risen over four times more rapidly for convenience users than for revolvers. The median monthly charge for convenience users has increased by about $130 (from $233 in 1991 to $363 in 2001), whereas the average charge of revolvers is substantially smaller and has increased more slowly, rising only $30 during that same time period (from $117 to $147). Again, much of this growth in the median size of transactional purchases probably results from a rise in cash-back and cobranding benefits. In addition, because convenience users do not have to pay for their purchases until the end of the billing period *plus* the grace period after receiving their bill, they have the opportunity to take advantage of interest rate "float" during the time between their purchase and payment of the obligation, which may be as long as forty-five to sixty days. During that period, a transactional user essentially receives a free loan from the credit card issuer at zero percent interest,[11] during which time those same funds can be invested in assets that generate a positive return, even if only a money market account or similar safe, short-term investment. In fact, empirical evidence tends to suggest that consumers do exactly this—convenience users tend to carry smaller precautionary balances in their checking accounts than revolvers, suggesting that they are taking advantage of this float. In addition, revolvers are more likely to make use of debit cards than are non-revolvers, which can be explained by the fact that revolvers do not receive the benefit of interest-rate float because they are required to pay the full interest on the account.[12]

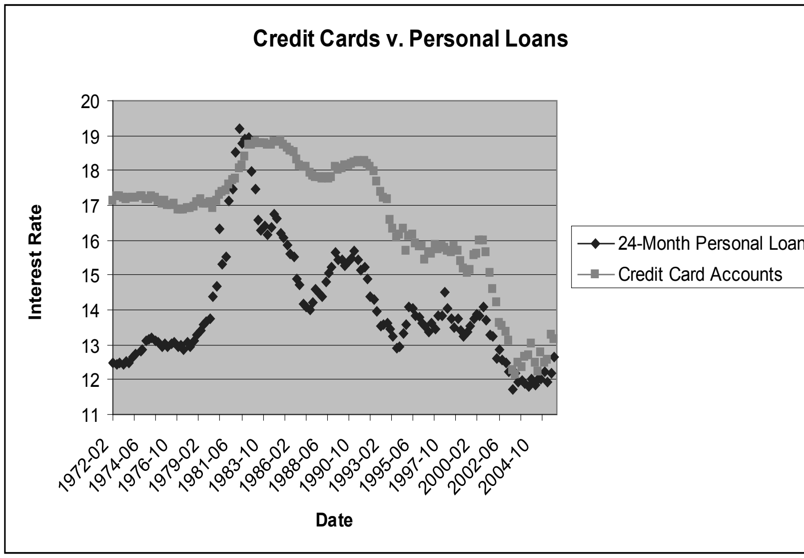
Overall, therefore, there is no evidence that increased use of credit cards has caused consumers as a whole to become overindebted. In fact, the rise in credit card use is the result of a substitution away from other less-attractive forms of credit (because of cost, flexibility, or other drawbacks such as the need to pawn personal goods) to credit cards.

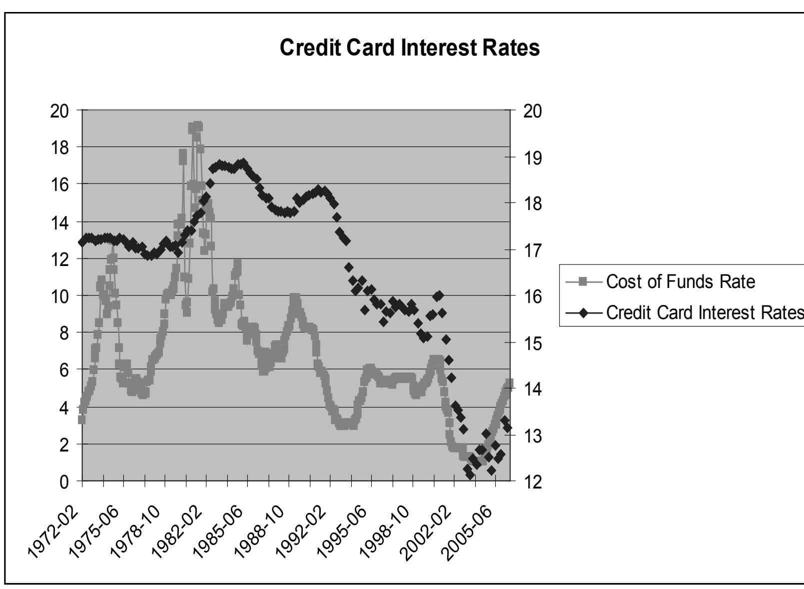
B. "High" Credit Card Interest Rates

Many commentators insist that the growth in credit card use as a source of revolving credit is irrational in light of the "high" interest rates charged on credit cards.[13] But credit card interest rates have fallen substantially over the past fifteen years (see graph above).

Annual fees, which were once a standard component of credit card contracts, virtually disappeared from credit cards during this period, except for those cards that offer frequent flier miles or some other benefit program that requires some administrative activity.[14] This elimination of annual fees, which were in the range of $20-$50 per year, was a massive across-the-board price reduction that not only reduced the cost of credit cards to consumers, but also increased competition in the credit card market by making it easier and less-expensive for consumers to carry multiple cards and to use the cheapest or most appropriate card for any given transaction.

This rapid decline in credit card interest rates explains the substitution from other types of consumer credit. Compare credit cards to the closest alternative to credit card borrowing, the traditional short-term unsecured installment loan, such as from a personal finance company. The following figure (top) displays interest rates on 24-month unsecured installment loans versus credit card interest rates for the past thirty years.





As can be readily observed, the difference between interest rates on short-tem personal installment loans and credit card accounts has narrowed over time. Indeed, in recent years the interest rate on credit card accounts has frequently fallen below that of short-term personal loans. A recent survey of consumer banking rates in the Washington, D.C., area found the prevailing interest rate on credit cards was 8.16%, whereas the prevailing rate for personal loans was 10.45%.[15] Moreover, once up-front initiation fees on personal loans are taken into consideration the overall cost of personal loans is almost certainly higher overall.[16] And this does not even consider the time, inconvenience, and more limited usefulness of a personal finance loan, or the more flexible repayment option of credit cards. According to one survey conducted by the Federal Reserve, 73% of consumers report that the option to revolve balances on their credit cards makes it "easier" to manage their finances versus only 10% who said this made it "more difficult."[17]

This decline in credit card interest rates has resulted from robust competition in the credit card market and savvy shopping by consumers. Survey evidence indicates that consumers who revolve credit card balances are extremely likely to be aware of the interest rate on their credit cards and to comparison shop among cards on that basis, and those who carry larger balances are even more likely to be aware of and comparison shop on this term than those who revolve smaller balances.[18] By contrast, those who do not revolve balances tend to focus on other aspects of credit card contracts, such as whether there is an annual fee, the grace period for payment, or benefits such as frequent flier miles. In fact, consistent with the observation of more aggressive interest rate shopping by revolvers, those who revolve balances are charged *lower* interest rates on average than those who do not.[19]

Empirical evidence indicates that credit card interest rates also generally reflect changes in the riskiness of credit card lending. Thus, when credit card chargeoffs increase, the spread charged between the underlying cost of funds and the interest rate rises.[20]

Furthermore, credit card interest rates have become less "sticky" over time, indicating that technological and risk-scoring innovations as well as more flexible risk-based pricing (as detailed below) has made credit cards even more responsive to competitive pressures. According to the General Accounting Office 93% of the cards they examined in 2005 had variable interest rates—a rise of nine percentage points in just two years.[21] As a result, interest rates on credit cards have become more closely tied to overall interest rates in the economy, as illustrated in the bottom figure.

As can be seen, interest rates on credit cards historically were relatively "sticky," when compared to other types of interest.[22] But note in particular that interest rates on credit cards were *equally sticky* throughout the *entire period* of 1972-1989. The era of the 1970s, of course, was an era of dramatically *increasing* interest rates—essentially the mirror opposite of the falling interest rates of the 1980s. During the period 1972-1982, the federal funds rate rose form a monthly low of 3.29%

in February 1972 to a high of 19.10% in June 1981. Annual averages ranged from 4.43% in 1972, steadily increasing to 16.38% in 1982, before they started falling again. Thus, credit card interest rates were *also* sticky during the 1970s and early-1980s despite a *rising* cost of funds rate. Regardless of whether the cost of funds rate is rising or falling, for a period of twenty years the interest rate on credit cards has remained relatively constant, until the decline in interest rates in recent years. If credit card issuers were reaping large profits off the "spread" between the cost of funds and interest rates in the 1980s, they by definition were suffering equally large losses during the 1970s and the early 1980s. In fact, during this period, the average return on credit card operations was lower than for other sectors of banking activity. So, in general, whether the cost of funds rate has been rising or falling, interest rates on credit cards have been much less responsive to changes in the cost of funds than have other forms of consumer credit.

In recent years, however, credit card interest rates became much more responsive to changes in the cost of funds rate during this period. Beginning with the final quarter of 1994 to the present, the interest rates on credit cards became tied much more closely to the cost of funds rate rose, and for credit card accounts actually assessed interest, the fit is even tighter, again likely reflecting the higher emphasis placed on this term by revolvers when shopping for cards.[23]

On the whole, therefore, there appears to be no evidence of any market failure with respect to interest rates on credit cards. Competition and increasingly sophisticated consumer choice have brought about lower and more responsive interest rates over time. Alternative types of consumer credit offer similar interest rates, but often higher fees and more inconvenience than do credit cards.

C. Fees and Other Price Terms

Interest rates on credit cards have fallen and become more flexible during the past decade, but during that same time period late fees, overdraft fees, and other fees have risen in frequency and amount. These fees remain only a relatively small percentage of issuers' revenues, however, only amounting to about 10% of issuers' revenues, whereas interest payments still amount to about 70% of revenues.[24] The remainder of revenue is generated by merchant discount fees and the like. Moreover, although the GAO was able to find some isolated instances where assessment of these fees imposed an undue hardship on particular consumers, it was unable to find any systematic evidence of categorical abuse or misuse of these fees.

This increased use of penalty fees arose during the same time period that credit card interest rates both became lower and more flexible. This does not appear to be a coincidence. Evidence indicates that, in general, these fees are risk-based fees triggered by actual borrowing behavior and when used in combination with interest rates provides issuers with greater flexibility in pricing credit terms than relying on interest rates alone. Interest rates are generally an *ex ante* before the fact estimate of a given borrower's likelihood of default. Late fees, over-limit fees, and other similar fees, by contrast, are more tightly tied to the borrower's exhibited risky behavior. The only systematic empirical study of these fees of which I am aware concludes that these fees are risk-based and complement interest rates for efficient risk pricing.[25] Massoud, Saunders, and Scholnick find, for example, that a one standard deviation in bankruptcy per capita leads to an increase in penalty fees of $0.62 to $1.31. Similarly, a one standard deviation change in the chargeoff ratio was found to change late fees in a range of $4.35 to $7.57. In addition, they find that a 1 basis point reduction in card interest rates will result in an increase in penalty fees of between 0.88 and 4.11 cents. Thus, in their study, a one standard deviation in credit card interest rates (273 basis points) was estimated to change late fees by $2.40. Moreover, they found no evidence that assessed penalties were larger for low-income borrowers.

The increased use of risk-based fees has occurred at the same time as increased variable-rate pricing on credit cards, as the combination of these two pricing mechanisms is evidently more efficient than interest rates alone. In addition, it appears that consumers who pay these fees are not surprised by their existence, but are aware of them before they enter into the transaction that triggers the fee.[26]

In addition, if credit card penalty fees were actually some sort of new form of consumer abuse, rather than simply a more accurate pricing scheme, then this tradeoff between higher risk-based fees and lower interest rates would result in larger economic rents or "economic profits" to the banking industry. In fact, return on assets has been largely constant for credit card banks over the past two decades, even though there has been a steady rise in the returns of other commercial banks.[27] Thus, during the early days of credit cards, issuers relied heavily on annual fees that were assessed on all cardholders, regardless of risk. During the 1990s, issuers phased out widely-disliked annual fees and moved toward greater emphasis on interest rates that were more closely tied to borrower risk. The gradual increase in the use of risk-based fees to supplement interest rates has made credit pricing reflect risk still further. This suggests that the transition to more risk-based pricing has come about through market competition, resulting in more efficient pricing of credit terms to consumers. First, there was a general phasing out of annual fees and greater emphasis on interest rates, then recent years has seen a gradual increase in the use of penalty fees to further more closely tailor price to cardholder risk.

II. Cost-Benefit Analysis and Unintended Consequences

Available evidence indicates that the credit card market is competitive and responsive to consumer choice. Understanding the economics of the credit card market therefore raises serious challenges for any proposals to heighten regulation of the credit card market. In fact, misguided regulation can have serious unintended consequences that will end up reducing consumer welfare; thus, any proposal for additional regulation should be studied carefully to ensure that the benefits of any such regulation exceed the costs, including any unintended consequences that such regulation is likely to spawn. In addition, it would be wise to examine the continuing relevance and utility of existing regulations before proposing new regulations.

There are three basic manners in which credit can be regulated: substantive regulation, disclosure regulation, or market and common law "regulation." Each has costs and benefits.

A. Substantive Regulation

The oldest and hoariest type of regulation of consumer credit is substantive regulation of credit terms, such as usury restrictions that cap the rate that can be charged on interest rates. Substantive regulation of terms is generally frowned upon today, as thousands of years of economic history have generally demonstrated that the costs of substantive regulation generally exceed any benefits that it would generate.

In particular, there are three predictable unintended consequences that result from substantive regulation of consumer credit terms: (1) term substitution and repricing, (2) product substitution, and (3) rationing. Each of these three would likely manifest themselves in response to efforts to place new regulations on credit cards.

(1) Term Substitution and Re-pricing: Credit card contracts are complicated, multiple-term contracts. Term substitution refers to the phenomenon that regulation of some terms of this multiple-term contract will cause issuers to adjust other terms in order to reach the market clearing "price." Even in the relatively short history of credit cards, history is littered with examples.[28] Prior to the Supreme Court's decision in *Marquette National Bank v. First of Omaha Corp.*,[29] most consumer credit card contracts were governed by usury restrictions that capped the interest rate that could be charged on credit cards. As interest rates generally rose during the 1970s, this rate ceiling meant that card issuers could not charge a market rate of interest on their consumer loans. The era witnessed a number of offsetting term re-pricing adjustments by credit card issuers, all of which almost certainly made consumers worse off. First, issuers imposed annual fees on all cards to make up for the shortfall from the inability to charge a market rate of interest. Not only was this an inefficient pricing mechanism because it wasn't calibrated to borrower risk, it also forced transactional users of credit cards to subsidize revolvers who were able to borrow at the sub-market interest rate. Similarly, retailers would bury their credit losses by marking up the price of the goods they sold on credit; for instance, states with stricter usury ceilings also had higher retail prices for appliances. Usury restrictions also had a number of other unfortunate negative impacts on consumers. Customer benefits were lower in states with stricter usury ceilings, such as shorter banking hours and the elimination of other services such as free Christmas gift wrapping at department stores. Moreover, this term substitution also had the effect of making credit more heterogeneous in nature, making it more difficult and expensive for consumers to compare prices and shop. Most notably, annual fees made it more expensive for cardholders to carry more than one card, thereby making it difficult to switch from one card to another that presented a better deal.

The immediate aftermath of *Marquette* was the opportunity for credit card issuers to charge a market rate of interest for their products. In turn, this led to the rapid elimination of annual fees, which were no longer necessary to offset regulatory caps on interest rates. In turn, this enabled greater competition and consumer choice, which eventually resulted in a fall in a proliferation of card variety, lower interest rates, and heightened competition. According to a study by Thomas Durkin of the Federal Reserve, 90% of consumers report that they are "Very" or "Somewhat Satisfied" with their credit cards.[30] Given the ease of comparison shopping and the wide variety of cards in the marketplace, it should not be surprising that most consumers have found products and issuers with which they are largely satisfied.

Empirical evidence strongly suggests that efforts to place substantive limits on credit card pricing today would likely generate similar offsetting term substitution. As noted, empirical evidence indicates that penalty fees imposed by credit card issuers are generally tied to consumer risk and as a result have an offsetting effect on interest rates. Any regulatory efforts to cap or otherwise regulate late fees, overlimit fees, and the like, would therefore almost certainly lead to increased interest rates for all consumers, or other offsetting adjustments in credit contract terms. It is not readily apparent why regulators would seek to impose a regulatory scheme that forces responsible and less-risky borrowers to pay higher interest rates to subsidize irresponsible and risky borrowers who pay their bills late or exceed their credit limits. This cross-subsidization is especially unfair to low-income but responsible borrowers who would otherwise be lumped into the same interest rate category as these other borrowers. In fact, the GAO Report indicates that at least one credit card issuer is experimenting with a credit card that would eliminate all penalty fees—but in exchange would impose a much higher interest rate (above 30 percent) if the cardholder pays late or otherwise defaults on the terms of the card.[31] Thus, while there appears to be some isolated instances of penalty fees run amuck, blanket regulatory limitations on these fees will likely make credit card pricing less efficient and harm overall consumer welfare.

(2) Product Substitution: Notwithstanding the ability of credit card issuers to readjust uncontrolled terms of the credit card contract to try to price credit efficiently, in some situations the inability to charge efficient risk-based prices will make it impossible to extend credit card credit to some borrowers. Nonetheless, Americans need access to credit to deal with life's surprises, such as the need for unexpected car repairs, medical bills, to furnish a new apartment, or simply for a student to buy an interviewing suit to seek a job. If these individuals are unable to get access to credit cards, experience and empirical evidence indicates that they will turn elsewhere for credit, such as pawn shops, payday lenders, rent-to-own, or even loan sharks.[32] As noted above, there is no evidence that more widespread access to credit cards has worsened household financial condition because this growth in credit has been a substitution from other types of consumer credit.

It is hard to see how a college student or any young American is made better off by being denied a credit card and thus forced to furnish her apartment through a rent-to-own company. Nor is it readily apparent to me how a lower-income family who needs schoolbooks or a clarinet for their child is made better off by being forced to borrow from a payday lender or pawn shop to make ends meet. The young and the poor already have fewer and less-attractive credit options than middle class families—restricting their credit options still further by making it even more difficult for them to get access to attractive credit on competitive terms does not seem to be a plausible way of making their lives better.

(3) Rationing: Finally, if issuers are unable to re-price terms so as to reach a market-clearing price for all consumers, and those consumers are unable to get needed credit from pawn shops, loan sharks, and other less-attractive lenders, the eventual result will be that some Americans will lack access to much-needed credit. This is the well-established finding of thousands of years of economic history, going back at least to Ancient Greece. What of the person who needs access to credit to repair a broken transmission so that he can get to work? In the end, at least some consumers are going to be forced to survive without credit that will allow them to repair their car, buy braces for their children, or Christmas presents for their relatives. Simply wishing that he could have access to credit on terms favored by regulators will not make it so and it is not clear what policy benefit is gained by pretending otherwise.

III. Disclosure Regulation

The drawbacks of substantive regulation of consumer credit terms are well-understood. As a result, it has become increasingly common to mandate certain disclosures, rather than to impose substantive regulations on consumer credit. Evidence suggests that some disclosures, like the requirement of disclosing the APR for credit card loans, has tended to facilitate consumer awareness of competing credit offers and thus to shop for the best deal available.[33]

But as with substantive regulation, there is a trade-off to increased mandatory disclosures. Consumers have limited attention for reading disclosures and issuers have limited space and expense for making disclosures. Thus, mandating some disclosures necessarily makes it more difficult to disclose fully other card terms that some consumers may care more about or may make it more difficult for consumers to find the information that they care about.

For instance, approximately half of American consumers do not revolve a balance on their credit cards. For those consumers, the APR is a completely irrelevant term in shopping for and using a card. And the evidence suggest that in fact transactional users of credit cards pay much less attention to the APR and Finance Charge than do those who revolve balances (and the larger the balance the more attention is paid).[34] Transactors generally care more about other aspects of cards, such as grace periods, benefits (such as car rental insurance or purchase price protection), and any rewards they offer (such as frequent flier miles or cash back). Although requiring disclosure of information of interest rates is certainly useful for those who shop on that basis for the other half of card users who do not revolve balances it is simply unnecessary clutter that makes it more difficult for them to locate the information that they want from a card issuer.

Moreover, experience demonstrates that once disclosures are mandated, they become very difficult to update in light of changing circumstances. This can be a particular problem in rapidly-evolving markets such as the credit card market. For instance, the "Schumer Box" requires disclosure of useless or trivial information such as the amount of the minimum finance charge, which according to the GAO Report, was typically about 50 cents. Other mandatory disclosures, such as the method for computing balances, may be too complicated or of little importance to most consumers in choosing among cards.[35] The GAO Report observes that the outdated structure of the Schumer Box, TILA, and Regulation Z make it difficult to accurately and effectively disclose many of the new terms on credit cards that have been described, rendering such disclosures less helpful than would otherwise be the case.

Nonetheless, trivial, outdated, or irrelevant disclosures are given the same importance as other more important terms, and newly important terms are difficult to disclose at all. For mandatory disclosures to be an effective tool for facilitating consumer choice, rather than a counterproductive distraction and threat of information overload, regulators must be committed to updating them swiftly and regularly in order to keep up with rapid changes in the market and consumer preferences.

Still another problem with the actual practice of disclosure regulation is the apparent effort to use disclosure regulation as a "back door" version of substantive regulation, to try to guide consumers in the "right" direction. Thus, although it is recognized that usury restrictions are counterproductive, it is implicitly assumed that forcing disclosure of the "high" rate of interest will shock consumers into moderating their credit use, along the lines of "If consumers *only knew* how much they were paying in interest, they would borrow less." A related problem is mandating disclosures in order to advance some political or social goal, rather than to facilitate careful and responsible consumer borrowing. Thus, Congress recently mandated the disclosure of the amount of time it would take to pay off a cardholders existing balance assuming that only the minimum payment were made. Federal Reserve economist Thomas Durkin estimates that this disclosure actually will be useful to only 4% of cardholders who state that they actually intend to stop adding new charges to the card and to repay their balance by making only the minimum payment.[36] Although this disclosure effects a very small number of consumers—who could otherwise get the same information simply by calling their credit card issuers—it will necessitate still further expense by cardholders and further increase the costs to consumers of locating the information that they actually care about. Properly implemented, standardized disclosure may facilitate autonomous consumer choice by making it easier for consumers to comparison shop among credit products. But efforts to use disclosure as a back door version of substantive regulation is likely to be ineffective at bringing about the desired substantive outcome, while simultaneously failing to provide the useful information to consumers that disclosure regulation should produce.

Finally, according to another study by Durkin, two-thirds of credit card owners find it "very easy" or "somewhat easy" to find out information about their credit card terms, and only six percent believed that obtaining this information was "very difficult." Two-thirds of respondents also reported that credit card companies usually provide enough information to enable them to use credit cards wisely and 73% stated that the option to revolve balances on their credit card made it "easier" to manage their finances versus only 10% who said this made it "more difficult." Finally, 90% of credit card owners were "Very" or "Somewhat Satisfied" with their credit cards, versus only 5% who were "Somewhat Dissatisfied" and only 1%—or one out of 100—who were "Very Dissatisfied."

In short, consumers seem overwhelmingly satisfied with their credit cards, the information they receive from credit card issuers, and ease with which they can get information about their cards. Credit card issuers appear to have the incentives to provide timely and accurate information to consumers, and by all accounts appear to be doing so.

IV. Market Competition and Common Law as Regulation

It must also be kept in mind that market competition is a form of regulation as well. The credit card market is extremely competitive, with thousands of issuers constantly competing to woo consumers with better offers. Consumers routinely carry as many as four credit cards in their wallets, ready to switch immediately to the card that offers a more attractive package of benefits and terms. In such a market, it is unlikely that oppressive or unfriendly contract terms would last, and in fact this seems to be the case. The GAO Report found, for instance, that only three of the twenty-eight cards that they examined had "universal default" clauses in 2005.[37] The GAO Report also found that between 2003 and 2005 only a minority of credit card issuers used the so-called "double-cycle billing method" of calculating finance charges and even those issuers have eliminated that scheme today.[38] In addition, only 2% of cards charge annual fees, and virtually all of them provide some rewards program in return. In fact, annual fees traditionally have been the cost of credit cards most despised by consumers—in fact, when annual fees were first implemented in the 1970s, consumers cancelled 8% of their credit cards immediately.[39]

In addition, courts have used traditional common law rules and contract remedies to punish fraudulent or deceptive practices by card issuers. This has been quite efficacious in protecting consumers and raises further questions about the need for additional regulation.

Thus, although issuers may try to impose on consumers a variety of disagreeable terms, the ease with which consumers can shift from one card to another, and the heated competition among issuers for consumer loyalty, renders such a scenario relatively implausible. Whether annual fees, universal default clauses, or "double-cycle billing," the market appears to be quite self-correcting in terms of delivering to consumers the credit card products that they desire—which explains the 90% positive satisfaction rate described above.

Endnotes

1 *See* Zywicki, Bankruptcy Law and Policy, ch. 3.

2 Wal-Mart recently announced, for instance, that it was terminating its once-popular layaway program. Like other major department stores, Wal-Mart acknowledged that this form of credit had become irrelevant because of widespread access to credit cards. Unlike layaway, purchasing goods using a credit card permits the consumer to use the goods while paying them off, whereas under layaway the store keeps the goods until they are paid for.

3 *See* Thomas A. Durkin, *Credit Cards: Use and Consumer Attitudes, 1970–2000*, 86 Fed. Res. Bull. 623 (2000).

4 Alan Greenspan, Understanding Household Debt Obligations, Remarks Given at the Credit Union Nat'l Ass'n 2004 Gov'tal Affairs Conf. (Feb. 23, 2004), *available at* http://www.federalreserve.gov/boarddocs/speeches/2004/20040223/default.htm.

5 *Accord* Bd. of Governors of the Fed. Reserve, Rep. to Congress on Practices of the Consumer Credit Industry in Soliciting and Extending Credit and their Effects on Consumer Debt and Insolvency 5 (June 2006).

6 *Id.* at 13.

7 Wendy M. Edelberg & Jonas D. M. Fisher, *Household Debt*, Chi. Fed. Letter, Nov. 1997, at 1, 3 (1997); *see also id.* at 4 ("[I]ncreases in credit card debt service of lower-income households have been offset to a large extent by reductions in the servicing of installment debt."); Arthur B. Kennickell et al., *Family Finances in the U.S.: Recent Evidence from the Survey of Consumer Finances*, 83 Fed. Res. Bull. 17 (1997) (noting that the share of families using installment borrowing fell between 1989 and 1995 as a result of increased use of mortgages, credit cards, and automobile leasing); Glenn B. Canner & James T. Fergus, *The Economic Effects of Proposed Ceilings on Credit Card Interest Rates*, 73 Fed. Res. Bull. 1, 4 (1987) (noting that rise in credit card use may have been the result of "a substitution of credit card borrowing for other types of installment credit that do not provide flexible repayment terms").

8 *See* Thomas A. Durkin, Comment, *in* The Impact of Public Policy on Consumer Credit 36, 40 (Thomas A. Durkin & Michael E. Staten eds., 2002).

9 Kathleen W. Johnson, *Convenience or Necessity? Understanding the Recent Rise in Credit Card Debt*, Finance and Economics Discussion Series, Fed. Reserve Bd., 2004-47.

10 *See id.*

11 Technically the interest rate is slightly negative because of the time value of money.

12 Jonathan Zinman, *Why Use Debit Instead of Credit? Consumer Choice in a Trillion Dollar Market*. Brown and Plache find that 62 percent of revolvers who acquired a general purpose debit card actually used that card whereas only 37 percent of nonrevolvers used their debit card. *See* Tom Brown & Lacy Plache, *Paying with Plastic: Maybe Not So Crazy*, 73 U. Chicago L. Rev. 63, 84 (2006).

13 Note that if the interest rates really were higher on credit cards than on the types of credit that they supplant, then one would expect this to be reflected in a higher debt-service ratio, which as we have just seen, it is not.

14 U.S. Gen. Accounting Office, Credit Cards: Increased Complexity in Rates and Fees Heightens Need for More Effective Disclosures to Consumers 23 (2006). The GAO Report noted that some cards offered rewards but still did not charge annual fees.

15 The *Washington Times* reports area consumer banking rates each Friday. Data is drawn from those published reports.

16 Brito and Hartley reported, for instance, "A senior bank officer told us that the costs to the bank of processing a loan are so high that they cannot afford to make a loan of less than $3,000 for one year except at interest rates *above* those charged on credit cards." They also note, "inquiries in Houston in February 1992 revealed rates ranging from 17 percent and a $100 fixed fee for a collateralized 1-year loan at a branch of a major national finance company to over 50 percent for small loans ($300 maximum) at a local finance company." In short, bank loans of similar size and duration "either do not exist or are available only at terms more onerous than those offered by credit card issuers." By contrast, credit cards generally require no application fee and no minimum loan size. *See* Dagobert L. Brito & Peter R. Hartley, *Consumer Rationality and Credit Cards*, 103 J. Pol. Econ. 400, 402 (1995).

17 Durkin, *supra* note 3, at 623.

18 *See* Thomas A. Durkin, *Credit Card Disclosures, Solicitations, and Privacy Notices: Survey Results of Consumer Knowledge and Behavior*, Federal Reserve Bulletin, at A109 (2006).

19 Tom Brown & Lacey Plache, *Paying with Plastic: Maybe Not So Crazy*, 73 U. Chicago L. Rev. 63 (2006).

20 *See* Adam B. Ashcraft, et al., *The Bankruptcy Abuse Prevention and Consumer Protection Act: Means-Testing or Mean Spirited?* Working Paper, Federal Reserve Bank New York (Dec. 19, 2006).

21 GAO Rep., *supra* note 14, at 15.

22 An extended discussion of the explanation for the traditional stickiness of credit card interest rates is provided in Todd J. Zywicki, *The Economics of Credit Cards*, 3 Chapman L. Rev. 79 (2000).

23 *See* Kathleen Johnson, *recent Developments in the Credit Card Market and the Financial Obligations Ratio*, Fed. Res. Bull. 473, 477 (Autumn 2005) (noting that correlation between credit card interest rates and the prime rate was only 0.09 during the 1980s and early 1990s but has risen to 0.90 from mid-1990s to present).

24 GAO Rep., *supra* note 14, at 70-72.

25 *See* Nadia Massoud, et al., *The Cost of Being Late: The Case of Credit Card Penalty Fees*, working paper (January 2006).

26 *Supra* note 18, at A114.

27 *See* GAO Rep., *supra* note 14, at 76. For a discussion of the special difficulties in inferring credit card "profits" from the standard analysis of "return on assets" used in the banking industry, *see supra* note 22.

28 *See supra* note 22 for an extended discussion.

29 439 U.S. 299 (1978).

30 *Supra* note 18.

31 GAO Report, *supra* note 14, at 24.

32 *See* Susan Lorde Martin & Nancy White Huckins, *Consumer Advocates v. The Rent-to-Own Industry: Reaching a Reasonable Accommodation*, 34 Am. Bus. L.J. 385 (1997); Signe-Mary McKernan et al., *Empirical Evidence on the Determinants of Rent-to-Own Use and Purchase Behavior*, 17 Econ. Dev. Q. 33, 51 (2003); James P. Nehf, *Effective Regulation of Rent-to-Own Contracts*, 52 Ohio St. L.J. 751, 752 (1991); Eligio Pimentel, *Renting-To-Own: Exploitation or Market Efficiency?*, 13 Law & Ineq. J. 369, 394 (1995); Lendol Calder, Financing the American Dream; John P. Caskey, Fringe Banking: Check-Cashing Outlets, Pawnshops, and the Poor 37-67 (1994; Richard L. Peterson & Gregory A. Falls, Impact of a Ten Percent Usury Ceiling: Empirical Evidence (Credit Research Ctr., Working Paper No. 40, 1981); *see also* Robert W. Johnson & Dixie P. Johnson, *Pawnbroking in the U.S.: A Profile of Customers* 47 (Credit Research Ctr., Monograph No. 34, 1998)

33 *See supra* note 3.

34 *Id*. at A113.

35 GAO Report, *supra* note 14, at 54.

36 *Supra* note 3, at 623, 634.

37 GAO Report, *supra* note 14, at 26.

38 *Id*. at 28.

39 *See supra* note 22.

Treasury's Blueprint: Regulatory Efficiency or More Red Tape?

*By John Shu**

On March 31, 2008, the U.S. Department of the Treasury issued its "Blueprint for a Modernized Financial Regulatory Structure," the largest proposed revamping of federal financial regulation and oversight since the Great Depression.[1] Although released in the middle of a national financial crisis, the Blueprint is not designed to be a quick fix for the current economic situation. Instead, it proposes significant changes that will greatly affect financial institutions—and the entities with which they conduct business—for many years to come.

The Treasury Department developed its proposal over the past year, with some of the ideas already existing for many years and/or previously proposed. Divided into short, intermediate, and long-term plans, the Blueprint proposes to expand the Fed's responsibilities and to streamline the regulatory plan for depository institutions, securities firms, hedge funds, mortgage originators, and the insurance industry.[2] But despite this stated intent to streamline the regulatory framework, the Treasury's proposals call for the creation of several new regulatory agencies, such as "Mortgage Origination Commission," "Office of National Insurance," "Office of Insurance Oversight," "Mortgage Stability Council," "Federal Insurance Guarantee Corporation" (with a "Federal Insurance Guarantee Fund"), "Prudential Financial Regulatory Agency," and a "Conduct of Business Regulatory Agency."[3]

Mortgages

While practically every American is painfully aware of the current mortgage and housing sector difficulties, the Treasury Department's proposal is not specifically designed to address the nation's immediate real estate or financial problems. One may wonder, then, why the Department of the Treasury is getting involved with mortgages. Unlike 1933, mortgages today are deeply intertwined not only with Wall Street, but also the global economy. In fact, investors exposed to subprime mortgages that do not comply with state and federal consumer protection laws face the risk that the mortgages supporting some of their investments may not be enforceable, which would lead to extensive litigation. Thus, the Federal Reserve has long-held rulemaking power regarding the Truth in Lending Act, including the Home Ownership and Equity Protection Act.[4]

As an important part of its short-term planning, the Treasury Department, recognizing the importance of mortgages to the national and global economies, recommends creating a Mortgage Origination Commission (MOC) to "evaluate, rate and report on the adequacy of each state's system for licensing and regulating participants" (e.g., brokers and lenders) in the mortgage origination process. Presumably this would include creating licensing standards for state-regulated mortgage companies and oversee how states oversee mortgage origination. The MOC's board would have representatives from the Federal Reserve, the Office of the Comptroller of the Currency (OCC), Office of the Thrift Supervision (OTS), the Federal Deposit Insurance Corporation (FDIC), the National Credit Union Administration (NCUA), and the Conference of State Bank Supervisors (CSBS). Because each state currently sets its own minimum qualification standards, a certain number of mortgage brokers and lenders do not fall under federal regulation and instead are subject to varying degrees of state oversight.[5] These state-regulated entities are responsible for more than half of the subprime mortgages in the United States. The MOC would not replace the states' regulation of mortgage origination but, rather, would add an additional regulatory layer and provide the marketplace with information regarding each state's mortgage compliance standards.[6] The MOC, similar to the OTS, would be funded through assessments on mortgage originators that would be required to register through the National Mortgage Licensing System & Registry.

One problem with the Treasury Department's proposal is that the state agencies would certainly balk at having to operate under a federal minimum standard, instead of having the autonomy to create their own. Also, a substantial number of the troublesome subprime mortgages issued from federally regulated entities, not state-regulated ones.[7] Additionally, because of regulatory fragmentation, it remains unclear which regulatory agency would have oversight and enforcement of mortgage originators who are independent participants (i.e., unaffiliated with depository institutions), and those which are affiliates, not subsidiaries, or depository institutions. The Blueprint notes that state supervisory agencies must be given the clear authority to enforce federal mortgage lending standards, along with the appropriate federal regulator, and calls for federal legislation either to set uniform minimum licensing qualification standards for state mortgage market participant licensing systems or to give the MOC the power to develop and implement them.

The Federal Reserve

The Blueprint not only advocates that the Fed continue to have rulemaking authority over the Truth in Lending Act and the Homeowners Equity Protection Act, it proposes expanding the Fed's power and responsibilities. For the short term, the Treasury Department suggests clarifying and enhancing liquidity provisioning by the Federal Reserve and allowing the Fed to conduct on-site examinations of non-depository institutions under certain circumstances, presumably in conjunction with the U.S. Securities & Exchange Commission (SEC) and/or the U.S. Commodity Futures Trading Commission (CFTC). This is partially an acknowledgement of what the Fed has already done this year, e.g., opening up the discount window to non-depository institutions, establishing the Term Securities Lending Facility, and facilitating JPMorganChase's buyout of Bear Stearns.

For the intermediate and long-term, the Treasury Department envisions the Fed acting as a "market stability regulator," with broad authority to monitor risks and take

** John Shu is an attorney at Stradling Yocca Carlson & Rauth in Newport Beach (CA). He served as a law clerk to the Hon. Paul H. Roney, U.S. Court of Appeals for the Eleventh Circuit.*

"corrective action" under certain circumstances across the financial system, including investment banks, hedge funds and commodity operators—not just the financial holding companies, bank holding companies, and certain chartered banks it already monitors. In order to do so, the Fed presumably would want to be able to take "corrective action" sooner rather than later, and would be required to evaluate the capital, liquidity, and margin practices across the entire financial system, (not just with depository institutions), as well as to analyze the potential impact on overall financial stability. The Blueprint, however, does not define at what point the Fed would be able to step in and take "corrective action." Part of this vision, however, is the creation of a "Market Stability Council" which would serve as a check on the Fed, and would have the Fed oversee state-chartered banks and payment and settlement systems.[8]

Skeptics argue that the Fed already failed to properly regulate those banks already within its jurisdiction. A number of banks have had to write off billions of dollars in subprime mortgage losses. There is some question whether the Fed could handle expanded regulatory responsibilities (although the Fed, unlike federal agencies such as the SEC, does not depend on Congress for funding staffing increases). Further, the Treasury Department's proposal does not make clear what authority the Fed would have to intervene or enforce existing regulations. For example, if the Fed were to intervene only in times of "extreme market stress," then Congress or perhaps the Fed's Board of Governors would have to define the term and provide clear guidance as to not only the insertion point(s) but also the procedures that the Fed would take once it determined that it had to take action.

Insurance

Insurance, like banking, is a sector which permeates every aspect of American life. The subprime fallout adversely affected insurers. For example, bond insurers such as Ambac and MBIA suffered heavy losses as a result of their forays into the Collateralized Debt Obligation (CDO) world, affecting municipalities and investors across the country.

Recognizing this, the Treasury Department proposed an intermediate-term plan to establish a federal insurance regulatory structure, presumably with federal preemptive powers. Traditionally, insurance regulation is left to the states. As a result of the insurance industry's increasing national and international focus, the Treasury Department believes that the current state-based regulatory framework is too cumbersome for efficiently developing nationwide products and competing abroad.

The Treasury Department's proposed federal structure would require creating an optional federal charter similar to the current dual-charter banking system and would require a new regulator: the Office of National Insurance. There is some question as to how strict the federal regulatory structure would be. There is also some question as to whether it is appropriate to, in effect, nationalize a structure which traditionally has been within the role of the several states. The Blueprint also urges Congress to immediately establish an Office of Insurance Oversight (OIO) within the Department of the Treasury. The OIO would address international regulatory issues such as reinsurance and serve as the lead regulatory voice in promoting American international insurance regulatory policies and ensuring that the state insurance regulators achieve a uniform implementation of declared U.S. international insurance policy goals.

Agency Streamlining

Perhaps the most widely reported feature of the Blueprint is the Treasury Department's proposal to merge the CFTC with the SEC and the OTS with the OCC. These proposed mergers are part of the Blueprint's intermediate-term plan. Futures trading is no longer limited to agricultural commodities and the Treasury Department believes that product and market participant convergence, market linkages, and globalization warrant unifying the CFTC and the SEC. To maintain some of the CFTC's charter, the Treasury Department recommends that the SEC adopt core principles for exchanges and clearing agencies, expedite the SRO rule approval process, and provide a general exception under the Investment Company Act of 1940 for already actively traded exempted products such as Exchange-Traded Funds.[9]

Harmonizing the two agencies' regulatory philosophies and the many differences between securities and futures regulation will be a daunting task. The SEC and CFTC have separate, and sometimes disparate, rules regarding numerous issues, such as margin, segregation, insider trading, insurance coverage for broker-dealer insolvency, customer suitability, short sales, SRO mergers, implied private rights of action, portfolio managing, and the SRO rulemaking approval process. For example, "margin" in the securities context refers to the minimum amount of equity that must be put down in order to purchase securities on credit, whereas in the futures context "margin" means a risk-based performance bond system which acts like a security deposit.

The suggested SEC–CFTC merger is not new, and has drawn opposition from various trade associations, the CFTC, the states, and within Congress. For example, commodities, due to their agricultural history, fall under the jurisdiction of the House and Senate Agricultural Committees. Those committees would be reluctant to cede oversight power should the CFTC and SEC merge. Notably, the futures markets have always been known as more of a "Wild West" environment than the securities markets, and the SEC has fallen under repeated criticism for its failures to catch and stop behaviors ranging from stock option backdating to Enron to the current subprime fiasco. Critics point out that a combined agency which would be SEC-heavy would fail to properly regulate the already chaotic futures markets. Furthermore, state regulators, whose "blue-sky" laws regulating securities have already been somewhat eclipsed by the SEC and CFTC, may justifiably view this proposed consolidation as a further infringement of their regulatory authority.

Prudential Financial Regulatory Agency

For the long-term, the Treasury Department envisions a single "Prudential Financial Regulatory Agency" to focus on financial institutions with some type of explicit government guarantee associated with their business operations, capital adequacy requirements, investment limits, activity limits, and direct on-site risk management supervision. As an intermediate step to that ultimate vision, the Treasury Department proposed combining the Office of Thrift Supervision (the primary federal regulator of saving associations and savings and loan companies) with the Office of the Comptroller of the Currency (the national bank regulator). Part of the Treasury Department's reasoning is that U.S. consumers have sufficient access to residential mortgage loans and that thrifts and banks are nearly indistinguishable in the modern day. The proposed combined agency would oversee the safety and soundness of firms with federal guarantees, while having authority to deal with affiliate relationship issues.

Industry groups and the OTS have screamed foul. In fact, the OTS has suggested that its authority ought to be *expanded* in order to provide nationwide uniformity in regulating mortgage bankers and mortgage brokers. The OTS-OCC merger proposal, like the proposed SEC-CFTC merger, is also an old idea. Thrifts have historically focused more on mortgage lending, but they also offer various financial products. OTS's budget comes from assessments it levies on the more than 800 savings and loans it supervises. In a direct challenge to Treasury's proposal, OTS has suggested that it its role be expanded to include supervisory powers over mortgage brokers, a task for which Treasury proposed creating the Mortgage Origination Commission.

Business Conduct Regulatory Agency

For the long-term, the Treasury Department ultimately envisions combining a number of agencies, including the SEC, CFTC, and perhaps even the Federal Trade Commission (FTC) into a single "business conduct regulator" to protect consumers and investors. This agency would presumably have authority over disclosures, rule writing, business practices, and chartering/licensing. The future agency would subsume most of the roles of the SEC and CFTC, as well as the consumer protection and enforcement roles of insurance and banking regulators, with authority over mortgage disclosures. It is unclear as to whether Treasury envisions the SEC, CFTC, and other agencies to still exist, or to which agencies the non-subsumed duties would go.

CONCLUSION

It is interesting that the Treasury Department, and thus Secretary Henry M. Paulson, is the public face of the proposed changes, rather than the President. The Administration was bold in revealing the Blueprint, most of which, in principle, would help government regulators keep up with rapidly changing financial innovation. The devil, however, is in the details. Beyond the political difficulties of turning any of the Blueprint's proposals into law, many are skeptical that a new regulatory regime would be any better than the current one in terms of protecting consumers and the overall market. If anything, the nation's current political mood appears to want more government regulation, rather than less. There is also a structural disconnect between the nationalization, and globalization, of financial products and our federalist system of government. Individual state regulators and state legislatures may decide to flex their own political muscles in trying to maintain or even expand their powers and responsibilities.[10]

The irony of the Treasury Department's Blueprint is that it does not appear to streamline the regulatory environment all that much. Even if the SEC/CFTC and OTS/OCC mergers occurred, the Blueprint still proposes at least seven new entities, and would require multiple new charters. As noted earlier, the Blueprint also fails to adequately address several procedural and enforcement questions.

Very little in the Blueprint can happen without congressional action in the forms of legislation and appropriations. Interested or potentially affected parties already have attacked the Blueprint. Additionally, the political realities of an administration in its last months, combined with a Congress which is unlikely to be able to take any substantive action regarding the Blueprint in a presidential election year, means that, at minimum, the current regulatory structure will survive well into the latter half of 2009. Further, global economic difficulties related to the various subprime and exotic mortgage-backed instruments have led many observers to believe that more, not less, regulation and enforcement could have prevented those difficulties. While politically the Blueprint has forced Congress to take some action (e.g., holding hearings), Congress has often demonstrated a real lack of courage and thoughtfulness in tackling pressing issues. Should Congress pass new legislation as a result of the Blueprint, the affected administrative agencies would still have to engage in extensive rulemaking to satisfy the attendant administrative law requirements and to fill in any legislative gaps. Even without any new legislation, it is reasonably likely that certain agencies, the Fed, or even the several states will take preemptive steps to either maintain or expand their power and jurisdiction. Financial sector companies would be wise to monitor Congress and the relevant agencies to see what storm might be coming. Regardless of the final form of the legislation, there will certainly be significant debate, and litigation, in this arena for years to come.

Endnotes

1 The Blueprint is available at http://www.treas.gov/press/releases/reports/Blueprint.pdf.

2 The current financial regulatory structure includes five federal depository institution regulators, state-based regulators, separate federal securities and federal futures regulators, self-regulatory organizations, and state-based insurance regulators.

3 The Treasury proposal leaves in place the many Self-Regulatory Organizations (SRO's) and, in certain cases, enhances their powers.

4 The Fed's authority exists independent of Congress, which is currently debating bills regarding mortgage standards. The Fed has been sharply criticized for its perceived failures to better regulate the mortgage lending industry. Approximately four months ago, the Fed proposed new standards for exotic mortgages and for high-fee or high-cost loans to borrowers with lower credit scores. The Fed's proposal would require mortgage lenders to disclose hidden fees and mortgage companies to show that their customers could realistically afford their mortgages, and prohibit misleading advertising. The Fed was planning to issue final rules sometime this summer, but the Fed now appears to be leaning towards limiting the scope of its proposed rules in response to mortgage industry comments. Consumer protection groups,

Congress, and even some agencies such as the FDIC have publicly expressed their concerns about the Fed's seeming response to the various mortgage, home builders, and realty trade associations.

5 Only seven states actively participate in the National Mortgage Licensing System & Registry (NMLSR), which provides information regarding mortgage market participants' background, expertise and disciplinary history.

6 In proposing the Mortgage Origination Commission, Treasury borrowed certain aspects from the mission of the Federal Financial Institutions Examination Council (FFIEC), a formal interagency body created in 1979 which prescribes uniform principles, standards, and report forms for the federal examination of financial institutions and makes recommendations to promote supervisory uniformity. The FFIEC facilitates public access to data that depository institutions must disclose under the Home Mortgage Disclosure Act of 1975.

7 The Blueprint also does very little, if anything, to address the current and upcoming waves of mortgage and/or subprime-related litigation. Already there are a significant number of lawsuits, including large class actions, against almost every type of entity which is part of the mortgage lending process or residential real estate, including the investment banks. The litigation claims are diverse, but include categories such as inadequate disclosure, fraud and securities fraud, commercial contract, and bankruptcy cases.

8 Payment and settlement systems are the mechanisms used to transfer funds and financial instruments between financial institutions, and between financial institutions and their customers to discharge certain obligations. On a typical business day, U.S. payment and settlement systems settle transactions with a value of more than $13 trillion.

9 An Exchange-Traded Fund (ETF) is an investment vehicle traded on stock exchanges, much like stocks or bonds. An ETF represents a collection or "basket" of assets such as stocks, bonds, or futures.

10 In certain areas, federal regulatory requirements tend to be less restrictive than those of certain states, *e.g.,* consumer privacy.

Free Speech & Election Law

The Uncertain Future of the Presidential Public Financing System

*By Audrey Perry**

The future viability of the once robust presidential public funding system is now questionable. First implemented in 1976, the system has helped presidential candidates focus less on fundraising and more on campaigning, and given major-party candidates enough money in the general election to finance their campaigns.[1] In the 2008 election cycle, however, only one major candidate, Senator John Edwards, received public funding during the primary election.[2] Furthermore, it is possible that at least one major-party nominee will not accept the general election public funds in 2008—for the first time since the system began in 1976.

The system has seen its share of controversy this election cycle. In February 2007, Senator Barack Obama requested advice from the Federal Election Commission (FEC) about public funding in the general election.[3] Obama wanted to raise private funds at the time, but still preserve the option of taking public funding for the general election.[4] In its request for an advisory opinion, Obama's campaign stated that "should both major party nominees elect to receive public funding, this would preserve the public financing system, now in danger of collapse, and facilitate the conduct of campaigns freed from any dependence on private fundraising."[5]

Since then, however, Senator Obama has backed away from his commitment to the public funding system.[6] Even though he will, if eventually nominated by the Democratic Party, face in Senator John McCain an opponent who has repeatedly expressed his desire to accept public funds in the general election, Obama has recently called the system "creaky" and said that he may not accept federal funds.[7]

The currently non-functioning FEC further complicates the issue. At the moment, the Commission has only two of six Commissioners seated. It takes four Commissioners' votes to approve a public funding request. The remaining Commissioners' nominations have been held up in a Senate deadlock for months, with no end in sight. So, even if the presidential party nominees decide to take public funding for the general election, the FEC may be unable to certify the candidates' entitlement to the U.S. Treasury funds (raising the question of whether a court order or unusual Treasury Department action would be required to release the funds).

History of the System

The presidential public funding system was put into place as part of the sweeping election reform adopted in response to the Watergate scandal. The public funding portion of the Federal Election Campaign Act (FECA) was designed to reduce corruption in the political process by providing public funds to candidates to diminish candidates' reliance on raising private funds.

Since the 1976 election, presidential elections have been financed to some extent by public money, which is provided by voluntary tax check-off.[8] The tax check-off was meant to build the public funding system on a mass base of small donors. However, participation in the program has been waning—in 1980, the high point of the tax check-off was 28.7%, and in 2004 the check-off rate was 9.2%[9]

In 2000, George W. Bush became the first presidential candidate win a major-party nomination without taking matching public funding in the primary election.[10] In 2004, George Bush, John Kerry, and Howard Dean all did not accept public funds in the primary election.[11] 2008 may be the first year that a major party candidate does not accept public funding for the general election.

Primary Matching Funds

The primary matching funds portion of the public funding scheme has been less than useful to viable candidates in recent years. In fact, no eventual major party presidential nominee has received primary matching funds since Al Gore in 2000.[12]

To qualify for primary matching funds, a presidential candidate must establish that he or she is a serious candidate with broad base support by raising more than $5,000 in at least twenty states.[13]

Furthermore, if they receive primary matching funds, presidential candidates must agree to limit campaign spending. In 2008, for example, candidates are allowed to spend about $42 million. Their campaign spending in each state was limited to a specified amount based on the number of voting-age individuals in the state.[14] Candidates must also contribute no more than $50,000 to their own campaign and agree to a post-campaign financial audit by the FEC.

If a candidate meets these eligibility requirements, the federal government will match $250 of an individual's total contributions to the candidate. Only contributions received after January 1 of the year before the election will be matched. The first matching fund payments are not made until at least January 1 of the election year, and made monthly thereafter.

The increasingly frontloaded timing of state primaries and caucuses has contributed to the primary matching fund program's increasing irrelevancy. Frontloading has made the spending limits unrealistic by making the campaign longer and increasing the length of the primary period. In the 2008 election cycle, a candidate who accepted primary matching funds would not have been eligible to receive any such funds until a few days before the Iowa caucus. As of May 1, 2008, Republican candidates had spent over $250 million in the primary; the Democratic candidates have far surpassed that and are still spending.[15] The primary contribution spending limit is clearly inadequate for waging the sort of primary campaigns that have become standard in recent cycles.

....

** Audrey Perry is currently an attorney for John McCain's presidential campaign.*

Furthermore, the frontloaded 2008 primary season made the acceptance of primary matching funds impractical for serious candidates because the eventual major party nominee could have been chosen (and was in the case of the Republicans) by early February. If, for example, Senator McCain had accepted primary matching funds and the attached spending limits, he would have been crippled against the Democratic nominee by the spending caps which would have been in force until the September Republican Convention.

Party Convention Funding

The presidential campaign public financing system also provides national party committees with an optional public grant to pay for their nominating convention costs. If a party committee accepts public funds, it may not spend more than the amount provided, which is indexed for inflation. This year, both the Democratic and Republican Party have already applied for and received the full $16 million in available public money to fund their conventions.[16]

In addition to the public funds national parties may spend directly on their conventions, the FEC allows cities that host conventions to establish non-profit "host committees."[17] That committee may raise money to spend on convention-related projects that might otherwise have been sponsored by the national party with public funds. For example, a host committee may spend money to welcome convention attendees to the city with information booths or receptions, facilitate commerce, furnish a venue for the party's use, and offer construction and convention-related services, etc.[18]

FEC regulations regarding host committee fundraising are decidedly more relaxed than those for other committees. A host committee may receive unlimited donations from corporations, banks, labor unions, and individuals for permissible expenditures.[19] Since 2003, there has been no requirement that the organizations contributing to the host committee have a local presence in the host city.[20] The use of host committees has been criticized by some because the committees may accept unlimited money from corporations and other normally prohibited sources. Critics argue that this allows for the use of "soft money" for nominating conventions, and thus in effect circumvents the Bipartisan Campaign Reform Act (BCRA).[21]

In 2004, the New York Host Committee (for the Republican Convention) received $85.7 million and the Boston Host Committee (for the Democratic Convention) received $56.8 million[22]—which combined was more than four times the amount the national parties received in public funds.[23] Although presidential candidates may not "establish, finance, maintain, or control" host committees,[24] they may make a "general solicitation of funds" for these committees.[25]

General Election Funding

Major party presidential nominees are eligible for a public grant for the general election. In 2008, the Democratic and Republican nominees will be eligible to receive $84.1 million in public funds once their parties officially nominate them at their respective conventions.[26] Candidates may accept general election funding even if they turned down primary matching funds. In order to be eligible for general election public funding, a candidate must limit spending to the amount of the grant and may not accept private contributions for the campaign. Candidates are also not allowed to spend more than $50,000 of personal funds in the general election and must permit an FEC audit after the election.

The FEC allows candidates who accept general election public funds to establish a special account exclusively to pay for legal and accounting expenses associated with complying with campaign finance law. Expenses of this fund, commonly known as "GELAC," are not subject to normal expenditure limits, and the legal fund may accept private contributions within the federal limits and source prohibitions.[27] Candidates may fund GELAC accounts by re-designating primary-election contributions[28] or soliciting GELAC contributions.[29] Past major-party nominees have raised significant amounts in GELAC contributions. In 2004, George W. Bush raised $18.8 million and John Kerry raised $11.9 million.[30]

GELAC funds may be used to pay for many of a campaigns administrative, legal, and accounting expenses. For example, "payroll, overhead[,] ... computer services" and other costs associated with "legal and accounting services provided solely to ensure compliance with" federal campaign laws and regulations;[31] non-reimbursable costs incurred in providing transportation services for the Secret Service and national security staff;[32] costs associated with voting recounts;[33] and winding down expenses for legal and accounting compliance activities. A candidate may not solicit GELAC funds before April 1 of the election year, and GELAC contributions must be returned if a candidate does not accept the general election grant.[34]

Since the 1976 election, every major party nominee has accepted the general election grant. This year, however, could be an exception. As discussed above, Senator Obama has wavered on his commitment to accept public funding in the general election. Senator McCain, however, seems likely to take public funding, although he has said he will reconsider if his opponent does not accept public funds.[35] At this point, Senator McCain has not been raising general election funds, and has publicly expressed a desire to opt into the public financing system. It makes sense for Senator McCain to accept public funds this cycle.[36] The Republican National Convention is late this year, not concluding until September 4, 2008. Senator McCain would have just over two months to spend the $84.1 million public grant. Furthermore, Senator McCain's well known support of the public funding system and commitment to clean elections and campaign finance make it a natural fit for him to accept public funding in the general election.

CONCLUSION

Participation in the presidential public financing system has been declining in recent years, as the system has received increased attention—with periodic calls for either the program's abolition or reform. Several portions of the system are outdated, which has generally eroded the system's value. The way the presidential public financing system functions during the remainder of the 2008 election cycle could determine whether the program is revived or abandoned.

Endnotes

1 From 1976 through 2004, of eight-two candidates generally considered as "serious" (based on media standards, name recognition, etc.) only eight did not participate in the primary matching fund system. These eight were John Connally (1980), Ross Perot (1992), Steve Forbes (1996), Steve Forbes and George W. Bush (2000), and Howard Dean, John Kerry, and George W. Bush (2004). U.S. Congressional Research Service, *The Presidential Election Campaign Fund and Tax Checkoff: Background and Current Issues* (RL32786; Dec. 12, 2005), by Joseph E. Cantor at 5.

2 Although John McCain applied for and was approved for the receipt of primary matching funds, he determined not to accept primary matching funds. *See* Press Release, Fed. Elec. Comm'n, *FEC Approves Matching Funds for 2008 Candidates* (Dec. 20, 2007) *available at* http://www.fec.gov/press/press2007/20071207cert.shtml.

3 *See* Fed. Elec. Comm'n, Advisory Opinion 2007-03 (2007), *available at* http://saos.nictusa.com/saos/searchao?SUBMIT=ao&AO=604&START=576396.pdf.

4 *Id.*

5 *Id.*

6 *See* Jim Kuhnhenn, *Obama Says Public Financing System is 'Creaky' and Needs Reform*, Associated Press, April 11, 2008; *A Lapsed Principle*, Washington Post, April 14, 2008, at A14; *Creaky v. Cling: Public Financing of Elections,* St. Louis Post-Dispatch, April 16, 2008.

7 *See* Kuhnhenn, *supra note 6.*

8 U.S. Congressional Research Service. The Presidential Election Campaign Fund and Tax Checkoff: Background and Current Issues (RL32786; Dec. 12, 2005), by Joseph E. Cantor at 1.

9 *Id.* at 3.

10 Anthony Corrado, et al., The New Campaign Finance Sourcebook 181 (The Brookings Institution 2005).

11 *Id.*

12 *Id.* at 189.

13 *See* 11 C.F.R. § 9033.2(b)

14 *See* Fed. Elec. Comm'n,, *Presidential Spending Limits for 2008*, *available at* http://www.fec.gov/pages/brochures/pubfund_limits_2008.shtml.

15 *See* Center for Responsive Politics, Banking on Becoming President (2008), *available at* http://www.opensecrets.org/pres08/index.php.

16 Press Release, Fed. Elec. Comm'n, Both Major Parties to Receive Public Funding for 2008 Conventions (June 26, 2008) *available at* http://www.fec.gov/press/press2007/20070626conventions.shtml.

17 *See* 11 C.F.R. § 9008.50(b).

18 11 C.F.R. § 9008.52(b).

19 *See* Campaign Finance Institute, The $100 Million Exemption: Soft Money and the 2004 National Party Conventions 30-35 (2004), *available at* http://www.cfinst.org/legacy/eGuide/PartyConventions/financing/pdf/full_partyconventions.pdf (listing major donors to the 2004 New York Host Committee).

20 Corrado et al., *supra* note 10, at 191-92.

21 *Id.* at 192-93.

22 Fed. Elec. Comm'n, Report of the Audit Division on the New York City Host Committee 1 (2005), *available at* http://www.fec.gov/audits/2004/20060112_ny_host_cmt_2004.pdf; Fed. Elec. Comm'n, Report of the Audit Division on the Boston Host Committee 1 (2005), *available at* http://www.fec.gov/audits/2004/20050531host.pdf.

23 Corrado et al., *supra* note 10, at 192.

24 Fed. Elec. Comm'n, Advisory Opinion 2003-12 (2003)*, available at* http://saos.nictusa.com/aodocs/2003-12.pdf (stating that the "general solicitation" exception to the ban on "soft money" solicitation does not apply to those organizations "establish[ed], finance[d], maintain[ed], or control[led]" by a federal candidate).

25 11 C.F.R. § 9008.55(d). *See also* 2 U.S.C. § 441i(e)(4)(A).

26 11 C.F.R. § 9004.1. *See also* Fed. Elec. Comm'n, *Presidential Spending Limits for 2008*, *available at* http://www.fec.gov/pages/brochures/pubfund_limits_2008.shtml.

27 11 C.F.R. § 9003.3(a)(1)(i)(B).

28 *See* 11 C.F.R. § 9003.3(a)(1)(ii)-(v).

29 11 C.F.R. § 9003.3(a)(1)(i)(A).

30 Fed. Election Comm'n, *Report of the Audit Division on Bush-Cheney '04, Inc. and the Bush-Cheney '04 Compliance Committee, Inc. July 2, 2003 – Dec. 31, 2004 (*2007), *available at* http://www.fec.gov/audits/2004/20070322bush_cheney_compliance_04.pdf; Fed. Election Comm'n, *Report of the Audit Division on Kerry-Edwards 2004, Inc. and the Kerry-Edwards 2004 General Election Legal and Accounting Compliance Fund Feb. 18, 2003 – Dec. 31, 2004* (2007), *available at*

31 11 C.F.R. § 9003.3(a)(2)(i)(A)-(B), (F).

32 11 C.F.R. §§ 9003.3(a)(2)(i)(H); 9004.6.

33 Fed. Election Comm'n Advisory Opinion 2004-35 (2004), *available at* http://saos.nictusa.com/aodocs/2004-35.pdf.

34 11 C.F.R. § 9003.3(a)(1)(i).

35 *See* Kuhnhenn, *Creaky v. Cling, supra* note 6.

36 *Id.*

Intellectual Property

First Amendment Challenges to Copyright Laws: *Kahle v. Gonzales* and *Golan v. Gonzales*

*By David S. Olson**

There has been an attempt in the last few years to reverse some of the territorial creep of copyright law that has occurred in recent decades. These challenges have sought to undo copyright term extensions, increase the purview of fair use exceptions, and, in the cases discussed below, to constrict Congress's ability to expand copyright protections by use of First Amendment arguments. There is a reason for this recent pushback against copyright law. It is well known that the ways in which we create, distribute, critique, imitate, and copy writings and graphic arts have changed enormously in the three decades since the advent of the personal computer. Much ink has been spilled detailing the powerful changes that computers, cheap memory, digitization, the Internet, and increasing broadband adoption have wrought in the way that content is created and copyright enforcement challenged. But at least as important for copyright as technological changes are the profound changes to copyright law itself that have resulted in an unprecedented increase in the number of works copyrighted and the length of time copyright endures. In 1976—the same year that the Apple I personal computer was created[1]—Congress changed copyright law from an "opt-in" registration system, in which less than half of all new works were copyrighted each year and the average copyright lasted only a short time,[2] to a longer-term automatic copyright system from which it is relatively difficult to "opt-out."[3]

Prior to the 1976 Copyright Act, authors had to comply with formalities and register their works in order to receive copyright protection. The result was that about half of otherwise-qualifying works were never registered. In addition, the copyright term was fairly short before 1976, lasting only 14 years at first, which was eventually extended to 28 years. If an author wanted to renew his copyright, he had to pay a fee and officially renew the work. Approximately 85% of copyrighted works were never renewed.[4] This regime meant that for the entirety of United States history prior to the 1976 Copyright Act, (1) many works of authorship were never covered by copyright at all, and (2) the vast majority of copyrighted works lost copyright protection within 14-28 years.[5]

The 1976 Copyright Act worked a sea change on the practice of copyright in the United States. The Act made two important changes. It extended copyright terms to life of the author plus 50 years, and it did away with the registration requirement and all formalities for achieving copyright protection for a work. Thus, since January 1, 1978, virtually every bit of expression set down in a "tangible medium" has automatically received copyright protection for life of the author plus 50 years (75 years for certain older works and works made for hire). Subsequently, the Copyright Term Extension Act of 1998 extended copyright terms to life of the author plus seventy years (95 years for certain older works and works made for hire).[6]

Congress made these changes to copyright law out of concern for copyright owners and to comply with the Berne Convention for the Protection of Literary and Artistic Works.[7] Thus, when it came to formalities, Congress was concerned that copyright owners were losing their copyrights due to carelessness in complying with copyright formalities. Eliminating the formalities eliminated the problem. Likewise, when Congress extended copyright terms after intense lobbying from corporate copyright owners like the Walt Disney Company,[8] Congress was concerned with keeping valuable intellectual property in U.S. companies' hands. What Congress does not seem to have focused on is the interests of public users of copyrighted works. Making copyright owners register and renew their copyrights put the burden (which was low) of attaining and keeping copyright on the copyright owners, who had the best information as to copyright value and thus were the least cost avoiders. Likewise, although copyright term extension benefited a select group of copyright owners who owned work that had long-term value, it denied the public the right to unrestricted use of both commercially valuable and non-valuable copyrighted works for an additional 20 years.

The significance of these copyright law changes on those who would quote, copy, or otherwise use another's pre-existing work should not be underestimated. A researcher working before 1978 would safely assume that things published more than 56 years ago were in the public domain and could be used freely. He would also likely assume that anything out of print and published longer than 28 years ago was in the public domain, and he would be correct about 85% of the time. Finally, if a work bore no copyright notice, he would know immediately that the work was not covered by copyright and that he could freely use it.

Today's would-be user of copyrighted material faces a different scenario entirely. He knows that anything published since 1978 is automatically copyrighted for the life of the author plus 70 years. He knows that he may only safely use materials published over 85 years ago—prior to 1923.[9] For works published between 1923 and 1976 he must trace the history of registration and renewal to determine copyright status.[10]

Thus today's researcher or user of copyrighted materials faces a very different landscape than he did thirty years ago. While the pre-1976 Act researcher could make use of the vast majority of materials from just a generation or two earlier

** David S. Olson is an Assistant Professor at Boston College Law School and Non-Resident Fellow at the Stanford Center for Internet and Society.He had a very minor role in the appeal of Golan, and participated in a minor way during Professor Larry Lessig's preparation for oral argument of Kahle before the Ninth Circuit. The views expressed here are therefore emphatically his own and should not be attributed to plaintiffs in either case, the Stanford Law School Center for Internet and Society, Professors Lessig or Christopher Sprigman, or to the other attorneys involved in these cases.*

without running afoul of copyrights, today's researcher may need to seek copyright permission to use many materials from the last three or four generations. The researcher of 2108 will need to seek permission to use all materials created in the preceding 70 years and many works that are over 100 years old will remain covered by copyright.

Some may see little cause for concern because use of others' work has a flavor of theft to it—one thinks of college students lobbying for the destruction of copyright so that they may download music freely. But a growing body of literature points out that the more one looks at the concepts of novelty and originality in authorship, the more one realizes that, one way or another, there truly is (as Ecclesiastes tells us) "nothing new under the sun." For instance, in Judge Posner's recent book, *The Little Book of Plagiarism* (2007), the author discusses how Shakespeare could not have created his brilliant plays without extensive (and uncredited) use of numerous, and often recent, historical and literary sources.[11] Likewise, only a couple of generations ago, it was expected that persons of letters would quote without attribution from the works of others.[12] Such allusions were the mark of a well-read person.

But even when a user of copyrighted materials seeks to document each quote assiduously, and to add value to a work rather than to divert sales, copyright law can stymie non-exploitative use of copyrighted material. The case of *Shloss v. Estate of James Joyce* is a good example.[13] In that case, Carol Shloss, a Stanford professor and Joyce scholar, wrote a book on James Joyce's relationship with his daughter Lucia and the way that relationship and Lucia's artistic work impacted, and can be seen in, Joyce's *Finnegans Wake*. Shloss's book in no way competed for sales with *Finnegans Wake* or any of the other materials from which she quoted. If anything, her book served as a complementary good that increased interest in *Finnegans Wake*. Nevertheless, because the copyright in Joyce's works is controlled by his grandson (Joyce died in 1941), who takes a very dim view about any research whatsoever into his aunt, and thus refused any discussion of licensing Joyce's work for that purpose, Shloss's book was gutted of many of the supporting quotations before publication.[14] Shloss subsequently received pro bono representation from Larry Lessig and the attorneys of the Stanford Law School Center for Internet and Society Fair Use Project and sued for a declaratory judgment of fair use in order to post the supporting quotations on her academic website. After a period of intense litigation, Shloss won her right to quote Joyce's work as needed for her project.[15] The story ended happily for Shloss, though only after much time and aggravation. But she never would have had to fight that fight under the pre-1976 Act regime. And many authors never would have found the resources to fight.

Copyright does not only affect those wanting to quote written materials. Because dramatic arts such as music, choreography, and plays are also copyrighted, under the post-1976 Act regime, a community theater may not use dramatic works without license. Thus, many of the works of the last 85 years are off-limits to those who cannot pay a license fee, and going forward copyright will only be more restrictive, so that the majority of works from the previous century will be copyrighted.

The elimination of formalities with the 1976 Copyright Act had other profound effects. As discussed above, prior to 1978, over 90% of works were in the public domain by the time of their renewal. They had either never been copyrighted (about half of published works) or they were not renewed (about 85% of copyrighted works).[16] Of the small percentage of works whose copyrights were renewed, determining who owned the copyright was fairly straightforward. Not only did the work have to be published with notice and registered, but upon renewal the copyright owner was on record again. Thus, if one wanted to use a copyrighted work, generally one could quickly and easily determine with whom one should be negotiating. The abandonment of formalities and automatic term renewal has changed all of this and created a huge number of "orphan works."

"Orphan works" is the term that has been coined to describe the many works that are now in copyright but are out-of-print, and for which it is difficult to determine the copyright owner.[17] Because copyright now lasts so long, even when the original creator of a work can be easily identified, it may be very difficult to determine who owns the copyright after the originator has died. While the ownership of copyrights in valuable works by famous authors will usually be established at the time of death when other assets are divided, ownership of the copyrights of more ordinary people is not generally determined after death. Thus, such rights likely pass with the residue of the estate. And once copyright ownership has passed beyond the original author, it may continue to be sold, given, or devised in a series of private, non-centrally-recorded transactions. The puzzle of ownership can take significant time and energy to untangle, if the ownership interests can be untangled at all without an opinion from a court.[18]

It is against this very changed backdrop of copyright law that the Stanford Law School Center for Internet and Society litigated the recent cases of *Kahle v. Gonzales* and *Golan v. Gonzales*.

Background: *Eldred v. Ashcroft*

Kahle and *Golan* built on the litigation that began in *Eldred v. Ashcroft*.[19] In *Eldred*, plaintiffs challenged the Sonny Bono Copyright Term Extension Act of 1998 (CTEA).[20] The CTEA added twenty years to the term of copyright protection, so that for works produced after January 1, 1978, copyright protection lasts for life of the author plus 70 years.[21] For works for hire and works published before 1978 and still in copyright at the time of the CTEA, copyright was extended to 95 years.[22] Plaintiffs made two main arguments as to why the CTEA was unconstitutional, at least in part.[23] First, they argued that the CTEA's extension of copyright terms to existing works violated the Constitution's limitation of Congress's power to grant copyrights only for "limited times."[24] Plaintiffs argued that allowing Congress to retrospectively extend copyright terms of previously published works eviscerated the "limited times" limitation and allowed Congress the ability to enact perpetual copyright term a bit at a time. Second, plaintiffs argued that extensions of copyright terms for new and previously published works impacted the speech interests of those who would make use of work that would otherwise fall into the public domain.

Thus, plaintiffs argued, First Amendment scrutiny should be applied to the CTEA, and such scrutiny should result in the CTEA being held unconstitutional.

After losing at the district and circuit court levels, plaintiffs argued their case before the Supreme Court. The Supreme Court affirmed the lower courts, rejecting the plaintiffs' arguments. First, the Court addressed the plaintiffs' arguments that Congress could effectively grant perpetual copyright through a series of term extensions that would apply both prospectively and retrospectively.[25] The plaintiffs pointed out that a series of term extensions had kept certain works in copyright for much longer than originally expected. The Court acknowledged the plaintiffs' argument, but rejected that the CTEA's term extension had effectuated a perpetual copyright.[26] The Court seemed to think that Congress had extended copyright not out of a nefarious plan to achieve perpetual copyright via continued incremental term extensions but instead to harmonize copyright term in the U.S. with the term adopted by the E.U. in 1993, which is life of the author plus 70 years.[27] Notwithstanding that copyright term is now long, the Court held that on its face it is clearly still "limited."[28] In addition, the Court suggested that extending copyright terms to existing work could promote authorship because Congress's policy of parity assures authors that their works will get the benefit of future term extensions whether they are published before or after a statute extending copyright term.[29] Thus, according to the Court, the CTEA did not violate the Copyright Clause of the Constitution.[30]

Next, the Supreme Court rejected the plaintiffs' argument that the CTEA violated the First Amendment. The Court did "recognize that the D.C. Circuit spoke too broadly when it declared copyrights 'categorically immune from challenges under the First Amendment.'"[31] But said the Supreme Court, when Congress passes copyright legislation that does not "alter[] the traditional contours of copyright protection, further First Amendment scrutiny is unnecessary."[32] The Court noted two important features of copyright law that protect First Amendment interests—the idea/expression dichotomy and fair use.[33] The idea/expression dichotomy strikes a First Amendment balance in copyright law by allowing protection only of an author's expression of her ideas, but not of the ideas themselves.[34] Likewise, the fair use exemption to copyright protection allows use of copyrighted works, including direct copying and quotation, "for purposes such as criticism, comment, news reporting, teaching... , scholarship, or research,"[35] and for purposes of parody.[36] The Supreme Court also noted that the CTEA included specific additional protections for First Amendment interests including the rights of libraries, archives, and the like to make certain uses of works during their last twenty years of copyright for purposes such as preservation, scholarship or research,[37] and exemptions for small businesses and restaurants from paying performance royalties for playing televisions or radios in their businesses.[38]

The Supreme Court's decision in *Eldred* was an unequivocal defeat for the plaintiffs. But it was not a total defeat. The Court's rejection of the D.C. Circuit's holding that copyright statutes are "categorically immune" from First Amendment review left some hope for those seeking to restrict copyright expansion on First Amendment grounds. The logical converse of the holding that First Amendment scrutiny is not necessary where a statute does not "alter[] the traditional contours of copyright protection" is that, when a statute does alter those traditional contours, First Amendment review should apply.

But what are the traditional contours of copyright protection? The Court left this unanswered in *Eldred*. Are the traditional contours of copyright protection present so long as the idea/expression dichotomy and fair use protections remain in the Copyright Act? Or are the traditional contours of copyright protection altered whenever a statute passed under the copyright clause differs from historical precedent?

Kahle v. Gonzales

Two circuit courts have recently addressed these questions, each coming to a different conclusion.[39] Lead plaintiff in *Kahle v. Gonzales*, Brewster Kahle, runs the Internet Archive, whose goal it is to archive the oft-ephemeral data and content generated digitally on the Internet.[40] Co-plaintiff Richard Prelinger of Prelinger Associates, Inc., makes free copies of ephemeral movies available on the Internet. These plaintiffs made two challenges in *Kahle*. First, they challenged the Copyright Renewal Act of 1992, which eliminated copyright renewal requirements for works created between 1964 and 1977.[41] Second, and probably unfortunately, they argued that they should be allowed to present evidence that the Copyright Term Extension Act (already challenged in *Eldred*) violated the framers' understanding of the "limited times" restriction on copyright in the Constitution.

The Ninth Circuit quickly rejected the plaintiffs' CTEA argument. It stated that even though *Eldred* had only addressed in dicta what the framers would have thought the term "limited times" meant,[42] the Court had nevertheless upheld the constitutionality of the entire CTEA. True, the issue before the Court in *Eldred* was whether the retroactive application of the term extension was constitutional, while the plaintiffs in *Kahle* argued that the length of the term itself was unconstitutional. Still, the Ninth Circuit deemed this difference too small to merit review. The Ninth Circuit thought that the Supreme Court's reasoning as to why the CTEA's retroactive application was constitutional was broad enough to apply to the length of the CTEA's term as well. The Ninth Circuit stated that because "[a]rguments similar to Plaintiffs'" were presented in *Eldred*, "[t]he Supreme Court has already effectively addressed and denied Plaintiffs' arguments."[43]

The Ninth Circuit's disregard for the plaintiffs' argument as to copyright term may have encouraged the Court to give short shrift to their constitutional arguments about the elimination of the copyright renewal requirement. The plaintiffs argued that the elimination of the renewal requirement altered the traditional contours of copyright protection because renewal had long been a requirement for enjoying extended copyright term in the United States.[44] They argued that the effect the renewal requirement had of ensuring a large public domain of published works was a traditional contour of copyright. Plaintiffs pointed out, and the Court accepted, that renewal requirements both limited orphan works and made it so that "only a small percentage" of creative works were under copyright for the maximum term.[45]

The Court also acknowledged that "[e]liminating the renewal requirement dramatically increased the average copyright term and correspondingly decreased the number of works currently entering the public domain."[46] The plaintiffs characterized this as a change from an "opt-in" patent system to an "opt-out" system. Under the previous, opt-in system, formalities were required to gain copyright protection, and most copyrighted works were not renewed. Thus, the opt-in system created a large public domain. Under the so-called opt-out system, copyright attaches immediately to any expression fixed in a tangible medium, and it lasts for the full term of copyright without the author needing to do anything at all. If an author does not want the full extent of copyright protection, he must affirmatively opt-out of the default maximum copyright protection. An author may either disclaim copyright in his work or choose to reserve only limited rights via such licenses such as those provided by Creative Commons.[47] Only a small percentage of authors disclaim copyright or limit their copyright interest in their works, however, resulting in a vast majority of material that is automatically covered for the maximum term of copyright.

The plaintiffs in *Kahle* limited their challenge to the elimination of renewal requirements for works created between 1964 and 1977.[48] They did not challenge the elimination of the renewal requirement or of other formalities generally. The Ninth Circuit fixed on this fact, and instead of delving into whether the elimination of the renewal requirement changed the traditional contours of copyright law, the Court treated the question as whether Congress could place existing copyrighted work in parity with future works by eliminating the renewal requirement for both. The Ninth Circuit held summarily that *Eldred* disposed of this question. According to the Court, the Supreme Court had already ruled "that when Congress passed the CTEA, it 'placed existing and future copyrights in parity. In prescribing that alignment... Congress acted within its authority and did not transgress constitutional limitations.'"[49] Thus, a broad reading of the Ninth Circuit's reasoning in *Kahle* would mean that under *Eldred* any retroactive change in copyright law is constitutional so long as the change is aimed at providing parity for existing and future works.

Golan v. Gonzales

The Tenth Circuit in *Golan v. Gonzales* addressed whether a different change to copyright laws altered a traditional contour of copyright, thus necessitating First Amendment review.[50] Plaintiffs in *Golan* challenged section 514 of the 1994 Uruguay Round Agreements Act (URAA), which provided copyright protection to foreign works that were still in copyright in their country of origin, but were in the public domain in the U.S. because their authors failed to comply with U.S. copyright law formalities, or because the U.S. did not recognize copyright from the author's nation at the time the work was created.[51] By agreeing to section 514, the U.S. accepted Article 18 of the Berne Convention for the Protection of Literary and Artistic Works, which the U.S. previously had refused to join for more than 100 years.[52] The Berne Convention requires member countries to give equal copyright treatment to foreign and domestic authors. It also does away with copyright formalities and requires that signatory countries grant copyright protection to foreign works if those foreign works still have copyright protection in their countries of origin.[53] Thus, when the U.S. enacted legislation to comply with section 514 of the URAA the legislation granted copyright status to certain foreign works that were previously in the public domain in the U.S. due either to copyright holders' failure to comply with U.S. copyright formalities or failure to renew their copyrights.[54]

Plaintiffs in *Golan* each relied on artistic works in the public domain for their livelihood. They included orchestra conductors, educators, performers, publishers, archivists, and others who made use of works in the public domain. Many of the plaintiffs made use of works in the public domain because they could not afford to pay copyright licenses for uses that were often local or non-profit. Others had created derivative works based on foreign works in the public domain. The plaintiffs claimed that the URAA unconstitutionally interfered with their protected First Amendment interests in making use of public domain works.[55]

The district court disagreed, and held both that "Congress has historically demonstrated little compunction about removing copyrightable materials from the public domain" and that the plaintiffs had no First Amendment interests in the now-copyrighted foreign works.[56] The plaintiffs appealed to the Tenth Circuit, which took a different view.

The plaintiffs made three arguments in the Tenth Circuit. First, they argued that the CTEA's twenty-year extension violated the "limited Times" provision of the Copyright Clause because the founders would have found the extended term length "effectively perpetual."[57] Second, and similarly, the plaintiffs argued that the URAA's extension of copyright to works in the public domain violated the same "limited Times" provision of the Copyright Clause because allowing Congress to copyright works already in the public domain could enable Congress to repeatedly and perpetually copyright works.[58] Third, plaintiffs argued that the URAA violated the plaintiffs' First Amendment interests in making artistic use of the formerly public domain foreign works.[59]

The Tenth Circuit, like the Ninth, made short work of the plaintiffs' CTEA argument. The court agreed with the Ninth Circuit's conclusion that *Eldred* precluded a challenge to the CTEA's twenty-year term extension.[60] It also held that the URAA did not violate the Copyright Clause. The court rejected plaintiffs' argument that section 514 must be held unconstitutional to prevent Congress from forever keeping materials from the public domain. The court stated that, as in *Eldred*, "a regime of perpetual copyrights is clearly not the situation before us."[61] Moreover, the court rejected any invitation to second-guess Congress's determinations of how best to promote works of authorship under the Copyright Clause. The court stated firmly: "The clear import of *Eldred* is that Congress has expansive powers when it legislates under the Copyright Clause, and this court may not interfere so long as Congress has rationally exercised its authority."[62] The court believed that compliance with the Berne Convention, thereby assuring copyright protection for American works abroad, was a rational basis for enacting section 514.[63]

But, said the Tenth Circuit, the URAA's validity under the Copyright Clause did not make it immune to challenges based on other provisions of the Constitution, such as the First Amendment.[64] The Tenth Circuit began by addressing the Supreme Court's statement in *Eldred* that copyright laws should get the presumption of constitutionality *vis a vis* the First Amendment so long as the laws do not "alter[] the traditional contours of copyright protection."[65] After a detailed examination of the history of copyright laws in the United States, the Tenth Circuit concluded that "the traditional contours of copyright protection include the principle that works in the public domain remain there and that § 514 [of the URAA] transgresses this critical boundary."[66]

The Tenth Circuit analyzed how this alteration of the traditional contours of copyright affected the First Amendment interests of the plaintiffs. The Court found two First Amendment interests that plaintiffs had in the formerly public domain foreign works. First, the Court held that everyone has a non-exclusive right to use material in the public domain. Second, the Court held that "the First Amendment protects plaintiffs' right to unrestrained artistic use of the works at issue."[67]

Thus, said the Court, "at the moment that Dmitri Shostakovich's Symphony No. 5 entered the public domain, Plaintiff John Blackburn had a right to create a derivative work for a high school band to perform at an event commemorating 9/11."[68] Once created, the First Amendment protected Blackburn's right to perform his derivative work, according to the Tenth Circuit. Section 514 of the URAA impinged on Blackburn's and the other plaintiffs' First Amendment rights because the owners of the now-copyrighted original material could now charge fees for the performance of plaintiffs' derivative works.[69] The court found the plaintiffs' First Amendment interests greater than those of the plaintiffs in *Eldred*, because plaintiffs in *Eldred* had never enjoyed unfettered access to the works in question. Here, by contrast, the works at issue belonged to all when they were in the public domain, and the plaintiffs relied on their rights to use the works in performing or planning to perform the works and in building on the works for their own artistic productions. Thus, by removing works from the public domain, the court held that section 514 of the URAA "hampers free expression and undermines the values the public domain is designed to protect."[70]

The Tenth Circuit also held that "copyright's built-in free speech safeguards are not adequate to protect the First Amendment interests at stake."[71] The court held that neither the idea/expression dichotomy nor the fair use defense were adequate to protect plaintiffs' First Amendment interests in making use of the formerly public domain works. While the idea/expression dichotomy protects speech interests by allowing an author to copyright only her expression of an idea, but not the idea itself, in this case plaintiffs had previously had rights to the whole of the now-copyrighted works. The idea/expression dichotomy did not serve to protect these rights. Likewise, although fair use allows the use of a portion of a work for certain purposes such as "criticism, comment, news reporting, teaching... scholarship, or research," it could not serve to protect plaintiffs' interests in using the whole of the works that had been removed from the public domain.[72]

In addition, the court noted that unlike the CTEA, the URAA did not supplement the traditional First Amendment safeguards of copyright law. When Congress passed the CTEA, it gave additional protections to certain users of the works whose terms were extended. The CTEA gave certain rights to libraries, archives, and similar institutions for their actions related to preservation, scholarship or research in the last twenty years of a work's copyright. The CTEA also exempted small businesses and restaurants from paying performance royalties for music played from the radio, television or the like.[73] Unlike the CTEA, the URAA provided no supplemental First Amendment protections, other than a one-year safe harbor for using a restored work.[74]

The Tenth Circuit remanded the case to the district court for consideration of whether section 514 of the URAA is content-based or content-neutral and thus what level of First Amendment scrutiny is appropriate. If the district court finds that the URAA is content-based, then the district court is to apply strict scrutiny to determine "whether the government's interest in promulgating the legislation is truly 'compelling' and whether the government might achieve the same ends through alternative means that have less of an effect on protected expression."[75] If the district court finds that the URAA is content-neutral, then it must determine whether the restriction imposed by the URAA is "narrowly tailored to serve significant governmental interests."[76]

The URAA seems content-neutral. It allows foreign works to be copyrighted based on criteria that have nothing to do with the content of the works. Thus intermediate scrutiny will apply. In determining whether the URAA is "narrowly tailored to serve significant governmental interests," the district court may look to footnote 5 of the Tenth Circuit's opinion. There, the Court notes that in complying with the Berne Convention, the copyright laws of the United Kingdom, Canada, Australia, and India give parties who incurred any expenditure or liability in making use of a formerly public domain the right to continued use of that work even after the work is covered by copyright.[77] The copyright owner can only stop the reliance party from using the work if the copyright owner pays compensation to the user of the work, in an amount determined by negotiation or arbitration.[78] If the court determines that the government interest at issue is simply complying with the Berne Convention, then allowing users of formerly public domain work greater rights to continued use, as other countries have done, might be a constitutionally required narrowing of the URAA. If, on the other hand, the court determines that the government interest behind the URAA was more than just complying with Berne, and instead reflected Congressional judgment as to either the proper incentive for authors or the bargaining position of the United States with regard to international negotiations over copyright protections for authors, then the court is much less likely to second-guess whether Congress's enactment of the URAA was sufficiently narrowly tailored.

First Amendment Challenges to Copyright Laws After *Kahle* and *Golan*

So where are First Amendment challenges to copyright laws left after *Kahle* and *Golan*? First, the Tenth Circuit seems plainly correct in holding that the traditional contours of copyright protection must extend farther than the idea/expression dichotomy and the fair use defense. While these traditional safeguards do much to protect First Amendment interests, one can easily imagine legislation passed under the Copyright Act that would impinge on First Amendment interests notwithstanding the idea/expression dichotomy and fair use defense. The plaintiffs in *Kahle* gave one such example in their certiorari petition to the Supreme Court. Noting that, "copyright law has traditionally been viewpoint neutral," plaintiffs set forth the following scenario: "Imagine the European Union decided to deny copyright protection to 'hate speech,' and Congress, in an effort to 'harmonize' international copyright law, did the same." [79] The plaintiffs asked whether, in such a case, First Amendment review would be appropriate. It obviously would. Indeed, affirming that First Amendment review of copyright laws may sometimes be necessary does no more than affirm the general rule that legislation drafted under one constitutional provision must be consistent with the remainder of the Constitution. [80]

Eldred should not be understood as a case defining when the Copyright Clause trumps the First Amendment, or holding that a couple of safeguards within copyright law are sufficient to protect all First Amendment interests that might arise under any conceivable copyright law. Rather, *Eldred* is better understood as setting forth a simple rule of judicial economy. *Eldred* makes two simple and straightforward assumptions. First, the copyright laws that have developed over the last 200-plus years in the United States have adequately protected speech interests, and are therefore constitutional. Second, if a copyright law conforms to these "traditional contours of copyright protection" developed over the last 200 years, a court may presume that the law adequately protects speech interests, and may forgo First Amendment review. What are the "traditional contours of copyright protection" and whether the presumption of adequate First Amendment protection can be rebutted even when a law conforms to the traditional contours was not decided in *Eldred*, and is left for future courts to decide, as needed.

An interesting question is whether *Golan* is a one-off departure from the traditional contours of copyright, the likes of which we will not see again, It is hard to imagine future copyright laws that will remove more material from the public domain. Legislation under the URAA was enacted to harmonize U.S. law with foreign copyright law. Now that the U.S. has done away with formalities and "restored" copyright to those foreign works that were formerly denied copyright due to idiosyncrasies of U.S. law, it is extremely unlikely that other public domain works will be copyrighted.

Are there other current or future copyright laws that may depart from the traditional contours of copyright protection and thus require First Amendment analysis? The elimination of copyright formalities may be a candidate for direct assault. As the Ninth Circuit acknowledged in *Kahle*, formalities have long been a feature of copyright laws, and their elimination has resulted in a vastly diminished public domain. Although courts may be loath to second-guess Congress's determinations of the appropriate incentives for authors, the changes to copyright law that have resulted in most works being in copyright for a century or more plainly has an effect on the speech that others are allowed to make and the ways that others can communicate and discuss copyrighted expressions of ideas.

If litigants bring a First Amendment challenge to the elimination of copyright formalities (rather than just to the elimination of the renewal requirement for existing works, as plaintiffs challenged in *Kahle*), any First Amendment analysis should be under intermediate scrutiny, because elimination of formalities is content-neutral. Under intermediate scrutiny, it is hard to see how litigants could force the resumption of formalities given that the government has a significant interest in complying with Berne (which demands elimination of formalities that would deprive a copyright holder of "enjoyment and exercise" of the economic rights appertaining to copyright). But some reliance rights allowing continued use of works whose copyrights were not voluntarily registered or renewed before enforcement proceedings began might be both acceptable under Berne and serve to more narrowly tailor U.S. copyright law to account for First Amendment interests in the public domain.[81]

At the end of the day, the courts seem to recognize that determining the contours of copyright is a job for Congress. Likewise, courts are likely to be deferential to Congress in establishing the boundary between copyright and the public domain, even when the First Amendment interests of users of copyrighted material are considered. Whatever the success of future First Amendment challenges to copyright laws that depart from the traditional contours of copyright protection, focusing attention on the fact that copyright laws can injure First Amendment interests of users of copyrighted works is salutary. Hopefully, Congress will consider the First Amendment interests of copyright users in passing future copyright laws so that courts need not attempt to navigate between the Scylla of narrowly tailoring copyright laws so as to protect speech interests and the Charybdis of second-guessing congressional decision-making about how best to promote works of authorship.

Endnotes

1 Steve Wozniak and Steve Jobs created the Apple I computer in March 1976, and founded Apple Computer, Co. on April Fools' day. *See* Lev Grossman, 80 Days That Changed the World: April 1, 1976, Time.com (available at http://www.time.com/time/80days/760401.html (last visited April 21, 2008). In March 1977 they created the Apple II computer, which many consider to have been the first affordable and easily-usable personal computer. *See* http://web.mit.edu/invent/iow/apple.html (last visited April 21, 2008)

2 *See* Christopher Sprigman, *Reform(aliz)ing Copyright*, 57 Stan. L. Rev. 485, 491-93 (2004).

3 The Copyright Act of 1976 became Public Law number 94-553 on October 19, 1976 and went into effect on January 1, 1978. It is codified, as subsequently amended, at 17 U.S.C. §100 et seq.

4 Sprigman, *supra* note 2, at 519.

5 *Id.* at 522.

6 17 U.S.C. §§ 102, 302.

7 Berne Convention Implementation Act, Pub. L. No. 100-568, 102 Stat. 2853 (1988). The Berne Convention is available at http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html (last visited April 21, 2008).

8 *See* William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. Chi. L. Rev. 471, 483 (noting "Disney's successful efforts to lobby for the Sonny Bono [Copyright Term Extension]Act").

9 Works published prior to 1923 are in the public domain because prior to the 1976 Copyright Act, the 1909 Copyright Act provided a maximum term of 56 years (28-year original term plus 28-year renewal term). 17 U.S.C. § 24 (repealed 1976). The 1976 Act only extended copyright terms for works still in copyright. 17 U.S.C. § 304. Thus those works published before 1923 that had already entered the public domain remained there.

10 17 U.S.C. § 304.

11 Richard A. Posner, The Little Book of Plagiarism 51-74 (2007).

12 *Id.* at 54-56.

13 515 F. Supp. 2d 1083 (N.D. Cal. 2007). I served as counsel for Professor Shloss in this case. Accordingly, I emphasize that the opinions expressed here are my own, and should not be attributed in any way to Carol Shloss or to the Stanford Law School Center for Internet and Society or Stanford Law School Cyberlaw Clinic.

14 *Id*. at 1068, 1073.

15 *Id*. at 1085-86.

16 *See supra* notes 1-3 and accompanying text.

17 U.S. Register of Copyrights, *Report on Orphan Works* (2006), *available at* http://www.copyright.gov/orphan/orphan-report-full.pdf (last visited April 21, 2008).

18 *See id.* at 32.

19 537 U.S. 186 (2003).

20 Pub. L. No. 105-298, 112 Stat. 2827 (1998).

21 17 U.S.C. § 302(a).

22 *Id*. § 302(c).

23 I do not address plaintiffs' third argument, that the CTEA violated the public trust doctrine by transferring public property (the works about to enter the public domain) into private hands without a showing of government benefit. Plaintiffs abandoned this argument by the time they filed their certiorari petition with the Supreme Court.

24 U.S. Const. art. I, § 8, cl. 8 (stating, in relevant part, that "Congress shall have power... To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.").

25 *Eldred*, 537 U.S. at 208.

26 *Id.* at 209 ("As the Court of Appeals observed, a regime of perpetual copyrights 'clearly is not the situation before us.'") (quoting Eldred v. Reno, 239 F.3d 372 (D.C. Cir. 2001)).

27 *Id.* at 195-96 (citing H.R.Rep. No. 94-1476, p. 135-36 (1976), U.S.Code Cong. & Admin.News 1976, p.5659 (explaining that term extension of 1976 Copyright Act was done in part to harmonize United States law with foreign copyright terms); Council Directive 93/98/EEC of 29 October 1993 Harmonizing the Term of Protection of Copyright and Certain Related Rights, 1993 Official J. Eur. Coms. (L290), p. 9 (EU Council Directive 93/98)).

28 *Id.* at 209.

29 *Id.* at 203-04.

30 *Id.* at 204.

31 *Id.* (citing Eldred v. Reno, 239 F. 3d at 375).

32 *Id.* at 221 (citing Harper & Row Pub., Inc. v. Nation Ent., 471 U. S. 539, 560 (1985); San Francisco Arts & Athletics, Inc. v. U. S. Olympic Comm., 483 U. S. 522 (1987)).

33 *Id.* at 219-20.

34 *Id.* at 219.

35 *Id.* at 220 (quoting 17 U.S.C. § 107).

36 *Id.* (citing Campbell v. Acuff-Rose Music, Inc., 510 U. S. 569 (1994)).

37 *Id.* (citing 17 U. S. C. §108(h)).

38 *Id.* (citing 17 U. S. C. §110(5)(B)).

39 Like *Eldred*, both of these cases have been litigated by Larry Lessig and the Stanford Law School Center for Internet and Society.

40 487 F.3d 697 (9th Cir. 2007).

41 Pub. L. No. 102-307, 106 Stat. 264 (1992).

42 *Eldred*, 537 U.S. at 210 n. 16 (majority opinion stating "[i]t is doubtful... that those architects of our Nation, in framing the 'limited Times' prescription, thought in terms or the calculator rather than the calendar.").

43 *Kahle*, 487 F.3d at 698.

44 *Id.* at 698-700. *See also* Stewart v. Abend, 495 U.S. 207, 217 (1990) ("Since the earliest copyright statute in this country, the copyright term of ownership has been split between an original term and a renewal term.").

45 *Id.* at 699.

46 *Id.*

47 *See* http://creativecommons.org/license/ (last visited Apr. 21, 2008).

48 *Kahle*, 487 U.S. at 699-700.

49 *Kahle*, at 700 (quoting *Eldred*, 537 U.S. at 194).

50 501 F.3d 1179 (10th Cir. 2007).

51 Pub. L. No. 103-465, 108 Stat. 4809, 4976-80 (1994), codified at 17 U.S.C. §§ 104A, 109.

52 *Golan*, 501 F.3d at 1182, note 2; Berne Convention Implementation Act, Pub. L. No. 100-568, 102 Stat. 2853 (1988).

53 *Id.* at 1189; 17 U.S.C. §§ 104A, 109.

54 17 U.S.C. §§ 104A, 109.

55 *Golan*, 501 F.3d at 1182-83.

56 *Id.* at 1183 (quoting Golan v. Gonzales, No. Civ. 01-B-1854, 2005 WL 914754, at *14 (D. Colo. April 20, 2005)).

57 *Id.* at 1185.

58 *Id.* at 1185-86.

59 *Id.*

60 *Id.* at 1185 ("As the *Kahle* court reasoned, 'the outer boundary of "limited Times" is determined by weighing the impetus provided to authors by longer terms against the benefit provided to the public by shorter terms. That weighing is left to Congress, subject to rationality review.'") (quoting *Kahle*, 487 F.3d at 701).

61 *Id.* at 1186 (quoting *Eldred*, 537 U.S. at 209). The Court also noted that the D.C. Circuit had rejected an identical argument in *Luck's Music Library, Inc. v. Gonzales*, 407 F.3d 1262 (D.C. Cir. 2005) (holding § 514 of the URAA valid over the Copyright Clause). The D.C. Circuit did not address validity over the First Amendment.

62 *Id.* at 1187.

63 *Id.* ("we do not believe that the decision to comply with the Berne Convention, which secures copyright protections for American works abroad, is so irrational or so unrelated to the aims of the Copyright Clause that it exceeds the reach of congressional power") (citing Eldred at 208).

64 *Id.* (citing Saenz v. Roe, 526 U.S. 489, 508 (1999); Buckley v. Valeo, 424 U.S. 1, 132 (1976)).

65 *Eldred*, 537 U.S. at 221.

66 *Golan*, 501 F.3d at 1189. The D.C. Circuit in *Luck's Music Library, Inc. v. Gonzales*, 407 F.3d 1262, 1265-66 (D.C. Cir. 2005), came to the opposite conclusion—that removal of works from the public domain had been done several times in the past under federal copyright laws. But the Tenth Circuit's analysis seems better. The Tenth Circuit noted that the Government's purported examples of copyright laws that removed works from the public domain (the first Copyright Act (passed in 1790), laws allowing registration of foreign works that could not be registered during World War II, and a handful of private bills) were at best exceptions to the traditional rule in copyright that works in the public domain stay there. *Golan*, 501 F.3d at 1192.

67 *Id*. at 1193. The Court noted that "The Supreme Court has emphasized that the right to artistic expression is near the core of the First Amendment." The Tenth Circuit quoted *Ward v. Rock Against Racism*, 491 U.S. 781, 790 (1989):

Music is one of the oldest forms of human expression. From Plato's discourse in the Republic to the totalitarian state in our own times, rulers have known [music's] capacity to appeal to the intellect and to the emotions, and have censored musical compositions to serve the needs of the state.... The Constitution prohibits any like attempts in our own legal order. Music, as a form of expression and communication, is protected under the First Amendment.

68 *Id.*

69 *Id.*; 17 U.S.C. § 104A(d)(3) (allowing creator of derivative work created while foreign work was in public domain to continue using work if reasonable compensation is paid to copyright owner, which compensation is determined via normal infringement damages analysis). If the creator of the derivative work cannot pay the fee to the copyright owner, not only would they not be able to use their derivative work, but if the holding of cases such as *Anderson v. Stallone* is followed, the owner of the copyright in the original work could both use the derivative work at will, and is the only party who could license others to use the derivative work. *See* Anderson v. Stallone, No. 87-0592 WDKGX, 1989 WL 206431 (C.D. Cal. April 25, 1989) (giving copyright owner exclusive rights to derivative work created without authorization).

70 *Id.* at 1193-94 (citing Meade v. U.S., 27 Fed. Cl. 367, 372 (Fed. Cl. 1992)).

71 *Id.* at 1196.

72 *Id.* at 1195-96 (quoting 17 U.S.C. § 107).

73 *Id.* (citing *Eldred*, at 220 (quoting 17 U.S.C. § 108(h))).

74 *Id.* at 1196 (citing 17 U.S.C. § 104A(d)(2)).

75 *Golan*, 501 F.3d at 1196 (citing U.S. v. Playboy Entm't Group, Inc., 529 U.S. 803, 813 (2000)).

76 *Id.* (quoting *Rock Against Racism*, 491 U.S. at 791)).

77 *Id.* at note 5.

78 *Id.* at 1197 (citing Irwin Karp, *Final Report, Berne Article 18 Stud on Retroactive United States Copyright Protection for Berne and Other Works*, 20 Colum.-VLA J.L. & Arts 157, 178 (Winter 1996) (quoting "The Copyright (Application to Other Countries) Order in Council 998"; June 13, 1989 Article 7(2), *reprinted in* Copyright Laws and Treaties of the World, UK: Item 7c, at 2 (Unesco Supp. 1989-90))).

79 Kahle v. Gonzales Petition for Writ of Certiorari, 2007 WL 2323450 (2007) at *15.

80 *See* Saenz v. Roe, 526 U.S. 489, 508 (1999) ("Article I of the Constitution grants Congress broad power to legislate in certain areas. Those legislative powers are, however, limited not only by the scope of the Framers' affirmative delegation, but also by the principle that they may not be exercised in a way that violates other specific provisions of the Constitution.") (internal quotation marks omitted).

81 Christopher Sprigman makes a similar, and much more fleshed-out argument as to how "new-style formalities" can be adopted in the United States and still comply with the Berne Convention. Christopher Sprigman, *Reform(Aliz)Ing Copyright*, 57 Stan. L. Rev. 485, 555 (2004) (arguing for "new-style formalities" that "preserve formally voluntary registration, notice, and recordation of transfers (and reestablish a formally voluntary renewal formality) for all works... but then incent compliance by exposing the works of noncompliant rightsholders to a 'default' license that allows use for a predetermined fee.").

When "Exclusive" is not "Exclusive" and "Compulsory" not "Compulsory:" *eBay v. MercExchange* and *Paice v. Toyota*

*By David L. Applegate**

The source of American patent law, Article I, section 8, of the U.S. Constitution, empowers Congress to "promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[1] But what does it mean to grant inventors the "exclusive Right" over their respective inventions (albeit for limited times)? And have Congress and the courts been faithful to that grant of power?

In a series of recent cases, the Court of Appeals for the Federal Circuit, the district courts, and the Supreme Court of the United States have all given cause to question whether inventors indeed have "the exclusive Right" to their respective discoveries, even for limited times.[2]

I. The Statutory Framework

The current U.S. Patent Code, the Patent Act of 1952, mandates that the courts award compensatory damages upon a finding of infringement, but merely permits the courts to grant injunctions in accordance with traditional principles of equity. In its entirety, the first paragraph of section 284 of the Patent Act provides: "Upon finding for the claimant the court *shall* award the claimant damages *adequate to compensate for the infringement*, but in no event less than a reasonable royalty for the use made of the invention by the infringer, together with interest and costs as fixed by the court."[3] In pertinent part, section 283 provides: "The several courts having jurisdiction of cases under this title *may* grant injunctions in accordance with the principles of equity to prevent the violation of any right secured by patent, on such terms as the court deems reasonable."[4]

Just on the face of it, this statutory language presents a problem. First, the well-established "principles of equity" that section 283 explicitly incorporates require a plaintiff seeking a permanent injunction to demonstrate, in part, that "remedies available at law, such as monetary damages, *are inadequate to compensate for that injury*."[5] But because section 284 requires the court to award damages "*adequate to compensate for the infringement,*"[6] and the infringement is the injury to be compensated, sections 283 and 284 would seem to be mutually exclusive.[7]

One way to harmonize this apparent contradiction is to interpret section 284 to require the award of damages for "the [past] use made of the invention by the infringer" and to permit an injunction under section 283 to prevent "the [future] violation of any right secured by patent…."[8] This would in effect require courts to insert words that Congress did not include—but that is precisely what courts have generally done in the past. Patent holders that have proven infringement have typically received damages awards for past infringement, plus permanent injunctions against future infringement.[9] The parties are of course free to negotiate other arrangements, and frequently do; depending on the circumstances, many parties settle infringement suits with the grant of a license to the defendant to continue to practice the patented invention, subject to geographic and temporal limitations.

This linguistically unsatisfactory scenario at least secures to inventors the "exclusive Right to their respective… Discoveries" by enjoining others, for the duration of the patent, from exploiting the patented rights without the permission of the patent holder. But "exclusive," it seems, does not always mean "exclusive."

II. Tensions on the System – Patent Trolls and Business Method Patents

Recognized nearly contemporaneously with the first patent statute in 1790, "business method" patents have become more widespread since the Court of Appeals for the Federal Circuit (CAFC) decided in *State Street Bank & Trust Co. v. Signature Financial Group, Inc.*[10] that it was no longer necessary for the courts or the Patent Office to distinguish between "technology-based" and "business-based" patents.[11] Although the Federal Circuit has recently signaled its willingness to revisit this holding, the increased acceptability in recent years of business method patents and the increasing reliance by businesses on computers and the Internet have led to the realization that patents can be as valuable to owners that do not directly practice them as to owners that do, giving rise to patent holding companies.[12]

Sometimes disparaged as "patent trolls," patent holding companies typically own patents, but do not actually manufacture or sell a product or use the patented process. Instead, they license the rights to do so to others, sometimes after first charging others with infringement.[13] As the name "patent troll" implies, some view patent holding companies as illegitimate abusers of the patent system, waiting like trolls under the bridge to exact a toll. Others point out that patent holding companies are simply doing what the law and the Constitution permit. Regardless, the perceived inequity of permitting one who does not use an invention—and may not even be the inventor of that invention—to prevent others from using the invention has helped shape current efforts at patent law reform. Because injunctions are ultimately equitable remedies, this perceived inequity influences the grant of injunctions.

Legislation is currently pending before Congress, for example, that would change existing patent law in part by limiting the remedies available for infringement.[14] Proposals include limiting damages for the two-year period before actual notice of infringement for inventions not actually incorporated in articles made, offered for sale, or imported in or into the U.S.

* *David Applegate chairs the Intellectual Property Practice Group of Williams Montgomery & John Ltd., a Chicago-based firm of trial lawyers devoted to business litigation nationwide. The opinions expressed here are his own. Mr. Applegate discussed the meaning of "limited Times" in the Copyright Clause in* David Applegate, *Eldred v. Ashcroft: Just Another Mickey Mouse Copyright Case?* 3 Engage 3 (2002).

(including but not limited to business methods)—i.e., patents held by patent holding companies—as compared to upwards of six years for other inventions.[15] At the same time, the courts in general, and the Supreme Court in particular, have begun to approach the grant of injunctions in patent infringement cases more cautiously.

III. *eBay Inc. v. MercExchange L.L.C.*

In *eBay Inc. v. MercExchange L.L.C.*, plaintiff MercExchange obtained a jury verdict against online auction house eBay for willfully infringing three MercExchange patents, which eBay then appealed to the U.S. Court of Appeals for the Federal Circuit.[16] The CAFC affirmed the judgment of willful infringement and granted a permanent injunction against eBay, and eBay then took an appeal to the U.S. Supreme Court.[17]

A. Majority Opinion and Justice Roberts Concurrence

On May 15, 2006, the U. S. Supreme Court repudiated the notion that a prevailing patent infringement plaintiff is automatically entitled to a permanent injunction. Instead, the High Court redirected lower courts to consider carefully, in deciding whether to enjoin infringing conduct, the four traditional equitable factors: (1) adequacy of remedy at law, (2) likelihood of success on the merits, (3) balancing of harms, and (4) the public interest.[18] In a unanimous opinion by Justice Clarence Thomas, the Court found that the Patent Act did not replace traditional equitable principles with an "automatic injunction" rule, but also noted that a district court still has discretion to order a permanent injunction, even if a patent holder unreasonably declines to use its patent.[19]

Concurring separately and joined by Justices Ginsburg and Scalia, Chief Justice Roberts agreed that equitable considerations indeed govern the grant of injunctions, but cautioned that a patent holder's right to exclude all others cannot be protected through monetary damages that allow continued use of the patent, and that a major departure from a long history of equity should not be "lightly implied."[20]

B. Justice Kennedy Concurrence

Also concurring separately, Justice Kennedy strongly suggested that part of the Court's concern in *eBay* stems from the emergence of business method patents and patent holding companies, and that a twin-tiered system may therefore be appropriate. "In cases now arising trial courts should bear in mind that in many instances the nature of the patent being enforced and the economic function of the patent holder present considerations quite unlike earlier cases," he began.[21] The difference, he noted, has been the emergence of patent holding companies:

> An industry has developed in which firms use patents not as a basis for producing and selling goods but, instead, primarily for obtaining licensing fees.... For these firms, an injunction, and the potentially serious sanctions arising from its violation, can be employed as a bargaining tool to charge exorbitant fees to companies that seek to buy licenses to practice the patent.[22]

As a result, Justice Kennedy continued, instead of an injunction, continuing damages in the form of a royalty may be the appropriate remedy:

> When the patented invention is but a small component of the product the companies seek to produce and the threat of an injunction is employed simply for undue leverage in negotiations, legal damages may well be sufficient to compensate for the infringement and an injunction may not serve the public interest. In addition injunctive relief may have different consequences for the burgeoning number of patents over business methods, which were not of much economic and legal significance in earlier times. The potential vagueness and suspect validity of some of these patents may affect the calculus under the four-factor test.[23]

But if a patent holder—including a patent holding company—is not automatically entitled to enjoin a proven infringer from continuing to infringe, then in what sense have Congress and the courts secured for the inventor the "exclusive rights" to the invention? In a case with no injunction, the patent holder is left with only its damages remedy for continuing infringement. That means, in practice, either the patent holder has to bring multiple lawsuits to recover damages for what in the future will become past infringement or, more sensibly, the court will have to impose an ongoing royalty.[24] In either case, however, the patent owner loses the "exclusive right" to exploit the patent and is merely compensated for the partial taking of that right. In effect, then, in derogation of the general disapproval in the United States of compulsory licensing,[25] a court that does not enjoin future infringement compels the patent holder to license its patent to the infringer.[26]

IV. Applying the Equities

In some cases, a compulsory license seems fair and equitable, but in others it appears less so. Suppose, for example, that a researcher holds the patent on a life-saving drug, but either chooses not to produce the drug or lacks the resources to do so. Assume further that a pharmaceutical company is willing and able to produce the drug, but that the patent holder irrationally refuses to license the patent. After balancing the harms, even assuming the plaintiff's likely success on the merits, the public interest would fairly clearly call for denying a permanent injunction and instead awarding damages in the form of a lump sum or a continuing royalty—in effect, a compulsory license.

In practice, of course, the equities are rarely so stark. One can imagine, perhaps, a hypothetical situation in which an inventor motivated enough to invent a life-saving drug would be irrational enough to refuse to license its production, but that is rarely the way the world works. Even if the inventor for some reason disliked the drug company that approached him or drug companies in general, a competitor that did not share those disabilities would likely arise and the drug would eventually find its way to market.

On the other hand, suppose an online shopping innovator has a patent that permits repeat shoppers to store their payment and shipping information online in order to make subsequent purchases with a single mouse click, and that the patent holder refuses to share it with other merchants. Does the resulting inconvenience and annoyance to thousands, perhaps millions, of shoppers, together with the attendant value of their lost time, justify compelling the owner to license the patent to competitors?

Or suppose instead a patent holder has invented a new technology that enables cars to run more cleanly, and that a car company has infringed the patent, either because it was unaware of the patent or because it took a different interpretation of its claims. Should the court, upon finding infringement, enjoin the car company's production of infringing models, knowing that doing so will put autoworkers out of work, disrupt the supply chain, and increase pollution of the environment? Or do the public interest and the balancing of harms once again justify compelling the patent owner to license it?

In either case, supposing that the patent owner did not want to share the patent at any price, should the court that compels the patent holder to license the patent also force the parties to set the price, or should the court directly set the price? To what extent, in other words, is the court justified in ignoring market incentives in the name of the public good? Or does the answer depend on whether, in general, the patent holder has shown a willingness to license the patent to others?

A. Commercializing vs. Licensing

Even though the Supreme Court in *eBay* admonished lower courts to evaluate each case individually, post-*eBay* experience suggests that the result of applying those equities may still be in practice a two-tiered system.[27] In part, post-*eBay* cases have found that the patent-holder's willingness to license its patents and its failure to commercialize the patented invention itself—that is, whether the patentee is a patent holding company—mitigate in favor of deciding that damages are adequate.[28] Post-*eBay* cases also suggest that, on the other hand, district courts are more likely to grant permanent injunctions against continuing infringement, where the infringers compete directly with the patent holders in manufacturing and selling the patented invention.[29] In effect, what is emerging in the post-*eBay* world is a two-tiered system in which patent holders that manufacture or sell their inventions are more likely to obtain injunctions, and patent holders that merely license their inventions are more likely to get continuing royalties instead.

B. Relative Contribution of Patent to Product

Justice Kennedy's concurrence in *eBay* also noted that legal damages may "well be sufficient" to compensate for the infringement when "the patented invention is but a small component" of the accused infringing product, and both Congress and the lower courts have followed suit.[30] For example, both S. 1145 and H.R. 1908 would require courts to take into account, in assessing a reasonable royalty, the economic value of the infringing product or process that is attributable to the claimed invention's specific contribution over the prior art.[31]

District courts have also begun to emphasize the relative contribution of the patented invention to the infringing product. In *IMX, Inc. v. LendingTree, LLC*, for example, the District Court of Delaware found that the infringing product was based primarily on the asserted patent, which weighed in favor of granting an injunction.[32] But in *z4 Technologies, Inc. v. Microsoft Corp.*, the Eastern District of Texas denied an injunction where it found that Microsoft, the accused infringer, "only uses the infringing technology as a small component of its own software, and it is not likely that any consumer… purchases these products for the [patented] functionality."[33]

C. Treble Damages and Injunctions

The U.S. Supreme Court did not address the potential relationship between an injunction and treble damages in *eBay*, but in the context of patent infringement actions both are properly considered equitable remedies.[34] In deciding whether to award enhanced damages under section 284, courts balance equitable considerations, *SRI International, Inc. v. Advanced Technology Labs., Inc.*,[35] and their award is not subject to the Seventh Amendment right to a jury trial.[36]

In the context of willful patent infringement, therefore, a court might still order an ongoing "treble royalty" as "equitable relief" for future infringement that in most cases might have the practical effect of an injunction.[37] This approach would mitigate the harshness of an injunction against infringement, while still providing patent holders with an equitable remedy. Infringers with substantial non-patented interests at stake could also still infringe and pay enhanced damages without risking contempt of court, and would not face potentially exorbitant demands by a hold-out patent owner. It appears, however, that no court has yet awarded a patent owner treble damages based on future infringement.

V. *Paice LLC v. Toyota Motor Corp.*

A. The District Court

Among the most prominent post-*eBay* cases to deal with forced licensing in a patent infringement case is *Paice LLC v. Toyota Motor Corp,* involving a transaxle used in Toyota Motor Company's environmentally friendly "hybrid" vehicles: the Prius II, the Highlander, and the Lexus RX400h.[38] After a jury found that Toyota had infringed certain Paice patents for hybrid electric vehicle transaxles, the district court nonetheless denied an injunction against future infringement. Instead, on its own initiative, the district court imposed an "ongoing royalty" of $25 per infringing vehicle, with specified payment terms, while leaving the parties free to negotiate other terms if they wished.[39]

At trial, the district court rejected the customary equitable arguments in favor of an injunction. Regarding irreparable injury, the district court found that not granting an injunction would not adversely affect Paice's ability to practice or to license the patented technology, because Paice had allegedly adduced only "vague testimony" that it had been "sidelined" in its business dealings during litigation, and because Paice did not actually manufacture any goods.[40] Given the relatively small royalty awarded by the jury in relation to the overall value of the vehicles, and the fact that Paice had offered a license to Toyota during the post-trial period, the court also concluded that monetary damages provided an adequate remedy.[41] The district court further found that the public interest favored neither party, and that the balance of hardships favored Toyota because an injunction would disrupt "related business, such as dealers and suppliers," could have an adverse effect on the "burgeoning hybrid market," and might damage Toyota's reputation.[42]

B. The Court of Appeals

On appeal, the Federal Circuit upheld the jury's finding of infringement, the district court's denial of an injunction, and the district court's order of an ongoing royalty without affording Paice a right to a jury trial, but remanded the case because the "order provides no reasoning to support the selection of $25..." as the ongoing royalty.[43] On remand, the CAFC observed that the district court might consider additional necessary evidence to account for economic factors arising out of an imposed ongoing royalty and "may determine that $25 is, in fact, an appropriate royalty rate going forward," but "without any indication as to why that rate is appropriate, we are unable to determine whether the district court abused its discretion."[44]

The CAFC began its analysis by observing that "the most apparent restriction imposed" by the statutory language of 35 U.S.C. § 283 was not that injunctions be granted in accordance with the principles of equity, but that "injunctions granted thereunder must 'prevent the violation of any right secured by patent.'"[45] "The more difficult question raised by this case," the court therefore said, "is whether an order *permitting use* of a patented invention in exchange for a royalty is properly characterized as *preventing* the violation of the rights secured by the patent"—in effect asking whether a patent indeed gives an inventor "the exclusive right" to the invention "for limited times."[46]

Under some circumstances, the court continued, "awarding an ongoing royalty for patent infringement in lieu of an injunction may be appropriate," citing its own precedent of *In Shatterproof Glass Corp. v. Libbey–Owens Ford Co.*, in which the court had upheld a 5% court-ordered royalty on sales "for continuing operations."[47] The Court also relied on precedents in the field of antitrust law, in which "'mandatory sales and reasonable-royalty licensing' of relevant patents are 'well-established forms of relief… particularly where patents have provided the leverage for or have contributed to the antitrust violation adjudicated.'"[48] But awarding an ongoing royalty "where 'necessary' to effectuate a remedy," the *Paice* court concluded, "does not justify… such relief as a matter of course whenever a permanent injunction is *not* imposed."

Instead, in most cases, the court continued, "the district court may wish to allow the parties to negotiate a license amongst themselves regarding future use of a patented invention before imposing an ongoing royalty."[49] The *Paice* majority's focus, in other words, was not on whether to grant an injunction, or even whether to compel an ongoing license, but merely whether the parties or the court should set the terms of that license.

1. "Compulsory" License?
Majority View

With considerable disagreement in a concurring opinion from Judge Rader, the CAFC nonetheless insisted that the court-ordered royalty of $25 per vehicle that it upheld was not a "compulsory" license. Unlike, for example, a compulsory license under the copyright laws, the CAFC observed, the license that the district court had compelled Paice to grant to Toyota was restricted to Toyota rather than being generally available on demand to other parties:

> The term "compulsory license" implies that *anyone* who meets certain criteria has congressional authority to use that which is licensed. See, e.g., 17 U.S.C. § 115…[50] By contrast, the ongoing royalty order at issue here is limited to one particular set of defendants; there is no implied authority in the court's order for any other auto manufacturer to follow in Toyota's footsteps and use the patented invention with the court's imprimatur.[51]

2. "Compulsory" License? – Judge Rader View

In a separate concurrence, Judge Rader emphatically disagreed: "calling a compulsory license an 'ongoing royalty' does not make it any less a compulsory license."[52] Rather, he said, "this court should require the district court to remand this issue to the parties, or to obtain the permission of both parties before setting the ongoing royalty rate itself."[53] Not even Judge Rader, however, would require the district court to enjoin future infringement; his only concern seemed to be that the parties have a chance to negotiate an ongoing royalty rate before the district court imposed the rate itself.

"District courts have considerable discretion in crafting equitable remedies," Judge Rader continued, "and in a limited number of cases, as here, imposition of an ongoing royalty may be appropriate.… [But to] avoid many of the disruptive implications of a royalty imposed as an alternative to the preferred remedy of exclusion, the trial court's discretion should not reach so far as to deny the parties a formal opportunity to set the terms of a royalty on their own."[54]

3. Unanswered Questions

But suppose, knowing the district court has already ordered a $25 per unit rate, either Paice or Toyota refuses to agree to anything else? Can the district court then not safely reinstate a $25 per unit royalty rate, after explaining in sufficient detail to satisfy the court's majority why it finds that rate appropriate? Will the CAFC then not sustain the award as well-reasoned? And if that should happen, then has not the court, in fact as well as in theory, imposed a compulsory license?

VI. Post-*Paice v. Toyota*

Since the CAFC decided *Paice v. Toyota*, it has vacated a trial court's grant of a permanent injunction where the jury had already awarded a "reasonable royalty" that included an amount equal to what plaintiff's damages expert had testified would, over time, amount to a reasonable licensing fee.[55] In *Innogenetics N. V. v. Abbott Laboratories*, as in *Paice*, the plaintiff did not actually sell or distribute any products employing the patented invention, and had shown a willingness to license its patents. The patent holder made arguments at trial from which the court could infer that the jury's award of damages included the cost of future infringement.[56] The CAFC therefore vacated as an abuse of discretion the trial court's grant of a permanent injunction against future infringement, and remanded the case to the district court to determine the terms of a "compulsory license," such as conditioning sales on the payment of a running royalty.

CONCLUSION

So when under current law is a "compulsory" license not "compulsory"? Not, in the opinion of the *Toyota v. Paice* majority, when the equities of permitting the infringer to continue infringing the patent outweigh the interests of the patent owner, such as possible damage to the infringer's reputation.[57] And not, in the concurring opinion of Judge Rader, when the patent holder is given a chance to negotiate a royalty rate for future infringement before the court unilaterally imposes one, for "[w]ith such an opportunity in place, an ongoing royalty would be an ongoing royalty, not a compulsory license." [58]

And what of the Patent Clause of the Constitution? The *Toyota* majority pays lip service to the right to exclude others in asking whether permitting an infringer to use a patented invention in exchange for paying a royalty is properly characterized as preventing the violation of rights secured by a patent.[59] Although Judge Rader recognized exclusion of infringers as "the preferred remedy," he also recognized a patent owner's "opportunity to negotiate its own ongoing royalty [as providing at least] a minimal protection for its rights extending for the remainder of the patent term."[60]

Perhaps, in the end, the best that can be said is that the Constitution empowers Congress to grant inventors the exclusive rights to their inventions, but on its face does not require it.[61]

Endnotes

1 U.S. Const., art. I, § 8, cl. 8. We generally understand today that the Founders used "Science" to mean "knowledge," "useful Arts" to mean "science," and "Discoveries" to mean "inventions." Cf. Application of Joliot, 270 F.2d 954, 959 (C.C.P.A. 1959) (Rich, J., concurring) ("This might be a good time to point out that the purpose of the patent system is to promote the progress of the useful arts, which is done by granting patents for completed inventions.… It might also be well to note that, as pointed out in the House and Senate reports on the bill which became the Patent Act of 1952 (82nd Cong., House Report No. 1923, May 12, 1952, p. 4; Senate Report No. 1979, June 27, 1952, p. 3, to accompany H.R. 7794), it is quite unlikely that the authors of the Constitution had patents, inventors or inventions in mind when in Art. 1, section 8 they referred to 'science.' What we now know as 'science' did not exist when the Constitution was written.").

2 547 U.S. 388 (2006).

3 35 U.S.C. § 284 (emphasis added).

4 *Id.* § 283 (emphasis added).

5 eBay Inc. v. MercExchange L.L.C., 547 U.S. 388 (2006), emphasis added. In addition, the plaintiff must show: (1) that it has suffered an irreparable injury; (2) that the balance of hardships favors the plaintiff; (3) that the public interest "would not be disserved" by the injunction. *Id.*

6 35 U.S.C. § 284 (emphasis added).

7 *E.g.,* Paice LLC v. Toyota Motor Corp., No. 2:04-CV-211-DF, 2006 WL 2385139 at *1, *5 n.3 (E.D. Tex. Aug. 16, 2006) (denying injunction on grounds that "[i]rreparable harm lies only where injury cannot be undone by monetary damages" and stating that "sales of infringing products can be remedied via monetary damages in accordance with a reasonable royalty set by the jury.").

8 35 U.S.C. §§ 283, 284, *supra* nn. 3, 4.

9 Although the Court of Appeals for the Federal Circuit in eBay stated the "general rule that courts will issue permanent injunctions against patent infringement absent exceptional circumstances," 401 F.3d 1323, 1339 (2005), the point is actually in some dispute. According to one empirical study, "[permanent] injunctions are most commonly found in consent judgments and even formal settlements, most likely as a mechanism for formalizing the agreement. They are also sometimes employed in default judgments as a way of controlling an infringing party who has not presented himself in court. However, they are rare in adjudicated cases: only 19% of cases ending in trials and only 4% of those terminating in summary judgments included an injunction." Jay P Kesan. & Gwendolyn G.Ball, *How Are Patent Cases Resolved? An Empirical Examination of the Adjudication and Settlement of Patent Disputes*, 84 Wash. U. L. Rev. 237, 279 (2006), *available at* SSRN: http://ssrn.com/abstract=808347.

10 149 F.3d 1368, 47 USPQ 2d 1596 (Fed. Cir. 1998).

11 Wikipedia reports that the USPTO granted the first financial patent on March 19, 1799, for an invention for "Detecting Counterfeit Notes," the details of which were lost in the great Patent Office fire of 1836, and that the first financial patent for which any detailed written description survives is for "A Mode of Preventing Counterfeiting," granted Apr. 28, 1815. (This at least shows that counterfeiting has long been a recognized problem.) See http://en.wikipedia.org/wiki/Business_method_patent, visited Jan. 17, 2008. Perhaps the most famous "business method" patent in the world today is U.S. Pat No. 5,960,411, granted in Sept. 1999 to Amazon.com for its "one-click" Internet shopping method.

12 On February 15, 2008, the Federal Circuit set for May 8, 2008, a rehearing en banc in *In re* Bilski, No. 2007-1130, 2008 WL 417680 (Fed. Cir. Feb. 15, 2008) to determine the extent to which "business methods" are eligible for patent protection under U.S. law. Specifically, the Federal Circuit invited amici to address the following five sets of questions: (1) whether a claim of the application at issue addressed to a method practiced by a commodity provider for hedging the "consumption risks" associated with a commodity sold at a fixed price claims is patent-eligible subject matter under 35 U.S.C. § 101 ("Section 101"); (2) what standard should govern in determining whether a process is patent-eligible subject matter under Section 101; (3) whether the claimed subject matter is not patent-eligible because it constitutes an abstract idea or mental process, and when a claim that contains both mental and physical steps creates patent-eligible subject matter; (4) whether a method or process must result in a physical transformation of an article, or be tied to a machine, to be patent-eligible subject matter under Section 101; (5) whether it is appropriate to reconsider *State Street Bank & Trust Co. v. Signature Fin. Group, Inc.*, 149 F.3d 1368 (Fed. Cir. 1998), and *AT&T Corp. v. Excel Commun's, Inc.,* 172 F.3d 1352 (Fed. Cir. 1999) and, if so, whether those cases should be overruled in any respect. In *State Street Bank*, the CAFC had held that a method of transforming data representing discrete dollar amounts into a final share price was patentable where the claims recited computer processor means, storage means, and other means corresponding to an arithmetic logic unit. In *Excel*, the Court had found that claims directed to a method of generating a message record and including in the message record an indicator of a "primary interexchange carrier" for use in a telecommunications system were patent-eligible subject matter. *Excel* in turn had specifically referred to the Court's own en banc understanding expressed in *In re* Alappat, 33 F.3d 1526, 1543 (Fed. Cir. 1994), that the Supreme Court had "never intended to create an overly broad, fourth category of subject matter excluded from § 101," but rather simply "to explain a rather straightforward concept, namely, that certain types of mathematical subject matter, standing alone, represent nothing more than abstract ideas until reduced to some type of practical application, and thus that subject matter is not, in and of itself, entitled to patent protection." But in a dissent in *Lab. Corp. of Am. Holdings v. Metabolite Labs., Inc.*, 548 U.S. 124, 126 S.Ct. 2921, 2928 (2006), Justices Breyer, Souter, and Stevens suggested that the Federal Circuit may view patentability too broadly: "[State Street Bank] does say that a process is patentable if it produces a 'useful, concrete, and tangible result.' But this Court has never made such a statement and, if taken literally, the statement would cover instances where this Court has held the contrary." (citation omitted)

13 A patent holding company can be anything from a garage inventor who lacks the capacity to manufacture the patented product to a large company or foundation that holds many patents, sometimes buying them from others; in between are many a university or research institute. Among the most famous is Jerome Lemelson, who went from garage inventor to founding The Lemelson Foundation in roughly twenty years. *See* http://www.lemelson.org/about/bio_jerry.php, visited January 21, 2008. Supreme Court Justice Bradley both anticipated and disparaged the existence of patent holding companies over a century earlier, warning in *Atlantic Works v. Brady,* 107 U.S. 192, 200 (1882) that "an indiscriminate creation of exclusive privileges… creates a class of speculative schemers who make it their business to watch the advancing wave of improvement, and gather its foam in the form of patented monopolies, which enable them to lay a heavy tax upon the industry of the country, without contributing anything to the real advancement of the arts."

14 S. 1145, 110th Cong. (as reported out of S. Comm. on the Judiciary, July 19, 2007); H.R. 1908 (as passed by the H. of Reps., Sept. 7, 2007).

15 S. 1145, § 4; H. R. 1908, § 5.

16 401 F.3d 1323 (Fed. Cir. 2005).

17 126 S. Ct. 1837.

18 126 S. Ct. at 1839.

19 126 S. Ct. at 1839-40.

20 126 S. Ct. at 1841-42 ("a page of history is worth a volume of logic") (quoting N.Y. Trust Co. v. Eisner, 256 U.S. 345, 349 (1921)).

21 *eBay*, 126 S. Ct. at 1842 (Kennedy, J., concurring).

22 *Id*.

23 *Id*.

24 In other areas of the law, the Supreme Court has long recognized the equitable ability of a court to hear a suit based on "future damages" in order to avoid the burden of multiple future actions. *E.g.,* Root v. Lake Shore & M.S. Ry. Co., 105 U.S. 189, 207 (1881); Great-West Life & Annuity Ins. Co. v. Knudson, 534 U.S. 204, 211 (2002) (quoting Bowen v. Massachusetts, 487 U.S. 879, 918 (1988) (Scalia, J., dissenting)).

25 *See, e.g.,* Colleen Chien, *Cheap Drugs at What Price to Innovation: Does The Compulsory Licensing Of Pharmaceuticals Hurt Innovation?,*18 Berkeley Tech. L.J. 853, 862-64 & n.35 (2003) (noting that "courts have emphatically resisted compulsory licenses merely because a patent holder chooses not to use her invention"). For an interesting discussion of conflicting policy arguments for and against compulsory licensing of patents, sometimes by the same party, see James Love, "Brazil Puts Patients Before Patents," (May 4, 2007), *at* http://www.huffingtonpost.com/james-love/brazil-puts-patients-befo_b_47651.html; James Love, "Recent Examples of the Use of Compulsory Licenses on Patents," KEI Research Note 2001:2[1] (Mar. 8, 2007, revised May 6, 2007), *at* http://www.keionline.org/index.php?option=com_content&task=view&id=41&Itemid=1; and James Love, "Abbott Recently Sought Compulsory License in U.S. Patent Dispute," (May 1, 2007), *at* http://www.keionline.org/index.php?option=com_content&task=view&id= 43&Itemid=1.

26 *See* IMX, Inc. v. Lending Tree, LLC, 469 F. Supp. 2d 203, 226 (D. Del. 2007) (referring to compulsory license as a possible result); Transocean Offshore Deepwater Drilling, Inc. v. GlobalSantaFe Corp., No. H-03-2910, 2006 WL 3813778, at *5 (S.D. Tex. Dec. 27, 2006) (failure to enter a permanent injunction "will force a compulsory license").

27 *eBay*, 126 S. Ct. at 1840.

28 *See, e.g.,* Sundance, Inc. v. DeMonte Fabricating Ltd., No. 02-73543, 2007 WL 37742 at *2 (E.D. Mich. Jan. 4, 2007); IMX, 469 F. Supp. 2d at 225 (successful licensing weighs against injunction); z4 Technologies, Inc. v. Microsoft Corp., 434 F. Supp. 2d 437, 440-442 (E.D. Tex. 2006), *but see* Commonwealth Scientific & Industrial Research Organisation v. Buffalo Technology, Inc., No. 6:06-324, 2007 U.S. Dist. LEXIS 43832 (E.D. Tex. June 15, 2007) (granting permanent injunction in favor of patent holding company).

29 *See Transocean*, 2006 WL 3813778 at *3 (defendant "has not cited any case in which a continuing infringer in direct competition with a patent holder has not been permanently enjoined from using the patented invention"); Tivo Inc. v. Echostar Commun's Corp., 446 F. Supp. 2d 664, 669-70 (E.D. Tex. 2006) (*rev'd in part by* No. 2006-1574, 2008 WL 249155 (Fed. Cir. Jan 31, 2008)) (direct competition weighs in favor of an injunction; on reversal and remand, awarding additional damages for period that permanent injunction was stayed pending appeal).

30 126 S.Ct at 1842; *see supra* note 21.

31 S. 1145, § 4; H.R. 1908, § 5, *supra* note 15.

32 469 F. Supp. 2d at 225.

33 434 F. Supp. 2d at 440.

34 *See e.g.,* Paice LLC v. Toyota Motor Corp., 504 F.3d 1293, 1316 (Fed. Cir. 2007); *but see* Tull v. United States, 481 U.S. 412, 426 (1987); Swofford v. B&W, Inc., 336 F.2d 406, 412-413 (5th Cir. 1964).

35 127 F.3d 1462, 1468 (Fed. Cir. 1997).

36 *Swofford*, 336 F.2d at 410.

37 *See, e.g., id.*, 336 F.2d at 411 (recognizing that a court could award "compensatory damages incidental to an injunction to avoid multiplicity of suits"); *see also,* Georgia-Pacific Corp. v. U.S. Plywood Corp., 243 F. Supp. 500, 517 (S.D.N.Y. 1965) (discussing Patent Act of 1870).

38 No. 2:04-CV-211-DF, 2006 WL 2385139, at *5 (E.D. Tex. Aug. 16, 2006).

39 In part, the district court ordered that Toyota pay quarterly royalties, accompanied by an accounting of infringing product sales, beginning three months after the judgment was signed; that the first payment include royalties for all infringing vehicles sold and not accounted for in the jury's verdict; that payments not made within fourteen days of the due date accrue interest at 10%, compounded monthly; and that *Paice* could request audits, although "[i]t is anticipated that the parties may wish to agree to more comprehensive and convenient terms." *Paice*, 504 F.3d at 1314.

40 *Id*. at 1302.

41 *Id*. at 1303.

42 *Id*.

43 *Id*. at 1293.

44 The CAFC seems perhaps unduly uncharitable towards the District Court, which manifestly had based its ongoing royalty on what the jury had found reasonable for past infringement. In its August 16, 2006 Order, the District Court noted that the jury had, "based on the entire record, determined an appropriate reasonable royalty rate that can be easily calculated on future sales of the accused devices thereby removing uncertainty from future damages calculations," Paice LLC v. Toyota Motor Corp., No. 2:04-CV-211-DF, 2006 WL 2385139, at *5 (E.D. Tex. Aug. 16, 2006). This suggests at least implicitly that (a) the District Court found the jury's determination of a reasonable royalty for past infringement adequately supported by the record at trial, and (b) the District Court believed that the same royalty should apply to both past and future infringement. What the CAFC seems to criticize is step (b), the District Court's failure to explain why Toyota should compensate Paice for past and future infringement at the same rate. *Cf. Paice,* 504 F.3d at 1315 ("Upon remand, the court may take additional evidence if necessary to account for any additional economic factors arising out of the imposition of an ongoing royalty."). But perhaps this was merely an excuse to let the parties negotiate their own royalty rate, as the concurrence had suggested. *See Paice,* 504 F.3d at 1315, n.15 ("This process will also, presumably, allow the parties the opportunity to present evidence regarding an appropriate royalty rate to compensate Pace and the opportunity to negotiate their own rate prior to the imposition of one by the court, as the concurrence suggests.").

45 504 F.3d at 1314 (citing Joy Techs. v. Flakt, Inc., 6 F.3d 770, 777 (Fed. Cir. 1993) (for the proposition that non-infringing acts may not be enjoined)).

46 *Id*. (emphasis added).

47 Although the parties in that case had contested the amount of the royalty, which the court had styled a "compulsory license," the parties did not dispute the district court's authority to impose it. *See* Shatterproof Glass Corp. v. Libbey-Owens Ford Co., 758 F.2d 613, 628 (Fed. Cir. 1985).

48 504 F.3d at 1314 (citing United States v. Glaxo Group Ltd., 410 U.S. 52, 59 (1973)).

49 *Id*. at 1314-15 (emphasis added).

50 Under the copyright laws, musicians who record and distribute their works thereby give up exclusive performance rights. 17 U.S.C. §115 provides "When phonorecords of a nondramatic musical work have been distributed... under the authority of the copyright owner, any other person... may, by complying with the provisions of this section, obtain a compulsory license to make and distribute phonorecords of the work."

51 504 F.3d at 1313 n. 13.

52 *Id*. at 1316.

53 *Id*.

54 *Id*.

55 Innogenetics N.V. v. Abbott Labs., 512 F.3d 1363 (Fed. Cir. 2008).

56 512 F.3d 1363, at *12-13.

57 504 F.3d at 1303.

58 *Id*. 1316.

59 *Id*. at 1313 n.13; *see supra* note 45**.**

60 *Id*. at 1317.

61 "The exclusive right which Congress is authorized to *secure* to authors and inventors owes its existence solely to the acts of Congress securing it, from which it follows that the rights granted by a patent or copyright are subject to such qualifications and limitations as Congress, in its unhampered consultation of the public interest, sees fit to impose." The Constitution of the United States of America, Analysis and Interpretation, Sen. Doc. No. 92-82 (Library of Congress, 1972) at 317 (citing Sears, Roebuck and Co. v. Stiffel Co., 376 U.S. 225 (1964)); Compco Corp. v. Day-Brite Lighting, Inc. 376 U.S. 234 (1964); Wheaton v. Peters, 8 Pet. (33 U.S.) 591, 662 (1834); Evans v. Jordan, 9 Cr. (13 U.S.) 199 (1815).

Antitrust Implications of Patent Settlements: Balancing Patent Policy, Antitrust Law, and the Practical Limits of Litigation

*By Michael K. Friedland**

In a perfect world, all patents would be valid and none infringed. In a near-perfect world, a mechanism would exist to rapidly and efficiently determine whether a patent is valid and infringed. We live in neither world. Our method—litigation—is not rapid or efficient. When disputes arise, parties litigate, and then, most often, negotiate and settle.

Resolution by settlement is not unique to patent disputes. However, it has a unique feature. The right to exclude competition is a central right of patent owners. Very often, the competitor will concede in the terms of the settlement to discontinue selling its product. In entering such an agreement, the parties raise issues of patent and antitrust law. They have served the interest of the patent law; they have promoted "the progress of science and the useful arts" by securing for the patent owner "the exclusive right" to its invention.[1] However, they have also implicated an issue of antitrust law: two competitors have reached an agreement that might be characterized as a restraint of trade.

When confronted with such agreements, courts have attempted to fashion an appropriate test to distinguish between legitimate settlements and those antitrust violations masquerading as settlements. In doing so, the courts have recognized that traditional antitrust rules are not well-suited to the task. In their stead, courts have created new tests that seek a balance between patent and antitrust interests. At the same time, they have been mindful of a critical and practical consideration: the judiciary cannot afford to discourage settlements.

I. Patent Settlement Agreement: Pretext for Collusion?

Patent law and antitrust law have the same goal ultimately: to promote economic growth. The former seeks to achieve that goal by rewarding innovators with monopoly-type rights. Under patent law, an inventor has "the right to exclude others from making, using, offering for sale, or selling [his] invention throughout the United States or importing the invention into the United States" for a limited term of years.[2]

Antitrust law, on the other hand, seeks to achieve the same goal by prohibiting behavior that would interfere with competition. The Sherman Act prohibits "[e]very contract, ... or conspiracy, in restraint of trade or commerce among the several States"[3] and "monopoliz[ation] or attempt[s] to monopolize, or combin[ations], or conspir[acies]... to monopolize any part of the trade or commerce among the several States."[4]

As the Eleventh Circuit put it, when evaluating patent settlements, "a delicate balance must be drawn between the two regulatory schemes."[5] If the interests of antitrust law were ignored, patent law could be used as a pretext for collusion. A company could use an invalid patent for, among other things, cover for a price-fixing scheme.

> Suppose a seller obtains a patent that it knows is almost certainly invalid (that is, almost certain not to survive a judicial challenge), sues its competitors, and settles the suit by licensing them to use its patent in exchange for their agreeing not to sell the patent produce for less than the price specified in the license. In such a case, the patent, the suit, and the settlement would be devices—masks—for fixing prices....[6]

Ignoring the interests of patent law would lead to an equally undesirable result. Under patent law, and the Constitution, patent owners are granted the exclusive right to exploit their inventions. A patent owner wanting to exercise that right would not have the option of obtaining a settlement that includes an agreement by its competitor to withdraw from the market. Because such a settlement would be too vulnerable to antitrust challenge, the patent owner would be forced to litigate his patent suit to final judgment or give up on his exclusive right.

> Requiring parties to a lawsuit to either litigate or negotiate a settlement in the public interest... is, as a practical matter, tantamount to establishing a rule requiring litigants to continue to litigate when they would prefer to settle and to act as unwilling private attorneys general and to bear the various costs and risks of litigation."[7]

II. Balancing Patent and Antitrust Law

Traditional antitrust law tests, such as the per se rule and the rule of reason, are difficult to adapt to evaluation of patent settlements because they do not take into account a patent owner's legitimate right to exclude. As stated by the Eleventh Circuit in *Schering-Plough Corporation v. FTC*,[8] both the per se rule and the rule of reason are

> ill-suited for an antitrust analysis of patent cases because they seek to determine whether the challenged conduct had an anticompetitive effect on the market. By their nature, patents create an environment of exclusion, and consequently, cripple competition. The anticompetitive effect is already present.[9]

Early last century, in *Standard Oil Co. v. United States*,[10] the Supreme Court stated that the starting point in establishing a balance between the two interests is to determine whether the dispute between the patent owner and the competitor is legitimate.[11] Courts since have generally followed this principle. In *In re Tamoxifen Citrate Antitrust Litigation,*[12] the Second Circuit stated, "[u]nless and until the patent is shown to have been procured by fraud, or a suit for its enforcement is shown to be objectively baseless, there is no injury to the market cognizable under existing antitrust law, as long as competition is restrained only within the scope of the patent."[13]

Sitting by designation, Judge Posner set forth a test for determining whether a suit is objectively baseless.[14] The test focuses on whether "a neutral observer would reasonably think either that the patent was almost certain to be declared invalid, or the defendants were almost certain to be found not to have infringed."[15]

** Michael K. Friedland is a partner of Knobbe, Martens, Olson & Bear, LLP in Irvine (CA). The views expressed here are his own.*

In reviewing settlement agreements, courts have generally been willing to give the settlement agreements the benefit of the doubt. In *Tamoxifen,* a settlement agreement was reached after a district court had invalidated a patent, and while the judgment was on appeal to the Federal Circuit. The Second Circuit held that the settlement was legitimate. The court found that the risk of reversal on appeal was sufficient to justify the agreement. "There is a risk of loss in all appeals that may give rise to a desire on the part of both the appellant and the appellee to settle before the appeal is decided."[16]

III. Reverse Payments

An additional issue is raised by a settlement agreement known as a "reverse payment" settlement or "exit payment" settlement. In a reverse payment settlement, the competitor agrees to discontinue making and selling the accused product. As part of the bargain, the patent owner agrees to make a payment to the competitor. Such payments can be very large, and they are most common in the pharmaceutical context.[17]

Courts have acknowledged that such settlements may appear to be "suspicious" on their face.[18] At the same time, they have not been willing to create legal tests that would make it easier to subject these settlements to antitrust liability.[19] In *In re Tamoxifen Antitrust Litigation*, the Second Circuit held that a reverse payment settlement does not violate antitrust law "so long as the patent litigation is neither a sham or otherwise baseless."[20]

The Second Circuit in *In re Tamoxifen* also dismissed the suggestion that an antitrust violation should be found where the amount of the reverse payment exceeds the profit the competitor could have earned had it continued to manufacture the accused product. The Second Circuit stated that a large reverse payment might betray the patent owner's doubts regarding its ability to prevail on the merits in its case. According to the court, however, those doubts do not mean that the litigation is a sham or baseless.

> Of course, the law could provide that the willingness of the patent holder to settle at a price above the generic manufacturer's projected profit betrays a fatal disbelief in the validity of the patent or the likelihood infringement, and that the patent holder therefore ought not to be allowed to maintain its monopoly position. Perhaps it is unwise to protect patent monopolies that rest on such dubious patents. But even if large reverse payments indicate a patent holder's lack of confidence in its patent's strength or breadth, we doubt the wisdom of deeming a patent effectively invalid on the basis of a patent holder's fear of losing it.[21]

The court also reasoned that placing a cap on the amount of a reverse payment would only benefit the patent owner.

> We are unsure, too, what would be accomplished by a rule that would effectively outlaw payments by patent holders to generic manufacturers greater than what the latter would be able to earn in the market were they to defend successfully against an infringement claim. A patent holder might well prefer such a settlement limitation—it would make such a settlement cheaper—while a generic manufacturer might nonetheless agree to settle because it is less risky to accept in settlement all the profits it expects to make in a competitive market rather than first to defend and win a lawsuit, and then to enter the marketplace and earn the profits. If such a limitation had been in place here, [the patent owner] might have saved money by paying [the competitor] the maximum such a rule might allow.... But the resulting level of competition, and its benefit to consumers, would have been the same. The monopoly would have nonetheless endured—but to no apparent purpose, at less expense to [the patent owner] and less reward for [the competitor].[22]

The court rejected the argument that permitting the reverse payment settlement would disserve the public interest in having the validity of patents litigated. As the court observed, "[t]he Settlement Agreement was a virtual invitation to other generic manufacturers" to challenge the validity of the patent.[23]

In *Schering-Plough,* the Eleventh Circuit rejected the Federal Trade Commission's assertion that reverse payments settlement agreements violate section 1 of the Sherman Act and section 5 of the Federal Trade Commission Act.[24] In reaching this conclusion, the court acknowledged that some settlements could result in economic inefficiency, but recognized that the alternative to settlement is not a rapid and efficient method of determining whether a patent is valid or infringed. Rather, it is the ordeal of patent litigation.

> Patent litigation breeds a litany of direct and indirect costs, ranging from attorney and expert fees to the expenses associated with discovery compliance. Other costs accrue for a variety of reasons, be it the result of uncompromising legal positions, differing strategic objectives, heightened emotions, lawyer incompetence, or sheer moxie.[25]

A rule that encouraged patent litigation, the court reasoned, could end up hindering innovation. "[T]he caustic environment of patent litigation may actually decrease product innovation by amplifying the period of uncertainty around [an inventor's] ability to research, develop, and market the patented product or allegedly infringing product."[26]

Thus, the court reasoned that, although, a rule too liberal in allowing settlements might not be perfect, such a rule would be preferable to the next-best alternative, litigation. "The intensified guesswork involved with lengthy litigation cuts against the benefits proposed by a rule that forecloses a patentee's ability to settle its infringement claim."[27]

CONCLUSION

Patent settlements raise unique issues for federal courts. Courts are called upon to question whether a settlement is legitimate or a sham. A rule that is too permissive risks allowing competitors to collude, using patent litigation as cover. A rule that is too restrictive risks forces competitors to litigate against their will, potentially discouraging innovation and burdening an already taxed judicial system with a new species of litigation. Although some circuits have weighed in, these issues seemed destined to reach the Supreme Court. Until that time, the risk of government and private antitrust enforcement will cause parties to be cautious in crafting terms of settlement agreements.

Endnotes

1 *See* U.S. Const,, art. I, § 8.

2 35 U.S.C. § 154(a)(1)-(2).

3 15 U.S.C. § 1.

4 15 U.S.C. § 2.

5 Schering-Plough Corp. v. FTC, 402 F.3d 1056, 1067 (11th Cir. 2005)).

6 *Id.*

7 *In re* Ciprofloxacin Hydrochloride Antitrust Litigation, 363 F. Supp. 513, 532 (E.D.N.Y. 2005) (citations and internal quotations omitted).

8 402 F.3d 1056 (11th Cir. 2005).

9 *Id.* at 1065-66 (citing Valley Drug Co. v. Geneva Pharms., 344 F.3d 1294, 1312 (11th Cir. 2003). *Cf. In re* Cardizem CD Antitrust Litigation, 332 F.3d 896 (6th Cir. 2003) (an interim agreement in which the brand name patent holder agreed to pay a potential generic entrant to delay entry pending resolution of a lawsuit was per se illegal).

10 283 U.S. 163 (1931).

11 *Id.* at 171 ("Where there are legitimate conflicting [patent] claims, ... a settlement by agreement, rather than litigation, is not precluded by the Sherman Act.")

12 Joblove v. Barr Labs, Inc. *(In re* Tamoxifen Citrate Antitrust Litigation*)*, 429 F.3d 370 (2nd Cir. 2005), *reprinted as amended*, 466 F.3d 187 (2nd Cir. 2006), *cert. denied*, 127 S. Ct. 3001 (2007).

13 *Id.* at 213 (quoting *In re* Ciprofoxacin Hydrochloride Antitrust Litigation¸ 363 F. Supp.2d 514, 535 E.D.N.Y. 2005).

14 Ashai Glass Co., Ltd. v. Pentech Pharmaceuticals, Inc., 289 F. Supp.2d 986, 991 (N.D. Ill. 2003) (Posner, J. sitting by designation)

15 *Id.* at 993.

16 *In re* Tamoxifen Citrate Antitrust Litigation, 466 F.3d at 205.

17 *See In re* Tamoxifen Citrate Antitrust Litigation, 466 F.3d at 206 ("[R]everse payments are particularly to be expected in the drug-patent context because the Hatch-Waxman Act created an environment that encourages them").

18 *Id.* at 208.

19 *Cf. In re Cardizem*.

20 *In re* Tamoxifen Citrate Antitrust Litigation, 466 F.3d at 208.

21 *Id.* at 210.

22 *Id.* at 211.

23 *Id.* at 212, n.25. Reverse payment settlement agreements in the pharmaceutical industry have significantly increased since the Eleventh Circuit's *Schering-Plough* decision. In 2004, none of the 14 patent settlement agreements filed with the Federal Trade Commission included reverse payment provisions. *See Agreements Filed with the Federal Trade Commission Under the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Summary of Agreements Filed in FY 2006*, A Report by the Bureau of Competition at 4. By 2005, three of 11 settlements contained such payments. *Id.* By 2006, that number had increased to 14 of 28 settlements. *Id.*

24 *Schering-Plough Corp.*, 402 F.3d at 1076.

25 *Id.* at 1075.

26 *Id.*

27 *Id.*

Saying What the Law Is: Arguments for an ICJ That Is Less Deferential to Security Council and General Assembly Resolutions

*By Mark Angehr**

The United Nations (UN) system operates differently than the three branches of American government. Its political organs, the Security Council ("Council") and the General Assembly ("Assembly"), cannot make law. Its judicial organ, the International Court of Justice (ICJ), has no power of full-scale judicial review over the resolutions of the political organs. Instead of checks and balances, the United Nations Charter proposes a system of shared responsibility and common political agenda. Such intra-organizational cooperation may help advance United Nations policy, but it threatens the judicial independence of the ICJ. When the political organs pass resolutions that resemble legislation or contain interpretations of international law, the ICJ cannot maintain its institutional legitimacy without to some extent reviewing these legal interpretations in cases before it. This article explores the legal contours of ICJ review over the factual and legal determinations underlying Council and Assembly resolutions. This situation arises when the ICJ must rule, in accordance with international law, on a dispute that the Council or Assembly have already treated in their political processes. As the political organs flex their muscle with quasi-legislative and quasi-judicial resolutions, the ICJ must reassert itself by reviewing the legal and factual determinations underlying those resolutions in the course of deciding its cases.

The ICJ should not undermine the political organs' roles in maintaining international security, but at the same time it should not exhibit excessive deference to the judicial determinations of political processes. The political process is not suitable for adjudication of legal obligations under international law. That role should be left to the ICJ, which has the safeguards in place to ensure the proper adjudication of the rights and obligations of parties before it. An ICJ that is less deferent to the political agenda of the Council and Assembly will be more legitimate in the eyes of Member States. This increased legitimacy will provide a better enforcement mechanism against countries that violate international legal norms, and thus contribute to the healthy functioning of the United Nations system.

The ICJ is the "principal judicial organ" of the UN.[1] In addition to deciding contentious disputes between States, the ICJ possesses an advisory jurisdiction, under which it considers legal questions received from the Council and Assembly.[2] The UN Charter mandates that the ICJ decide these legal questions in accordance with international law.[3] The ICJ's advisory jurisdiction has become a dustbin for intractable political and humanitarian conflicts that the political organs have failed to solve with their own resolutions. Therefore, to answer an advisory opinion request, the ICJ often deals with resolutions closely related to the underlying legal question. The Court should retain its legitimacy as a judicial organ, but nonetheless further the political goals of the larger UN system. It should scrutinize the legal determinations of the UN's political organs or risk lending its judicial imprimatur to decisions based on political, non-judicial processes.

The judicial autonomy and legitimacy of the ICJ are centrally important to the healthy functioning of the UN system. The structural principles embedded in the UN Charter mandate that the ICJ not cede control over judicial determinations to political organs guided by the national interests of its Member States. Consequently, the ICJ must reclaim its judicial independence in the exercise of its advisory jurisdiction by rehabilitating its fact-finding capabilities and ensuring the correct application of international law to the specific factual situation before it. Regardless of previous legal action taken by the political organs, the ICJ must bring to bear on international disputes the inherent advantage of the judicial process—namely, an adversarial process to find true facts and the ability to ensure the correct application of international law.

The ICJ's recent advisory opinion on Israel's construction of a security wall in the disputed West Bank illustrates the danger to the Court's institutional legitimacy posed by cases that the Council and Assembly have already dealt with in their political processes.[4] In this opinion, the ICJ held that the wall violated international humanitarian law and international human rights law, and ordered Israel to remove the portions of the wall located in the West Bank.[5] However, Council and Assembly resolutions had already reached the same legal conclusion before the request for the advisory opinion. Just nineteen days before the submission of the advisory request, the Assembly passed Resolution 10/13, declaring the wall in violation of international law and ordering its removal.[6] A lengthy dossier submitted with the advisory request formed the factual basis of the ICJ's decision. This dossier included relevant Council and Assembly resolutions, as well as UN-commissioned fact-finding reports. Aside from written statements mainly opposing the Court's exercise of jurisdiction, the ICJ did not gather evidence outside of this dossier. The Wall Opinion illustrates the risks to the Court's institutional legitimacy, as well as to international law as a whole, when the ICJ defers to the factual and legal determinations of the political organs. If the ICJ is to expand its role in fact-intensive disputes such as those involving international human rights, it should increase its fact-finding capacity so that it may act less like an appellate body and more like a trial court of first instance.

....

** Mark Angehr is a student at Northwestern University School of Law.*
He would like to thank John O. McGinnis and Rahul Kolhatkar for their invaluable comments in preparing this article. An expanded version of this article will appear in a forthcoming issue of the Northwestern Law Review.

Organizational Dynamic of the ICJ's Advisory Jurisdiction

The ICJ is largely modeled on its predecessor court, the Permanent Court of International Justice (PCIJ), established by the League of Nations.[7] However, unlike the PCIJ, which was not formally part of the League of Nations, the ICJ is a principal organ of the UN as well as the UN's principal judicial organ.[8] Only States may be parties in cases before the fifteen-member Court, though the State need not be a member of the UN in order to appear.[9] Member States may request that the Court exercise jurisdiction over any dispute involving interpretation of a treaty or international law, or the "existence of any fact which, if established, would constitute a breach of an international obligation."[10] Once jurisdiction has been established, the Court must decide disputes in accordance with international law, which is limited to international conventions, custom, and general principles of law.[11]

The Assembly and the Council are authorized to submit advisory opinion requests to the ICJ on "any legal question," which the Court has broadly construed to include complex factual disputes or political issues.[12] The advisory opinion request must be "accompanied by all documents likely to throw light upon the question."[13] The advisory opinion, while truly a peculiar notion to federal courts in the United States, is permitted in many U.S. courts.[14] However, the advisory jurisdiction as exercised in the World Court differs from the practice in the United States of a state legislator requesting a court's opinion on the constitutionality of a proposed law.[15] The ICJ's advisory opinions have often involved hotly debated political disputes[16] and legal questions embedded in broader bilateral disputes.[17] State consent, while required for the exercise of contentious jurisdiction, is not required for the ICJ to exercise advisory jurisdiction over a dispute.[18]

The ICJ's status as "principal judicial organ" of the UN has been characterized as an "organic link" to the shared goals of the UN system.[19] The ICJ, like all other principal organs in the UN system, has a duty to further the purposes and principles of the UN These purposes are to "maintain peace and security," and "take collective measures for the prevention and removal of threats to the peace."[20] The advisory function of the ICJ, even more than its contentious jurisdiction, serves as a vehicle for the Court's participation in the "Purposes and Principles" of the UN Charter.[21]

Proponents of the advisory jurisdiction argue that by rendering advisory opinions, the Court is able to place another organ's operation upon a firm and secure foundation. Judge Bedjaoui has written that the Court's advisory function assists the political organs by taking into account "its preoccupations or difficulties and by selecting, from all possible interpretations of the Charter, the one which best serves the actions and objectives of the political organ concerned."[22] In the Wall Opinion, the Court explained that its obligation to clarify a legal issue for the Assembly outweighed any concerns about the judicial propriety of adjudicating an ongoing political dispute and armed conflict between Israel and Palestine.[23] Accordingly, the Court stressed the organizational purpose of the advisory opinion: "The Court's Opinion is given not to the States, but to the organ which is entitled to request it."[24] The ICJ characterized the opinion as that which "the General Assembly deems of assistance to it in the proper exercise of its function."[25] Accordingly, the Court placed the matter "in a much broader frame of reference than a bilateral dispute," as it was "of particularly acute concern to the United Nations."[26]

The Court is strongly inclined to not only answer a request for an advisory opinion, but to facilitate the larger aims of the UN by arriving at a conclusion in line with the preference of the political organ.[27] Judge Azevedo has stated that the Court "must do its utmost to co-operate with the other organs with a view to attaining the aims and principles that have been set forth."[28] The closer the institutional connection of the ICJ to the requesting organ, he argues, the greater the usefulness of that opinion to the operation of the requesting organ. However, the advisory function threatens the institutional legitimacy of the Court because it often resolves disputes without the consent of the relevant States,[29] and the political organ making the request has often already ruled on the issue.[30]

Organizational theory helps to explain why the ICJ is not functioning as a check on the actions of the political organs in its advisory jurisdiction. By examining the benefits and drawbacks of coordination among organizations and within organizations, organizational theory predicts the most efficient modes of cooperation.[31] Studies of coordination mechanisms within organizations suggest that the ICJ is likely motivated to undertake advisory opinions out of a fear of institutional isolation and marginalization.[32] An organization might "seek[] to forestall or prevent future crisis which may imperil its success or even continuation."[33] Because organizations have incentives to increase their authority and prestige, the Court is unlikely to decline the opportunity to contribute to the progress of international law by rendering an advisory opinion.[34] Given the institutional incentives for rendering advisory opinions, the ICJ will continue to do so as long as the perceived benefits of cooperation outweigh the loss in judicial autonomy.[35] Similarly, the political organ will make the request as long as the perceived advantage to its operations outweighs any loss to its political autonomy.

The ICJ's reliance on the political organs to enforce compliance with its decisions incentivizes the Court not only to take on advisory opinions, but to give opinions in accordance with the political preferences of the requesting organ. The main impediment to coordination between the ICJ and the political organ is the line between cooperation and competition. If the degree of interdependence is high, and the degree of antagonism is high, the result will be competition and conflict.[36] By contrast, if the degree of interdependence is high, and the degree of antagonism is low, the result will be cooperation. The ICJ has an incentive to reduce competition and increase smooth cooperation in order to avoid alienating the requesting organ and risking institutional isolation.

If we map the interaction of the ICJ and the Assembly in the Wall Opinion onto this organizational dynamic, we see a high level of interdependence due to their "organic link" and a low level of antagonism due to the Court's incentive to contribute to the shared goals of the UN as reflected in the stated policy preference of the Assembly. The resultant "cooperation" between the two organs reduces the need for information processing and furthers the shared mission of the UN. By systematizing coordination through a process that provides the Court with "an exact statement of the question" as well as a "voluminous dossier"[37] of documents "likely to throw light on the question,"[38] the Court is unlikely to conduct its own investigation outside of the given universe of documents. From an organizational theory perspective, the Court will not engage in its own extensive review of the background material and facts, because such a duplicative inquiry would bring the Court into competition with the functioning of the requesting organ. In relying on the resolutions and factual studies made by the political organs, the likelihood that the Court will render an opinion in line with the policy preferences of the political organ is thus greater.

The results of such a model have been borne out in the Court's case law. In 1949, the Court held in an advisory opinion that South Africa had no legal obligation to place its mandate, South West Africa (now Namibia), under a trusteeship with the UN[39] The Assembly had advocated for South Africa's withdrawal from South West Africa, but the Court found in favor of South Africa's continued occupation. The opinion weakened the Court's credibility, especially among African nations.[40] The loss of political capital to the Assembly outweighed any potential benefit of further coordination with the Court on the issue, and, as a result, the Assembly never revisited the issue with the Court.

Then, in 1971, the Council requested an advisory opinion on the "legal consequences" of South Africa's continued presence in Namibia.[41] The request was seen as an opportunity for the Court to "redeem its impaired image," since its advisory jurisdiction had been unused since 1962.[42] The Council had in fact already passed Resolution 276, which strongly condemned the "illegal" presence of South Africa in Namibia.[43] The Court in this iteration of coordination produced an opinion in line with the clear political preference of the Council by holding that South Africa's presence in Namibia was illegal.[44] The Court's interaction with the Council was thus cooperative, and in rendering an opinion that mirrored the effect of the Council's resolution on the issue, the Court avoided conflict with the political organ. The Court consequently repaired its image and staved off institutional marginalization by indicating its willingness to cooperate with the political organs.

Although this coordination effect has positive value as an explanation of the ICJ's behavior, it should not be seen as normative. The ICJ overestimates the institutional benefits it receives from such coordination. The fear of institutional isolation motivates the ICJ to defer to the political organ, but there is little evidence that behaving in such a way increases in the long-term the number of advisory requests that the Court receives. If the Court were correct in the assumption that advisory opinions deferent to the preferences of the political organs lessen the court's marginalization and increase the volume of its advisory jurisdiction caseload, there would be an increase in advisory opinions after the ICJ rendered a deferent advisory opinion. Although advisory requests two and four years later followed the deferent South West Africa opinion, a statistical breakdown of the Court's advisory docket shows no long-term changes in the number of opinions rendered from its first opinion in 1947 to its last in 2004. The Court averages about four advisory opinions a decade. As of 2008, the Court has not received another advisory request since the Wall Opinion, and it would appear that the Court will have a below-average number of advisory opinions this decade, despite the accommodation it provided the Assembly in the cooperative Wall Opinion.

While the ICJ is concerned about institutional marginalization and orders its behavior in rendering advisory opinions accordingly, the motivation of the political organs in requesting advisory opinions proves to be more complex. First, the Council or Assembly may refer a dispute to the ICJ's advisory jurisdiction when the intractability of the dispute does not lend itself to political resolution. Second, a referral to the ICJ's advisory jurisdiction can take place if the particular dispute is susceptible to judicial resolution, that is, if the ICJ can help the organ overcome a political impasse by settling a question of international law. Third, if the political organ doubts the utility of the advisory opinion it will receive, or if it fears an opinion not in line with its political preferences, it can take steps to make known its preferences before the Court composes its opinion. Therefore, the political organ's perception of the ICJ's propensity to render an opinion not in line with the organ's political preference is just one of three factors that determine when the ICJ will be asked to exercise its advisory jurisdiction. The Court's fear of marginalization is thus overblown; the factors determining when the organs refer a dispute to its advisory jurisdiction depend more on the peculiar nature of the dispute itself than on the Court's perceived deference to the political will of the Council or Assembly. In other words, the Assembly's decision to refer to the ICJ the question of the legality of the wall in Palestine depended more on the exigencies of that particular situation—namely, the need for a legal and not political resolution—than on the ICJ's recent record of deference to the Assembly in its advisory jurisdiction.

In light of the cost in loss of judicial autonomy and reduced institutional benefits, a new calculation shows that the Court should defer less to the requesting organ. The Court should thus be more competitive by undertaking its own fact-finding and by rendering decisions that may not line up with the political preferences of the requesting organ. The result of such an undertaking is more independent and legitimate advisory opinions. As more authoritative statements of the law, the opinions would provide a better enforcement mechanism against the political organs to police the behavior of States that have violated their legal obligations. By asserting its jurisdiction over fact-finding and legal interpretation, the ICJ would signal to the requesting organ that the function each organ was to perform had changed. In the long-term, the functional differentiation of each organ would shift to accommodate the Court's new role, and the organs could ultimately resume a cooperative interaction. The political organ would continue to request opinions, because the benefit of receiving truly independent advisory opinions would outweigh the risk of an opinion not in line with its political preference. A revitalized advisory jurisdiction could aid the political organs in providing another strong enforcement mechanism against States that violate international norms.

This model has the additional advantage of better serving the shared goals of the UN system. In reclaiming its judicial autonomy within its advisory jurisdiction, the Court is aiding the UN's settlement of international disputes "in conformity with the principles of justice and international law."[45] In contrast, an opinion that reproduces the politically-determined legal conclusion of the requesting organ does not further this goal, because it abdicates judicial responsibility to a political organ.

The ICJ's Review of Assembly and Council Resolutions in the Exercise of its Advisory Jurisdiction

The Charter contains no provisions directly addressing the ICJ's review, interpretation, or invalidation of Council and Assembly resolutions. The legal effects of these resolutions must therefore be interpreted from structural principles in the Charter, the competencies of the organs, and the case law of the ICJ. Although there is no hierarchical relation between any of the principal organs of the UN, the ICJ is the principal "judicial" organ,[46] and its primacy in that arena should privilege its legal determinations over those of the political organs. In exercising its jurisdiction, the ICJ cannot compromise its judicial character.[47] The Court cannot disregard the proprieties of judicial behavior, which under Article 39(2) of its Statute require jurisdiction over matters of international law.[48] Review of any resolutions interpreting international law is, therefore, incidental to this Charter-derived responsibility.

As long as the political organs do not extend their authority into quasi-judicial and quasi-legislative matters, the model of cooperation functions correctly. As Judge Ni in the Lockerbie case stated, "The Council has functions of a political nature assigned to it, whereas the Court exercises purely judicial functions. Both organs can therefore perform their separate but complementary functions with respect to the same events."[49]

However, as the political organs have flexed their muscle in the post-Cold War world, their functions have begun to overlap with those of the Court. The Court's judicial criteria provide a necessary complement to the political methods employed by the other organs. Without ICJ review, the political organs will continue to create legal norms through non-legal processes.

The bipolar world of the Cold War stymied the Charter's vision of an executive Security Council policing a new world order based on enforceable international norms.[50] The Council in that period, paralyzed by the static ideologies of a bilateral power structure, could not act on matters of importance because of the reciprocal veto.[51] Although there are no formal checks and balances in the Charter, the Cold War veto served as an effective proxy.[52] The demise of the Cold War witnessed a reinvigorated UN, as the Council reasserted its authority in areas of collective security.[53] While the UN may now fulfill its intended role as sovereign over a new world order, there is no check on its power analogous to that of the reciprocal Cold War veto.

As the Charter contains no formal provisions for judicial review of the political organs' action,[54] the Council and the Assembly must engage in auto-interpretation of the validity of their actions.[55] The political organs must act within their competence as set forth in the Charter or risk the loss of institutional legitimacy and the cooperation of Member States.[56] The onus is on the political organs to police their own behavior, which is notoriously difficult for a political body that makes determinations according to the national interest of its members.[57]

The political organs' relation to international law in carrying out their Charter-related duties is fundamentally different than that of the ICJ. The current President of the ICJ has characterized the political organs' decision-making process as decision not "according to the law," but decision "within the law."[58] While Article 38 of the ICJ's Statute requires that it rely exclusively on international law when settling disputes, the political organs resolve disputes "primarily according to political criteria."[59] The political basis for their decision-making is appropriate when resolving political disputes but problematic when the organs act in a judicial or legislative capacity.[60] There is a "dissonance between juridical decisions and political decisions in the international system" that renders the two forms of decision-making incompatible.[61] Unlike the political process, the law is "primarily based on considerations of fairness and normative application of rules."[62] Because the Court uses legal concepts and legal methods of proof, its "tests of validity and the bases of its decisions are naturally not the same as they would be before a political or executive organ of the U.N."[63]

In contrast, the function of a political organ is to examine issues in their political aspect, and it therefore "follows that the Members of such an organ... are legally entitled to base their arguments and their vote upon political considerations."[64] Despite these differences in purpose and process, when a political organ of the UN applies an international norm to a specific factual situation, it is a "law-creative" act.[65] Even if the members insist their resolution applies only to that one specific situation, the resolution has entered into the "stream of law" and considerations of equal treatment will favor the rule's application in equivalent situations.[66]

International acts of aggression are susceptible to political resolution by the political organs of the UN Situations such as Iraq's invasion of Kuwait in 1990 that lead to the First Gulf War are uniquely unsuited to legal resolution. Indeed, one of the purposes of the UN Charter is to facilitate prompt and effective action to counter threats to international security.[67] However, not only does the Assembly and Council counter security threats, but they also create legal obligations and interpret international law in passing their resolutions. As such, these resolutions deal with questions of law that should be "resolved through the normative application of rules."[68] The normative application of rules requires procedural safeguards to apply the law correctly and to ensure the due process rights of affected parties. However, the political organs do not have procedural mechanisms to ensure due process of parties and the normative application of the law. By contrast, the ICJ has twenty-six articles in its Statute concerning the procedural safeguards necessary to ensure the due process rights of the parties before it.

The political organs thus disregard the due process rights of affected parties when they pass resolutions that create legal obligations and foreclose possible legal defenses. When the Council must adjudicate issues of law and fact without these procedural safeguards, it is more likely to decide the issues in an arbitrary and capricious manner. Consequently, when confronted with the political organs' creation of legal obligations and clarification of international treaties, the ICJ should exercise its Charter-derived jurisdiction over the interpretation of treaties and not defer to the resolutions' factual and legal determinations. Such reconsideration of resolutions would bring the ICJ into more organizational conflict with the requesting organs, but it is necessary to ensure the due process rights of affected States and the correct application of international law.

Uncertain Legal Status of Full-Scale Judicial Review

The issue of the ICJ's review of the legality of the political organs' resolutions was raised at the founding conference of the UN.[69] The Belgian delegate proposed an amendment by which a Member State could ask the ICJ to rule on the legality of a Council or Assembly resolution.[70] The Conference rejected the amendment on the ground that the political organs were already required to act in accordance with the Charter, and judicial review would weaken or delay action by the political organs.[71] However, the significance of this rejection is unclear. Because the proposal was made in reference to the Council's non-binding, recommendatory power to settle disputes under Chapter VI of the Charter, the withdrawal of the proposal could mean that Belgium realized judicial review was unnecessary.[72] In any case, the Belgian proposal has not ended the debate about the permissible scope of judicial review.

Judicial review may be used in different contexts, and it is important to note that the review envisioned in the Belgian proposal involved a "constitutional" review of the legality of the resolution. This type of judicial review would examine whether the organ acted *ultra vires*, or beyond its power, and thus exceeded its competency as provided for in the Charter. While the ICJ has never directly invalidated a resolution of the political organs, it has ruled on the authority of the organs to pass certain resolutions.[73] The uncertainty surrounding judicial review is well-illustrated by the Namibia case, where the Court considered two direct challenges to Assembly and Council resolutions.[74] After declaring that it "does not possess powers of judicial review," the ICJ proceeded to examine the competency of the political organs to pass the resolutions, and held that they "were adopted in conformity with the purposes and principles of the Charter."[75]

Full-scale judicial review should be distinguished from the review this article proposes of the political organs' resolutions in the exercise of the ICJ's advisory jurisdiction. Reconsideration of a resolution's interpretation of international law differs from an examination of the political organ's Charter-derived authority to pass the resolution in the first place. The former is an assertion of the Court's jurisdiction over the interpretation of international law, while the latter is an inquiry into the Charter-based authority of the political organ to take action.

Nevertheless, both commentators and judges have promoted the concept of judicial review, in whatever form, within the context of the ICJ's advisory jurisdiction. Due to the nature of the judicial inquiry in the advisory jurisdiction, some level of judicial review is simple pragmatic necessity. The ICJ has stated that once the political organ has requested an opinion, any limitations on the "logical processes to be followed in answering it" would be "unacceptable because it would prevent the Court from performing its task in a logically correct way."[76] Also, some form of judicial review is necessary to satisfy the exigencies of the judicial function. Judge Onyeama has stated, "I do not conceive it as compatible with the judicial function that the Court will proceed to state the consequences of acts whose validity is assumed, without itself testing the lawfulness of the origin of those acts."[77] An inquiry into the lawfulness of the acts—their "legality, validity, and effect"—follows from the ICJ's function as the "principal judicial organ of the United Nations."[78] As a result, the review of a resolution's legality in the ICJ's advisory jurisdiction is incidental to its judicial function of deciding the dispute in accordance with international law.

Legal Effects of Resolutions in the Wall Opinion

In reviewing the resolutions that form the basis of the Wall Opinion, the Court is engaging in a legal interpretation of the political organs' application of international law. This review differs from a political interpretation of the normative content of the resolution. The ICJ has no authority to reconsider the political content of the resolutions, because to do so would undermine its judicial character. If the Assembly or Council acts within its Charter-based authority in the area of collective security, and that action is based on political imperatives rather than legal interpretation, the ICJ should not review or invalidate the decision. The ICJ could not conduct such review without infringing on the political organs' responsibilities to maintain world peace. However, the ICJ can and should review the underlying legal determinations of the political organs' resolutions, so as to fulfill its Charter-derived duty to decide disputes in accordance with international law.

Part of the Wall Opinion involved an interpretation of whether Israel had violated the Geneva Conventions in transferring parts of its population into Palestinian territory. Prior to the referral of the dispute over the wall to the ICJ, the Council had passed Resolution 446 declaring the Israeli settlements in the West Bank in violation of the Geneva Convention.[79] While there were undoubtedly overriding political considerations at play in Resolution 446, close examination of the language of Resolution 446 reveals a determination of illegality under the Geneva Conventions. Paragraph three of the resolution explicitly invokes Israel's responsibilities under the Geneva Conventions.[80] It proceeds to label Israel the "occupying power," thus mirroring the six provisions of Article 49 of the Geneva Convention, which all begin, "The Occupying Power...." The last sentence of paragraph three also shows that the court is relying on the Convention to establish the illegality of the settlements, because it imputes a violation to Israel that is a virtual transcription of Article 49 of the Convention.[81] The Resolution, therefore, relies on a legal interpretation of an international treaty. Accordingly, the Charter mandates that the ICJ review this legal interpretation, and any reconsideration of the issue underlying Resolution 446 is incidental to its duty to decide the case according to international humanitarian law.

The ICJ's reconsideration of the factual and legal determinations underlying the Council and Assembly resolutions is not meant to undermine the institutional efficiency in allowing those political organs to react quickly to political crises. The speed necessary to react to certain situations of state aggression or humanitarian crises increases the need for the ICJ to act as a *post facto* corrective on any collateral legal determinations hastily made in the Council or Assembly's resolutions. This post facto reconsideration of the legal determinations contained in the political organs' resolutions, however, should not invalidate legal obligations placed on Member States bound by the Charter to follow Council directives.

CONCLUSION

A more active role for the ICJ in its advisory jurisdiction capacity will create short-term conflict with the political organs when it reconsiders the factual and legal determinations of Assembly and Council resolutions. However, the antagonism occasioned by this form of judicial review should evolve into a mutually beneficial relationship with the political organs. An advisory opinion that simply repackages the legal conclusions of the political organs will not serve as a deterrent to States who are in violation of their legal obligations. For an advisory opinion to have force independent of the political enforcement mechanisms of the Assembly or the Council, it must persuade by way of a convincing application of international law to a well-developed set of facts.

True participation in the "purposes and principles" of the UN Charter requires that the Court perform its Charter-imposed duties to rule on claims "well-founded in fact and law."[82] The Court contributes to the shared political goals of the UN not by ceding judicial responsibility to a political organ, but by fulfilling its Charter-derived role as the UN's principal judicial organ. The ICJ's reassertion of its role as principal judicial organ is necessary to ensure the rights of litigants and the normative application of law to factual disputes.

The developing areas of international human rights and international humanitarian law are detailed, fact-laden areas of law. If the Court is to develop the law in this growing area, it must shed its reluctance to properly develop the factual record. Refusal to find facts and review the legal conclusions of the political organs in this field of law will contribute to the ICJ's marginalization as an international tribunal. Specialized courts like the Human Rights Committee will take the lead in developing international human rights law because of their command of facts on both sides of the issue—humanitarian as well as security concerns.

This article has endeavored to prove that there is room to second-guess the political organs and clarify legal questions in a way that does not compromise the Court's judicial character or infringe on the duties of the political organs to maintain international peace and security. Although the Court's desire to remain institutionally relevant contributes to its overly deferent stance to the political organs, the long-term effect of this behavior is further isolation. To remain relevant, the ICJ must do the opposite: aggressively review the factual and legal determinations of the political organs.

Endnotes

1 UN Charter art. 92.

2 *Id.*, ¶ 1.

3 Statute of the International Court of Justice art. 38, June 26, 1945, 59 Stat. 1055, 33 UNT.S. 993 [hereinafter ICJ Statute].

4 Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136 (July 9).

5 *Wall Opinion*, 2004 I.C.J. at 201-02.

6 G.A. Res. 10/13, UN Doc. GA/RES/ES-10/13 (Oct. 27, 2004).

7 The ICJ Statute is "based upon the Statute of the Permanent Court of International Justice." UN Charter art. 92. The PCIJ was dissolved after it famously ignored Hitler's invasion into Poland and continued to decide cases. Terry D. Gill, The World Court and How it Works 8 (2003).

8 UN Charter arts. 7, 92. *See* Statute of the Permanent Court of International Justice arts. 1-6, Dec. 16, 1920, 6 L.N.T.S. 380 (1926).

9 *See* ICJ Statute, arts. 34, 35.

10 *Id.*, art. 36, ¶ 2.

11 *Id.*, art. 38, ¶ 1.

12 *Id.*, art. 65, ¶ 1. *See* Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136, 153-54 (July 9)(noting that complex factual issues and political elements in advisory requests do not preclude exercise of advisory jurisdiction).

13 ICJ Statute, art. 65, ¶ 2.

14 Richard H. Fallon, Jr. et al., Hart & Wechsler's The Federal Courts and the Federal System 98 (5th ed. 2003) (noting that eleven states permit their highest court to render advisory opinions: Massachusetts, New Hampshire, Maine, Rhode Island, Florida, Georgia, Colorado, South Dakota, Alabama, Delaware, and North Carolina).

15 *See, e.g., In re* Opinions of the Justices to the Senate, 802 N.E.2d 565 (Mass. 2004) (Massachusetts senate requested opinion on constitutionality of bill which prohibits same-sex couples from entering into marriage).

16 *See* Legal Consequences for States of the Continued Presence of South Africa in Namibia (Namibia), Advisory Opinion, 1971 I.C.J. 16, 23 (June 21) (noting the political pressure surrounding South Africa's withdrawal from Namibia).

17 Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136, 158-59 (July 9) (conceding that the dispute involves a bilateral matter between Israel and Palestine).

18 *See* ICJ Statute, arts. 26, 68 (requiring that States consent in contentious matters and that ICJ consider principles of propriety for contentious matters when giving advisory opinion); *see also* Western Sahara, Advisory Opinion, 1975 I.C.J 12, 20 (Oct. 16) (noting that lack of State consent does not deprive ICJ of jurisdiction, but is consideration in determining propriety of rendering advisory opinion).

19 Mohamed Sameh M. Amr, The Role of the International Court of Justice as the Principal Judicial Organ of the United Nations 49 (2003).

20 UN Charter art. 1.

21 *Id.*

22 Muhammed Bedjaoui, The New World Order and the Security Council: Testing the Legality of its Acts 22 (1994).

23 Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136, 186 (July 9).

24 *Id.* at 187.

25 *Id.* at 159.

26 *Id.*

27 *See* Mahasen M. Aljaghoub, The Advisory Function of the International Court of Justice 1946-2005, at 76 (2006) (noting the ICJ's readiness to contribute to the UN's larger goals and render an opinion that will be of use to the requesting organ).

28 Conditions of Admission of a State to Membership in the United Nations, Advisory Opinion, 1948 I.C.J. 73, 78 (May 28) (Azevedo, J., concurring).

29 The advisory function's capacity to circumvent the principle of State consent has been a central concern for the ICJ. *See* Western Sahara, Advisory Opinion, 1975 I.C.J 12, 23-25 (Oct. 16) (discussing Spain's lack of consent to an adjudication with Morocco in an advisory jurisdiction context).

30 *See* Legal Consequences for States of the Continued Presence of South Africa in Namibia (Namibia), Advisory Opinion, 1971 I.C.J. 16, 45-58 (June 21) (noting Security Council Resolution 276 (1970), which determined that South Africa's continued presence in Namibia was illegal); *see also* Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136, 148-58 (July 9) (discussing previous Assembly and Council resolutions).

31 *See* David L. Rogers & David A. Whetten, Interorganizational Coordination 3 (1982); *see also* Arthur G. Bedeian, Organizations: Theory and Analysis 143-44 (1984).

32 Dennis Dijkzeul, The Management of Multilateral Organizations 65 (1997).

33 Aljaghoub, *supra* note 27, at 75.

34 *See* Bedeian, *supra* note 31, at 110 (discussing formulation of goals such as authority and prestige).

35 *See* Dijkzeul, *supra* note 32, at 65 (noting that "mutual adjustment implies the loss of autonomy" and this loss must be weighed against benefits of coordination).

36 *Id.* at 63.

37 *See* Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136, 161 (July 9) ("[T]he Court has at its disposal the Secretary General's report, as well as a voluminous dossier submitted by him to the Court.") .

38 ICJ Statute, art. 65.

39 *See* International Status of South West Africa, Advisory Opinion, 1950 I.C.J. 128, 143 (July 11).

40 *See* Legal Consequences for States of the Continued Presence of South Africa in Namibia (Namibia), Advisory Opinion, 1971 I.C.J. 16, 180 (June 21).

41 *Id.* at 3. *See* GA Res. 2372, UN GAOR, 22d Sess., Supp. No. 16A, UN Doc. A/6176/Add.l (June 12, 1968) (renaming of South West Africa "Namibia").

42 UN SCOR, 26th Sess., 1529th mtg. at 38-39, UN Doc. S/PV.1550, (29 July 1970) (Statement of Nepal).

43 S.C. Res. 276, ¶ 2, UN Doc. S/RES/276 (Jan. 30, 1970).

44 *See Namibia*, 1971 I.C.J. at 65 ("The Court having arrived at the conclusion that the Mandate is terminated and that the presence of South Africa in South West Africa is illegal....").

45 UN Charter art. 1.

46 *Id.*, art. 92.

47 Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (Wall Opinion), Advisory Opinion, 2004 I.C.J. 136, 161 (July 9) (stating that in exercising advisory jurisdiction, it is "necessary for [the ICJ] to give an opinion in conditions compatible with its judicial character").

48 ICJ Statute, art. 36(b).

49 Questions of Interpretation and Application of the 1971 Montreal Convention arising from the Aerial Incident at Lockerbie (Libyan Arab Jamahiriya v. United Kingdom), 1992 I.C.J. 22, 134 (Ni, J., declaration).

50 Roger D. Scott, *Global Peace Norms and Creeping Interventionism*, 154 Mil. L. Rev. 27, 29 n.7 (1997) ("It is a common view that the UN Security Council was rendered ineffective during the Cold War by discord among the permanent members.").

51 Aiyaz Husain, *The United States and the Failure of UN Collective Security: Palestine, Kashmir, and Indonesia, 1947-1948*, 101 Am. J. Int'l L. 581, 582 (2007) (noting the perpetual, reciprocal veto threats of either the United States or the Soviet Union in the Security Council).

52 W. Michael Reisman, *The Constitutional Crisis in the UN*, 87 Am. J. Int'l L. 83, 84 (1993) (explaining that the reciprocal veto in the Cold War "created a system that was [the] functional equivalent" of checks and balances).

53 *See* Ian Johnston, *US-UN Relations after Iraq: The End of the World (Order) as We Know it?*, 15 Euro. J. Int'l L. 813, 813 (2004) (remarking on a reinvigorated UN in the post-Cold War era).

54 Geoffrey R. Watson, *Constitutionalism, Judicial Review, and the World Court*, 34 Harv. Int'l L.J. 1, 4 (1993) (explaining that no provision in the UN Charter directly addresses judicial review).

55 Keith Harper, *Does the UN Security Council Have the Competence to Act as Court and Legislature?*, 27 N.Y.U. J. Int'l & Pol. 103, 105 (1994)(noting that UN organs determine their own competency).

56 Tania Vons, *Closing the Gap Between the Legitimacy and Legality of Humanitarian Intervention: Lessons from East Timor and Kosovo*, 7 UCLA J. Int'l L. & Foreign Off. 31, 38 (2002) (explaining the gap between UN Charter-derived authority and Council action may undermine the Council's legitimacy and that of the UN body as a whole).

57 Therese O'Donnell, *Naming and Shaming: The Sorry Tale of Security Council Resolution 1530*, 17 Euro. J. Int'l L. 945, 960-61 (2004) (discussing the self-interest of Member States and the decreased legitimacy of Council action taken in that interest).

58 Rosalyn Higgins, *The Place of International Law in the Settlement of Disputes by the Security Council*, 64 Am. J. Int'l L. 1, 16 (1970).

59 Jost Delbrück, *Functions and Powers: Article 24, in* The Charter of the United Nations: A Commentary 442, 447 (Bruno Simma et al. eds., 2002).

60 This is not to suggest that the disputes the political organs consider cannot or should not have legal aspects, but rather that a political organ engaged primarily in a judicial enterprise will resort to extra-legal criteria in its decision-making. *See* Higgins, *supra* note 58, at 15-16 (discussing the "substantial difference between the use, the employment of international law by the Security Council, on the one hand, and by a purely judicial body" like the ICJ).

61 Harper, *supra* note 55, at 133-34.

62 *Id.* at 137.

63 Questions of Interpretation and Application of the 1971 Montreal Convention arising from the Aerial Incident at Lockerbie (Libyan Arab Jamahiriya v. United Kingdom), 1992 I.C.J. 22, 166 (Weeramantry, J., dissenting).

64 Conditions of Admission of a State to Membership in the United Nations, Advisory Opinion, 1948 I.C.J. 57, 61 (May 28). However, this is not to say there are no legal restrictions at all on Assembly and Council resolutions; rather, Member States may use non-legal criteria in reaching their decision.

65 Oscar Schachter, *The Quasi-Judicial Role of the Security Council and the General Assembly*, 58 Am. J. Int'l L. 960, 963 (1964) ("When an organ applies a Charter principle or any other rule of law to a particular set of facts, it is asserting, as a matter of logic, a new rule of a more specific character. This is a law-creative act.").

66 *Id.* at 964.

67 UN Charter. art. 24.

68 Harper, *supra* note 55, at 139.

69 For a study of the UN Charter and the San Francisco Conference, see Djura Nuicic, The Problem of Sovereignty in the Charter and in the Practice of the United Nations (1970).

70 UN Doc. 664, IV/2/33, 13 UNC.I.O. Docs. 47, 48 (1945).

71 *Id.* at 65.

72 *Id.* at 304.

73 Amr, *supra* note 19, at 304-309 (noting cases of explicit and implicit judicial review.).

74 Legal Consequences for States of the Continued Presence of South Africa in Namibia, Advisory Opinion, 1971 I.C.J. 16, 34 (June 21) (noting challenge to Council and Assembly resolutions).

75 *Id.* at 46.

76 Certain Expenses of the United Nations, Advisory Opinion, 1962 I.C.J. 197, 217 (July 20) (Morelli, J., concurring).

77 *Namibia*, 1971 I.C.J. at 344 (Onyeama, J., concurring).

78 East Timor (Portugal v. Australia), 1995 I.C.J. 90, 424 (Skubiszewski, J., dissenting).

79 S.C. Res. 446, at 2, UN Doc. S/RES/446 (Mar. 22, 1979).

80 *Id.* at 3.

81 *Id. See* Geneva Convention Relative to the Treatment of Prisoners of War art. 49(6), Aug. 12, 1949, 6 U.S.T. 3316, 75 UNT.S. 135 ("The Occupying Power shall not deport or transfer parts of its own civilian population into the territory it occupies.").

82 UN Charter pmbl.

Labor and Employment Law

The Supreme Court's ADEA Cases: How will the Nine Grapple with a Graying Workforce in the Wake of *Ledbetter*?

*By Alan F. Smith**

The Supreme Court heard five age discrimination cases this term which will have a significant impact on employment law. Of special interest in these cases is whether the Roberts Court will apply its allegedly pro-business/pro-employer jurisprudence in the ADEA context.[1] Not all of the High Court's recent decisions have pleased Congress, particularly the House, which has shown a keen interest in remedying some of the Court's allegedly anti-worker decisions. Depending on the outcome of this term's cases, Congress may be weighing in shortly with its opinion of the Court's decisions.

Congress did just that last summer in response to *Ledbetter v. Goodyear*.[2] *Ledbetter* drew sharp protests from powerful Democrats.[3] The House passed a legislative fix—by a 225-199 party-line vote—just two months after *Ledbetter* was handed down.[4] The House bill did not make it in the Senate, but Senator Edward Kennedy has introduced legislation this session to fix *Ledbetter*.[5] There is ample reason to think that the Democrat-controlled Congress will continue to seek legislative remedies for decisions that it believes limit employee rights or remedies.

Given the graying of the American workforce, no matter what the Court decides, it will likely face more ADEA cases in the future.[6] The number of older employees who are continuing to work is predicted to increase in the near future.[7] As the nation's overall population ages, these older employees will constitute an ever greater percentage of the American workforce.[8] The increased prevalence and importance of age-related cases will undoubtedly lead to increased review by the appellate courts and, presumably, by the Supreme Court itself.[9]

Such a development may, in turn, lead to more back-and-forth between the two branches over this increasingly important area of the law. The 110th Congress has taken a renewed interest in employment issues and a continued resurgence of employment legislation is likely, particularly if the Democratic candidate wins the White House and Democrats pad their existing congressional majorities next November.[10] Even with anticipated retirements, the Court's liberal/conservative makeup is expected to remain largely the same in the near term, although it may become slightly more conservative if a Republican wins the presidency this fall.[11] This interesting confluence may, therefore, continue for some time. Whether the Court's decisions this term will engender *Ledbetter*-like condemnation by Congress and accompanying legislation remains to be seen.

Sprint/United Management Co. v. Mendelsohn: Admissibility of "Me Too" Evidence of Discrimination

Background

Ellen Mendelsohn was laid-off by Sprint as part of a reduction in force ("RIF").[12] She sued under the ADEA and sought to introduce the testimony of five other former Sprint employees to support her claim. The other employees alleged variously: (1) hearing at least one Sprint supervisor denigrate older employees, (2) that Sprint used its intern program to discriminate against older workers, (3) seeing a spreadsheet indicating that age was a consideration in layoffs, (4) receiving an unwarranted negative evaluation, (5) being "banned" from working for Sprint because of age, and (6) that Sprint required permission before hiring anyone older than forty.

District Court

Sprint sought to preclude the other employees' testimony as irrelevant. None of the other employees worked in Mendelsohn's department, reported to the same supervisors, or alleged hearing discriminatory remarks by Mendelsohn's supervisors. The district court granted Sprint's motion, allowing "evidence of discrimination against Sprint employees who are similarly situated," but excluding evidence of "discrimination against employees not similarly situated to [Mendelsohn]." The Tenth Circuit held that the district court abused its discretion by improperly applying a per se rule that evidence of discrimination from employees with different supervisors is per se irrelevant.

Supreme Court

A unanimous Supreme Court held that evidence of discrimination by other supervisors is neither per se admissible nor per se inadmissible under the Federal Rules of Evidence.[13] Whether this "me too" evidence is relevant in an ADEA case is a "fact-intensive, context specific" determination based on numerous factors.[14] The Court specifically noted factoring in "how closely related the evidence is to the plaintiff's circumstances and theory of the case."[15]

The Court further held that the Tenth Circuit did not afford the district court the deference it deserved under abuse of discretion review in determining the admissibility of evidence under the Federal Rules.[16] Nothing in the district court's decision indicated application of a per se rule, the Court said,[17] and because the basis of the district court's decision was thus unclear, the Court vacated its ruling and remanded the case.

The Aftermath

Both sides, along with business interests and pro-employee groups in general, are already claiming victory.[18] Sprint, other businesses, and business organizations are pleased because the Court overturned the Tenth Circuit's order and analysis, allowing the "me too" testimony. Mendelsohn and other employee rights groups see the Court's ruling as a watershed, because the Court has now affirmatively stated that this evidence *may* be admissible on a case-by-case basis.

**Alan Smith is an Associate with Greehan, Taves, Pandak & Stoner, PLLC in Woodbridge (VA), representing local governments and public entities.*

Federal Express Corp. v. Holowecki: What Constitutes an EEOC "Charge" in ADEA Cases?

Background

In the mid-1990s, FedEx initiated two new programs which tied employee compensation to performance measures. Employees allege the programs are merely a "veiled attempt" to force older employees to leave before they can receive retirement benefits. Patricia Kennedy, a FedEx employee, filed with the EEOC, alleging age discrimination. Kennedy and fourteen other current and former FedEx employees, also over forty years old, then filed suit, alleging age discrimination.

The issue is whether Kennedy filed a "charge." After filing an EEOC charge alleging age discrimination, an individual must wait sixty days before filing suit.[19] During this time, the EEOC processes the charge, investigates, and attempts to resolve the matter without litigation. Kennedy did not file a formal charge form. Instead, she filed an intake questionnaire with a signed affidavit attached. The district court held that these documents were not a charge and dismissed the case, only to be reversed by the Second Circuit.

The Supreme Court

The Supreme Court, by a 7-2 vote, held that Kennedy's documents were a charge. Because the ADEA did not define "charge," the Court, led by Justice Kennedy, deferred to the EEOC's reasonable interpretation and regulations. The Court adopted the EEOC's permissive request-to-act standard to determine whether a filing is a charge: in addition to the name of the party charged with discriminating, the filing "must be reasonably [and objectively] construed as a request for the agency to take remedial action to protect the employee's rights or otherwise settle a dispute between the employer and the employee."[20]

The EEOC's own interpretation and application of this position was inconsistent.[21] Initially, for example, Kennedy's filing was not considered a charge by the EEOC field office that received it.[22] Because of this mistake, Kennedy sued before the EEOC began its conciliation process with FedEx.

But the EEOC's interpretation is still entitled to deference. Its position is supported by its purpose and operations. The EEOC is required to initiate informal dispute resolution proceedings when an ADEA-related charge is received. But the EEOC is also supposed to educate the public.[23] The EEOC must have flexibility to weed out information requests from enforcement cases. Otherwise, employees may not consult the EEOC for information, fearing that this consultation will initiate action against their employer. The ADEA was intended to be easily accessible to average citizens and not to require the use of attorneys.[24]

Kennedy's filing was a charge. Her intake questionnaire, filed instead of the formal charge form, combined with her "detailed," signed affidavit were sufficient.[25] Kennedy's request to "[p]lease force Federal Express to end their age discrimination plan" is a request for EEOC to remedy the discrimination on her behalf. To a reasonable extent, filings should be interpreted to protect employee rights and remedies, and ambiguity should not be held against the filer. Kennedy's filing also included the minimum information required by federal regulations.[26] The potential confidentiality of her affidavit did not prevent it from becoming a charge.[27]

Because the EEOC did not originally consider Kennedy's filing a charge, it did not initiate the ADEA's informal dispute resolution process. FedEx was not notified of the allegations until Kennedy sued and, therefore, it had no opportunity to resolve the matter without litigation. The Supreme Court advised the district court to stay the case on remand to give the parties an opportunity to settle.

Justice Thomas, a former EEOC Chairman, wrote a strongly worded dissent, which Justice Scalia joined, which stated that the majority erred by introducing and applying an overly broad, "vague," "malleable," and "vacuous" standard.[28] The Court simply did not provide a clear standard for distinguishing whether a filing is or is not a charge. "Charge," which is commonly defined as an "accusation" or "indictment," has an ordinary administrative law understanding, Thomas said, including "the employment discrimination context where a charge is a formal accusation that an employer has violated or will violate, employment discrimination laws."[29] The standard must be grounded in the meaning of the word "charge."

The dissent argued that Kennedy's documents did not clearly and objectively indicate her "intent to initiate the EEOC's processes" and ask it to "take *the particular form* of remedial action that results from filing a charge."[30] Kennedy's request for help "cannot be equated with [her] intent to file a charge."[31] Her choice not to use the official charge form, while not dispositive, was "strong evidence" of her intent. Unlike the charge form, aspects of her filing indicated that it should not be considered a charge; including the title, a statement that additional steps were expected, and the filer's ability to maintain confidentiality.[32] The documents, in their entirety, did not objectively indicate her intent to initiate EEOC enforcement proceedings.

The EEOC's handling of Kennedy's filings and its general complaint process bolster the dissent's position. Initially, the EEOC itself did not treat Kennedy's filing as a charge.[33] Typically, the EEOC does not consider intake questionnaires, like Kennedy's, to be charges. Instead, the EEOC uses questionnaires to investigate and develop charges. Charges are then completed and filed by the EEOC, not the employee. While an intake questionnaire could be considered a charge, Kennedy's filings were not sufficient.

The Fallout

The decision generally benefits employees and is likely to meet with the approval of Congressional Democrats.[34] The Court's "permissive" standard gives the EEOC and employees greater flexibility when filing, processing, and investigating charges.[35] Althought the EEOC's interpretation and application were imperfect, its "request to act" standard will benefit employees overall, at least as opposed to the alternatives. Employees will also be encouraged by the Court's emphasis on the totality of the filing, and its strong reminder that the process should be as "laymen-friendly" as possible. The Court's holding that post-filing conduct will not nullify an otherwise valid charge is an additional positive for employees. The EEOC was also directed to revise and clarify its forms and procedures to minimize future "misunderstandings."[36]

Gomez-Perez v. Potter: Federal Employee ADEA Retaliation Claims

Background

Gomez-Perez will decide whether the ADEA provides federal employees with a private cause of action for retaliation.[37] Myrna Gomez-Perez, a forty-five-year-old Postal Service employee, was granted her request to transfer from New York to Puerto Rico. She then requested and was granted a transfer to another office which was closer to her sick mother. Shortly after this request was granted, she requested a transfer back. This request was denied, and the available position was changed to part-time and filled that day. After Gomez filed a union grievance and EEO complaint claiming age discrmination, her supervisor and co-workers retaliated against her.

District Court

Gomez sued under the ADEA, claiming she was retaliated against because of her EEO complaint. The district court held that the suit was barred, because the United States had not waived sovereign immunity for ADEA retaliation claims.

First Circuit

The First Circuit reversed the district court on sovereign immunity, holding that sovereign immunity was waived and did not bar Gomez-Perez's ADEA claim. The court then turned to whether the ADEA's federal-sector prohibition against "any discrimination based on age" includes "retaliation."[38] Although Congress chose to explicitly provide a retaliation claim for private employees under the ADEA, it did not do so for federal employees.[39] Congress's conscious decision to include a retaliation cause of action for private employees under the ADEA, but not for federal employees, is important and dispositive.[40] Unlike Title VII, where the federal employee provisions explicitly incorporate the private sector provisions, the ADEA's federal employee provisions specifically do not incorporate the private sector provisions. Congress simply did not intend to apply the broad private employee ADEA provisions to federal employees.

The First Circuit would not imply a cause of action for retaliation, as the Supreme Court did in *Jackson v. Birmingham Board of Education*, because Title IX is significantly different from the ADEA.[41] The ADEA federal provision provides an explicit right of action that does not include retaliation. Title IX instead contains a broad prohibition against discrimination and the Supreme Court has interpreted an implied right of action from it. This expansive prohibition, and the unique ability of teachers and coaches to vindicate student rights, led the *Jackson* Court to include retaliation within the implied cause of action.[42] The ADEA does not need to protect other uniquely situated parties from retaliation in the same manner as Title IX. Finally, the ADEA federal sector provision was adopted in a different historical context than Title IX.

Supreme Court

In a 6-3 decision, the Supreme Court reversed the First Circuit, holding that the ADEA's broad prohibition against age discrimination in federal employment includes retaliation.[43] Justice Alito wrote the majority opinion and was joined by Justices Breyer, Ginsberg, Kennedy, Souter, and Stevens. The majority followed *Sullivan v. Little Hunting Park, Inc.* and *Jackson v. Birmingham Board of Education,* noting that "[a]ll three cases involve remedial provisions aimed at prohibiting discrimination."[44] The majority determined that the ADEA language in this case is "not materially different from" the Title IX language in *Jackson*, and is the "functional equivalent of" the section 1982 language in *Sullivan*. It further noted that the government did not ask the Court to overrule *Sullivan* or *Jackson*, nor did it question their reasoning. In fact, the government asked the Court to follow *Sullivan* in *Jackson* and in *CBOCS v. Humphries,* which was decided the same day as *Gomez-Perez*.[45] This inconsistency was not lost on Justice Alito, who, during oral argument, pointedly questioned the government's position that "a general ban on discrimination includes a ban on retaliation except when the government is being sued."[46]

The Court held that the First Circuit erred in distinguishing *Jackson* and *Gomez-Perez*. First, the First Circuit was wrong to distinguish *Gomez* from *Jackson* because Title IX's private right of action is implied.[47] The First Circuit improperly conflated the question of whether a statute grants a private right of action with the question of whether the statute substantively prohibits certain conduct. Second, the First Circuit erred in distinguishing *Jackson* because retaliation is more important under Title IX than under the ADEA. Both *Jackson* and *Gomez-Perez* are based on the statute's text, not policy considerations.[48] Third, the First Circuit incorrectly held that Title IX was a response to *Sullivan*, but the ADEA federal-sector provision was not.[49] The ADEA was amended in 1974, five years after *Sullivan*, and only two years after Title IX was enacted. Like *Jackson*, it is "appropriate" and "realistic" to presume that Congress expected the Court to interpret the ADEA federal provision as it did similar language in *Sullivan*.

The presence of the private sector anti-retaliation provision and absence of a similar federal provision did not persuade the majority.[50] Because private sector anti-retaliation was enacted seven years before the public sector provision was added, the negative implication here is "not the strongest." The prohibitory language in the two provisions also "differs sharply." The private sector provision specifically lists prohibited employer practices (including retaliation), while the federal sector provision instead broadly prohibits "discrimination." Again, Congress was presumably familiar with *Sullivan* and expected the Court to interpret this broad prohibition as it did in *Sullivan*. These differences yield a negative implication which is not "sufficient reason to depart from the reasoning of *Sullivan* and *Jackson*."

The majority finished by dismissing the argument that the federal civil service system provides the exclusive means of addressing retaliation. There is no direct evidence, only "unsupported speculation," that Congress intended for the civil service process to be the only remedial process for retaliation. Finally, the majority questioned why the civil service commission would assume that Congress expected it to prohibit and remedy retaliation in federal employment when Congress did not explicitly do so in the ADEA itself.

Chief Justice Roberts began by noting, even in dissent, that broad anti-discrimination provisions may also prohibit retaliation. But not "every express ban on discrimination must be read as a ban on retaliation as well," because anti-discrmination and anti-retaliation are distinct concepts and have different purposes.[51] The overall context here indicates that the ADEA does not include a retaliation claim for federal employees.

Congress "was not sloppy," it consciously and deliberately chose to include retaliation in the private sector provision, but not in the federal sector. The negative implication raised by the private sector anti-retaliation provision "may not be at its 'strongest,'" but "it is certainly strong enough."[52] Congress chose not to include retaliation because it intended for the civil service system to address it. Congress is "quite familiar with that detailed administrative system" and recognized that regulating civil service is "complex," "complicated," and requires the careful balancing of conflicting policy issues and priorities.[53] When Congress extended the ADEA to federal employees, it expected the civil service system's "comprehensive regulatory scheme" to administratively address retaliation as it had done with Title VII.[54] Although the Civil Service Reform Act was enacted after the federal sector provision, it prohibits retaliation and includes a "host of administrative remedies." Congress intended to provide an administrative, not a judicial, remedy for retaliation in federal employment.[55]

The Day After

Ironically, after the turnover on the Court and the expectation of a dramatic rightward shift, the outcome in *Gomez-Perez* is entirely consistent with *Jackson v. City of Birmingham*. Justice O'Connor wrote the majority opinion in *Jackson*, which some believed might be overturned, with the help of Justice Alito, after she left the Court.[56] Instead, Justice Alito authors the majority opinion that upholds, strengthens, and extends *Jackson*. Stare decisis and the role of precedent were an important and controversial part of Justice Alito's confirmation, and they appear to have played a significant role in this case.[57] *Gomez-Perez*, like *Holowecki*, is a pro-employee decision that should be a pleasant surprise for Congress' Democrat majority. It has already received praise for its "new tone and direction... in distinct contrast to [*Ledbetter*]."[58] Rather than restricting employee rights as many predicted, *Gomez-Perez* has expanded them.

Meacham v. Knolls Atomic Power Laboratory: Who Bears the Burden in ADEA Disparate Impact Cases?

Background

The issue in *Meacham* is whether an employee alleging disparate impact under the ADEA bears the burden of persuasion issue of whether the defendant employer had reasonable grounds other than age itself for its conduct.[59] Although this is not *Meacham*'s first trip to the Supreme Court, this one will be a bit more extensive. In 2005, the Court vacated the Second Circuit's judgment (affirming the district court's judgment in favor of Meacham and the other plaintiffs) and remanded the case for reconsideration in light of the Court's decision in *Smith v. City of Jackson*, which was decided while Meacham's certiorari petition was pending.[60]

After being laid off by Knolls during an RIF, Meacham sued alleging age discrimination. Specifically, he claimed that the RIF had a disparate impact on older employees because all but one of the employees were over forty years old. In the RIF, employees were ranked by their job performance, flexibility, and the importance of their skills on a scale of zero to ten. Managers identified the bottom-ranked employees in the RIF and conducted an adverse impact analysis on the identified employees, similar to the EEOC's four-fifths rule.[61] A board, followed by Knolls' general manager and general counsel, then reviewed the manager's selections and the impact analysis.

Second Circuit

The Second Circuit applied the *Wards Cove* burden-shifting framework for analyzing disparate impact cases under Title VII, in light of *Smith*.[62] In *Smith*, the Supreme Court held that disparate impact is cognizable under the ADEA, and that the appropriate test in ADEA cases is a "reasonableness test," not the *Wards Cove* "business necessity" test.[63] The reasonableness test requires that employers rely on reasonable, specific, non-age factors to accomplish the employer's legitimate goals. The reasonableness test emanates from the ADEA itself, which permits any "otherwise prohibited" action "where differentiation is based on reasonable factors other than age."[64] The reasonableness test recognizes that "age, unlike race or other classifications protected by Title VII, not uncommonly has relevance to an individual's capacity to engage in certain types of employment."[65]

The Second Circuit held that employees bear the burden of persuasion on the "reasonable factors other than age" defense.[66] Knolls had the burden of offering a legitimate business justification, while Meacham bore the burden of demonstrating that the proffered justification was unreasonable.[67] Knolls had to offer a justification for its "unaudited and heavy reliance subjective assessments of criticality and flexibility."[68] Knolls' personnel manager and its expert testified that these assessments were "ubiquitous" and "routinely used components of personnel decisionmaking systems in general and were appropriate to the circumstances."[69] Knolls satisfied its burden, so the burden shifted to Meacham to demonstrate the justification was unreasonable.

The Second Circuit began its analysis of Meacham's burden by explaining that it would not sit as a "super-personnel department" in judgment of the "[wide-]range of reasonable personnel systems." While Knolls's process was not perfect, it was also "not unreasonable."[70] Knolls established guidelines and standards that were intended to prevent and restrict arbitrary decision-making by individual managers. And Knolls actually did monitor the RIF process. The process will usually be reasonable , even if decisions are based on subjective standards, like flexibility and criticality, if these decision are made by managers who directly supervise the employees. Because age often correlates strongly with legitimate employment needs, the age-distribution of the laid-off employees is not, by itself, probative of the RIF's reasonableness, the court said. Meacham's and the other plaintiffs' ADEA claims failed to meet their burden and establish that Knolls' justification was unreasonable.

Supreme Court Outcome

Forecasting *Meacham*'s outcome is particularly difficult due to the Court's recent turnover and *Wards Cove*'s distance in time. Justices Stevens, Scalia, and Kennedy are the only members of the Court remaining who were also on the bench when *Wards Cove* was decided in 1989. Kennedy and Scalia were in the *Wards Cove* majority, while Justice Stevens authored an impassioned dissent.[71] In *Smith*, Stevens did acknowledge that *Wards Cove* governed ADEA disparate impact claims.[72] Justices Kennedy and Thomas, who did not recognize disparate impact in the ADEA context, agreed with Justice Stevens that if these claims are cognizable, they are governed by *Wards Cove*. They may have tipped their hand further by joining Justice O'Connor in stating that "once the employer has produced evidence that its action was based on a reasonable non-age factor, the plaintiff bears the burden of disproving this assertion."[73] Because Chief Justice Rehnquist took no part in Smith, one cannot even his decision as a rough proxy for that of Chief Justice Roberts. Finally, Justice Breyer recused himself from this case, leaving the more liberal members of the Court without a typically reliable vote in employment law cases.

Perhaps the most interesting aspect of *Meacham* is what the Court will not be deciding. This was apparent during oral argument when former Solicitor General Seth Waxman conceded that regardless of what objective standards and requirements are in place, the RIF process still calls for human judgment. Waxman stated that "ultimately you are relying on one person's judgment of another." This he said, goes to the heart of the second question presented on appeal, whether granting supervisors broad discretionary authority during the RIF constituted a RFOA. The Court did not grant certiorari on this question. Ultimately, this question may have a greater impact on ADEA cases, but it will have to wait for another day.

Kentucky Retirement Systems v. EEOC

Background

Kentucky Retirement Systems ("KRS") addresses whether using age as a factor in calculating retirement benefits is facially discriminatory in violation of the ADEA.[74]

The KRS plan provides two types of retirement benefits: (1) normal retirement benefits ("normal benefits") and (2) disability retirement benefits ("disability benefits"). Workers are eligible for normal benefits at fifty-five or sixty-five years old (depending on whether they work in a hazardous position) or after working twenty years.[75] Workers disabled after they are eligible for normal retirement still receive normal benefits. Generally, annual benefits are 2.5% of final compensation multiplied by years worked. But employees who become disabled and are not yet eligible for normal retirement (i.e., under fifty-five and less than twenty years of service), have years added to their actual service. Disabled employees receive the number of years remaining until eligible for retirement (i.e., until they reach fifty-five or work twenty years), but not more than the number of years they have already worked. An employee in a hazardous position who is injured in the line of duty and is eligible for disability benefits is also guaranteed additional monthly benefits.[76]

As a result, disability benefits may be greater than normal benefits for similarly situated employees: those with the same years of service and final compensation. Employees receiving normal benefits have their actual service multiplied by 2.5%, while those receiving disability benefits have their actual service plus the additional years multiplied by 2.5%. Workers entitled to disability retirement at a younger age will also receive greater benefits than other employees with the same length of service.

KRS conceded that, "assuming every factor, other than age... is identical," the annual benefit for a younger retiring worker "will frequently exceed (and will never be less than)" the annual benefit for a worker retiring at an older age because of disability.[77] KRS also acknowledged that an employee retiring due to disability (who is not eligible for normal retirement) will always receive greater annual benefits than an older employee with the same condition, service, and final compensation, disabled after fifty-five and therefore ineligible for disability benefits.

Charles Lickteig, a sixty-one-year-old deputy sheriff (with only seventeen-and-a-half years of service) applied for disability benefits.[78] Denied because he was older than fifty-five, Lickteig filed an EEOC charge. The EEOC investigated and determined that KRS violated the ADEA. Conciliation was unsuccessful and the EEOC sued. The district court granted summary judgment for KRS, holding that the EEOC did not establish that the plan was "discriminatory, either facially or through disparate treatment combined with intent."[79]

Sixth Circuit: Take One

A three-judge panel of the Sixth Circuit upheld the district court's decision granting KRS summary judgment. Although the panel was troubled by the case, it felt constrained by precedent.[80] In *Lyon v. Ohio Education Association & Professional Staff Union*, the Sixth Circuit upheld a strikingly similar retirement plan that also attributed unworked service to younger workers when determining benefits.[81] "Because the retirement plan at issue is materially indistinguishable from the early retirement incentive plan in *Lyon*, the Kentucky Retirement plan cannot be held to violate the ADEA."[82] If it were not controlling, the "difficulties with the *Lyon* rationale and holding" may have resulted in a different outcome.[83] But *Lyon* "foreclosed" inferring intent from the employer's knowledge of its own plan.[84]

Sixth Circuit: Take Two

The Sixth Circuit granted the EEOC's rehearing request and sitting en banc, it overruled the three-judge panel.[85] The EEOC established a prima facie ADEA claim, the court held. The facially discriminatory policy was, by itself, sufficient; additional proof of discriminatory intent was not necessary. [86] "[A]n employer's intent to discriminate is directly evidenced by the employer's writing or adoption of a facially discriminatory employment policy."[87]

Applying Supreme Court precedent renders the KRS plan facially discriminatory "in at least two ways." First, the plan "categorically excludes" employees who are 55 or older and still working from disability benefits because of their age. Second, similarly situated employees who are eligible for disability benefits will receive different benefits "for no reason other than their age."[88] The plan is a "formal, facially discriminatory policy requiring adverse treatment" based on age.[89] Buttressing this holding were the "persuasive" decisions of four sister circuit courts, recognizing prima facie ADEA claims, in analogous cases.[90]

There is also "compelling evidence" that Congress intended to prohibit the age-discrimination in this case and in *Lyon*. Responding to the Supreme Court's decision in *Public Employees Retirement System of Ohio v. Betts*, Congress enacted the Older Workers Benefit Protection Act.[91] By amending the ADEA, Congress intended "to prohibit discrimination against older workers in all employee benefits except when age-based reduction in employee benefit plans are justified by significant cost considerations."[92] Congress passed the ADEA knowing full well that there was no wide-spread discriminatory animus towards older workers.[93] It could not have intended to combat this non-existent strain of age discrimination. Instead, Congress meant to prohibit discrimination based on "assumptions," having no factual basis, that age adversely affects employee work.[94]

CONCLUSION

The decisions in these cases will indicate whether the Roberts Court will interpret many employee rights and remedies more expansively or restrictively. Whether or not the Court has a pro-business or pro-employer slant, it is important for practitioners, judges, and the public to have a clearer idea of where the Court is going in the future in this important area of employment law. With Congress in Democratic hands for the foreseeable future, it will be interesting to see not only how this Court addresses new, and presumably pro-employee, legislation, but also how Congress responds legislatively to the decisions of a more conservative Court. The Court's first three decisions appear unlikely to garner the wrath of Congress. These decisions are arguably pro-employee and have received the stamp of approval from Linda Greenhouse, the doyenne of the Supreme Court beat. Greenhouse has surmised that the sharp response to *Ledbetter* may have influenced the Court's rulings. Supreme Court watchers and employment law practitioners alike will have to wait and see if this harmony continues with a few more Fed Exes or whether the remaining decisions will go over like a "*Ledbetter*" balloon. If it is the latter, expect more congressional fireworks.

Endnotes

1 Jeffrey Rosen, *Supreme Court, Inc.*, N.Y. Times Mag., Mar. 16, 2008 ("The Supreme Court term that ended last June was, by all measures, exceptionally good for American business.... The current term... has also been shaping up nicely for business interests."); Tony Mauro, *Supreme Court Continues Pro-Business Stance*, Legal Times, Feb. 21, 2008, *available at* http://www.law.com/jsp/law/LawArticleFriendly.jsp?id=1203508159200; David G. Savage, *Supreme Court Gives Business 2 Wins*, L.A. Times, Feb. 21, 2008, *available at* http://www.latimes.com/news/nationworld/nation/la-na-scotus21feb21,1,856984,print.story?ctrack=2&cset=true; Linda Greenhouse, *In Steps Big and Small, Supreme Court Moved Right*, N.Y. Times, July 1, 2007, at A1, *available at* http://www.nytimes.com/2007/07/01/washington/01scotus.html?_r=1&pagewanted=print&oref=slogin; Tony Mauro, *High Court Reveals a Mind for Business*, Legal Times, July 2, 2007, *available at* http://www.law.com/jsp/law/LawArticleFriendly.jsp?id=1183107990440; Robert Barnes & Carrie Johnson, *Pro-Business Decision Hews to Pattern of Roberts Court*, Wash. Post, June 22, 2007, at D1, *available at* http://www.washingtonpost.com/wp-dyn/content/article/2007/06/21/AR2007062100803_pf.html.

2 Ledbetter v. Goodyear Tire & Rubber Co., Inc., 127 S. Ct. 2162 (2007), *available at* http://www.supremecourtus.gov/opinions/06pdf/05-1074.pdf.

3 Press Release, Chairman George Miller, House Comm. on Education and Labor, May 30, 2007, *available at* http://www.house.gov/apps/list/speech/edlabor_dem/rel053007.html ("Congress Needs to Address Supreme Court Decision on Employment Discrimination."); Statement of Majority Leader Steny Hoyer, June 12, 2007, *available at* http://www.speaker.gov/blog/?p=479 ("Congress Will Remedy The Supreme Court's Wrongly Decided Wage Discrimination Case.").

4 Final Vote Results for Roll Call 768, H.R. 2831, Lilly Ledbetter Fair Pay Act of 2007, *available at* http://clerk.house.gov/evs/2007/roll768.xml. Two hundred and twenty three Democrats and two Republicans voted in favor. Six Democrats and 193 Republicans voted against.

5 Major Congressional Action on H.R. 2831, *available at* http://thomas.loc.gov/cgi-bin/bdquery/z?d110:HR02831:@@@R; S.1843, Fair Pay Restoration Act, *available at* http://thomas.loc.gov/cgi-bin/bdquery/z?d110:s.01843. *See also* Lori Montgomery, *White House Threatens to Veto Discrimination Bill*, Wash. Post, Apr. 23, 2008, at A4.

6 *See* Dept. of Labor, Employment & Training Admin., Rep. on the Aging of the American Workforce, Exec. Summary, Feb. 2008, *available at* http://www.doleta.gov/reports/FINAL_Taskforce_Report_2-11-08.pdf ("The aging and retirement of the Baby Boomers generation is a dramatic demographic transformation for our county. According to the U.S. Census Bureau, by 2030, 19.7 percent of the population, or about 71.5 million Americans, will be 65 or older, compared with just 12.4 percent in 2000.").

7 *See* Mitra Toossi, *Labor Force Projections to 2012: The Graying of the U.S. Workforce*, Monthly Labor Rev., U.S. Dep't. of Labor, Bureau of Labor Statistics, Feb. 2004, v. 127 no. 2, at 47*, available at* http://www.bls.gov/opub/mlr/2004/02/art3full.pdf ("Increases in the [labor participation rate] are expected to be greatest for the 55-to-64 and 65-to-74 age groups" between 2002 and 2012.); *id*. at 9 ("This increase in the labor force participation rate of the 55 and older workers is projected to continue in the future.").

8 *See id*. ("The labor force will continue to age, with the annual growth rate of the 55-year-and older group projected to be nearly 4 times that of the overall labor force; as the participation rates of older age groups increase, the older population's share of the workforce will rise.").

9 Sara E. Rix, a Senior Pol'y Advisor with the Pub. Pol'y Institute of AARP, has cited the ADEA as one of the reasons for older workers' increased participation. Sara E. Rix, *The Aging of the American Workforce*, 81 Chi.-Kent L. Rev. 593, 599 (2006), *available at* http://lawreview.kentlaw.edu/articles/81-2/Rix.pdf.

10 Roger Clegg, *Equal Rights Nonsense*, Wall St. J., Feb. 8, 2008, at A16.

11 Tom Goldstein, "Analysis: The Court and the 2008 Election," SCOTUS-Blog, *available at* http://www.scotusblog.com/wp/commentary-and-analysis/analysis-the-court-and-the-2008-election/#more-5451 ("Even a Justice on the left [ex. Stevens and Souter] who is planning on leaving and would prefer to have his or her successor appointed by a Democrat will likely retire relatively early in a Republican presidency. The Senate will probably remain in Democratic hands in 2009, limiting the prospect of a very conservative replacement.... [I]f a Democratic President wins in 2008, the current conservative-leaning détente on the Court is likely to be enshrined for the indefinite future.").

12 Sprint/United Mgmt. Co. v. Mendelsohn, 128 S. Ct. 1140 (2008), slip op. *available at* http://www.supremecourtus.gov/opinions/07pdf/06-1221.pdf.

13 *Id*. at 1.

14 *Id*. at 8-9.

15 *Id*. at 8-9.

16 *Id*. at 4-5.

17 *Id*. at 6-7.

18 *See* Robert Barnes, *Ruling May Aid Those Charging Age Bias*, Wash. Post, Feb. 27, 2008, at A2. Linda Greenhouse, *A Case-by-Case Ruling on Discrimination*, N.Y. Times, Feb. 27, 2008, *available at* http://www.nytimes.com/2008/02/27/us/27scotus.html?ei=5124&en=e60c8c366116048e&ex=1361854800&partner=permalink&exprod=permalink&pagewanted=print;

David G. Savage, *No Ruling from Justices in Age-Bias Case*, L.A. Times, Feb. 27, 2008, *available at* http://www.latimes.com/news/nationworld/nation/la-na-scotus27feb27,1,7541668,print.story?ctrack=4&cset=true.

19 19 U.S.C. § 626(d) ("No civil action may be commenced by an individual under [the ADEA] until 60 days after a charge alleging unlawful discrimination has been filed with the [EEOC]...."); *see also* Fed. Express Corp. v. Holowecki, 128 S. Ct. 1147 (2008); slip. op. at 3, *available at* http://www.supremecourtus.gov/opinions/07pdf/06-1322.pdf.

20 Majority op., at 11.

21 *Id*. at 10-11.

22 *Id*. at 8-9.

23 *Id*. at 9.

24 *Id*. at 11-12.

25 *Id*. at 14. Kennedy filed a formal charge form with the EEOC, but only after she filed suit.

26 *Id*. at 4, 13; 29 CFR § 1626.8(b) (deeming a charge "sufficient" if it is a writing containing the name of the charged party and generally alleging the discriminatory acts). The preceding section, 29 CFR § 1626.8(a), states that a "charge *should* contain" the charged party's contact information, factual description of the discriminatory acts, the number of employees employed by the charged party, and whether the charging party has sued in state court. (Emphasis provided.)

27 Fed. Express, majority op., at 14-15.

28 Dissenting op., at 2, 10, 13.

29 *Id*. at 3.

30 *Id*. at 5-6 (emphasis in original).

31 *Id*. at 6.

32 *Id*. at 6-7.

33 *Id*. at 11-12.

34 *A Verdict for Workers, for a Change*, N.Y. Times, March 2, 2008) *available at* http://www.nytimes.com/2008/03/02/opinion/02sun2.html?_r=1&pagewanted=print&oref=slogin ("The decision is noteworthy because it suggests that this court could be pulling back from what has often seemed like a knee-jerk inclination to rule for corporations over workers.... The justices may have been chastened by the enormous criticism of the [*Ledbetter v.*] *Goodyear* ruling from Congress and the public.").

35 *Id.* ("It is surprising and welcome to see the court apply any sort of permissive standard....").

36 Majority op., at 16-17.

37 Gomez-Perez v. Potter, 461 F.3d 54 (1st Cir. 2007). The AARP filed an amicus brief on behalf of Gomez.

38 29 U.S.C. § 633a(a) ("All personnel actions affecting employees or applicants for employment who are at least 40 years of age... shall be made free from any discrimination.").

39 *Id*. § 623(d) ("It shall be unlawful for an employer to discriminate against any of his employees...because such individual… has opposed any practice made unlawful by this section, or because such individual… has made a charge, testified, assisted, or participated in any manner in an investigation, proceeding or litigation under this Chapter.").

40 *Gomez*, 476 F.3d at 59 (disagreeing with the D.C. Circuit's decision in Forman v. Small which found a federal employee retaliation cause of action). Contra Forman v. Small, 271 F.3d 285, 297 (D.C. Cir. 2001) ("Nothing in the plain language of § 633a suggests that Congress intended the federal workplace to be less free of age discrimination than the private workplace.").

41 Jackson v. Birmingham Bd. of Educ., 544 U.S. 167 (2005).

42 *Gomez*, 476 F.3d at 58 (citing *Jackson*, 544 U.S. at 173-175).

43 Gomez-Perez v. Potter, slip op. *available at* http://www.supremecourtus.gov/opinions/07pdf/06-1321.pdf.

44 *Sullivan*, 396 U.S. 229 (1969) (holding that 42 U.S.C. 1982 included a claim for retaliation); majority op. at 5.

45 CBOCS West, Inc. v. Humphries, slip op. *available at* http://www.supremecourtus.gov/opinions/07pdf/06-1431.pdf

46 Transcript of Oral Argument, Gomez-Perez v. Potter, Feb. 19, 2008, at 27, *available at* http://www.supremecourtus.gov/oral_arguments/argument_transcripts/06-1321.pdf. Justice Alito questioning Deputy Solicitor General Gregory Garre:

> [W]ould it be unkind to say that the government's position seems to be that a general ban on discrimination includes a ban on retaliation except when the government is being sued?" In *Jackson* the government argued that discrimination on the basis of sex included retaliation.... And yet here you're arguing exactly, what seem to me exactly the opposite position.

See generally Jackson, 544 U.S. 167 (federal government supported Supreme Court interpreting Title IX's discrimination prohibition to include retaliation); Humphries v. CBOCS West, Inc., 474 F.3d 387 (7th Cir. 2007) (federal government supports reading § 1981's (42 U.S.C. § 1981, as amended by the Civil Rights Act of 1991) discrimination prohibition to include retaliation); *but see Gomez.*

47 Majority op. at 6-7.

48 *Id*. at 8.

49 *Id*. at 8-9.

50 *Id*. at 10.

51 Dissent (C.J. Roberts) at 3-4. Justices Thomas and Scalia chose not to join this portion of the Chief Justice's dissent. Justice Thomas, joined by Justice Scalia, added a very brief, one-paragraph dissent criticizing *Jackson* and reiterating that the federal-sector provision does not cover retaliation. Dissent (J. Thomas) at 1.

52 *Id*. at 6.

53 *Id*. at 9-10, 12.

54 *Id*. at 11-12.

55 *Id*. at 12-13.

56 Marty Lederman, "The Profound Effect of Justice O'Connor's Retirement," SCOTUSBlog, *at* http://www.scotusblog.com/wp/uncategorized/the-profound-effects-of-justice-oconnors-retirement/ (citing *Jackson* as one of the cases in which "Justice O'Connor's has been the decisive vote or opinion, and in which a more conservative Justice might well vote to overrule the governing precedent").

57 Robert Barnes, *Justices Show Ability to Move to the Center*, Wash. Post, May 29, 2008, at A2.

58 Linda Greenhouse, Justices Say Law Bars Retaliation Over Bias Claims, N.Y. Times, May 28, 2008, at A1. *See also* Robert Barnes & William Branigin, *Justices Uphold Retaliation Lawsuits*, Wash. Post, May 28, 2008, at A1. *But see Flawed Victory: The Supreme Court Stretches the Law to Help Victims of Workplace Retaliation*, Wash. Post, May 29, 2008, at A18.

> The Supreme Court pleased workers this week . . . [y]et both [*Humphries* and *Gomez-Perez*] are deeply flawed and should make those applauding the results more than a little nervous…. Protecting employees from retaliation makes sense, but it is not the province of judges to create such protections on the basis of their own beliefs of what it right or wrong, or even on the basis of their intuitive sense of what Congress meant to do or should have done. (Emphasis provided.)

59 Meacham v. Knolls Atomic Power Laboratory, 461 F.3d 134 (2d Cir. 2006). The EEOC, AARP, and National Employment Lawyers Association have filed amicus briefs supporting the plaintiffs. The U.S. Chamber of Commerce has weighed in on Knolls' behalf.

60 Smith v. City of Jackson, 544 U.S. 228 (2005). In *Smith*, Justices Stevens, Souter, Ginsburg, Breyer, and Scalia, supported applying disparate impact to the ADEA. A plurality, led by Justice Stevens, based its decision on the ADEA's text and purpose, relevant precedent (including *Griggs v. Duke Power* and *Hazen Paper Co. v. Biggins*), and agency deference. Justice Scalia concurred in judgment, deferring to the EEOC's consistent and reasonable views recognizing disparate impact recovery. He did not join the portion of Stevens' opinion addressing prior Court decisions interpreting Title VII's identical language. Justices O'Connor, Kennedy, and Thomas were not persuaded by these arguments, concurring in the judgment because disparate impact is not cognizable under the ADEA. Justice Rehnquist did not participate in the case.

61 Under the EEOC's four-fifths rule, a "serious discrepancy" exists if a protected group's selection rate is more than 120% of the total population rate.

62 *Meacham*, 461 F.3d at 143; Wards Cove Packing Co. v. Atonio, 490 U.S. 642 (1989).

63 *Smith*, 544 U.S. at 243.

64 ADEA § 4(f)(1), 29 U.S.C. § 623(f)(1).

65 *Smith*, 544 U.S. at 240.

66 *Meacham*, 461 F.3d at 141-42.

67 *Id*. at 144.

68 *Id*. at 143.

69 *Id*. at 144 (internal quotations omitted).

70 *Id*. at 146 (citing *Smith*, 544 U.S. at 243).

71 *Wards Cove*, 490 U.S. at 644-45, 662.

72 *Smith*, 544 U.S. at 240-41.

73 *Id*. at 267-68.

74 Thirteen states and several government associations (National League of Cities, National School Board Assoc.) weighed in on Kentucky's behalf. AARP and the National Employment Lawyers Assoc. filed the only brief supporting the EEOC.

75 The EEOC did not challenge this distinction. The district court and three-judge panel only reviewed the disability retirement's impact on those employed in hazardous jobs.

76 The employee receives at least 25% of their final pay and an additional 10% for each dependent child, up to a maximum of 40%.

77 E.E.O.C. v. Jefferson County Sheriff's Dept., No. 06-1037, 467 F.3d 571, 575 (6th Cir. 2006).

78 E.E.O.C. v. Jefferson County Sheriff's Dept., 424 F.3d 467, 470 (6th Cir. 2005). Although eligible for normal retirement, Licketig continued working to support his family.

79 *Id*. at 468, 471.

80 *Id*. at 475.

81 Lyon, 53 F.3d 135 (6th Cir.1995). *See also E.E.O.C.,* 424 F.3d at 471.

82 *E.E.O.C.*, 424 F.3d at 475.

83 *Id*. at 475. Lyon could be criticized for its "reliance on the observation that the very purpose of offering an early retirement incentive plan is to buy out expensive workers" because having a justification for discriminating does not make disparate treatment legal. *E.E.O.C.,* 424 F.3d at 473 (quoting Lyon, 53 F.3d at 139) (internal quotations omitted).

84 *Id*. at 473.

85 *E.E.O.C.*, 467 F.3d at 573, 583.

86 *Id*. at 583.

87 *Id*. at 582.

88 *Id*. at 579.

89 *Id*. at 579.

90 *Id*. at 573, 580 (citing the Second, Seventh, Eighth, and Ninth Circuit).

91 *Betts*, 492 U.S. 158 (1989) (holding that the retirement plan was facially discriminatory, but it met an exemption because there was not proof of subterfuge to avoid the ADEA).

92 *E.E.O.C.*, 467 F.3d at 580 (quoting the Older Workers Benefit Protection Act, 29 U.S.C. § 621). While the original panel considered *Betts*, the Act, and the legislative history, it could not distinguish *Lyon* because this information was available to the court when *Lyon* was decided. *E.E.O.C.*, 424 F.3d at 474-75.

93 *Id*. at 582-83.

94 *Id*. at 577-78, 582-83.

Wide Latitude for Retroactive Litigation Proposed in the Lilly Ledbetter Fair Pay Act

*By Donald R. Livingston**

Title VII was built for speed. Discrimination charges are to be filed within a mere 180 days "after the alleged unlawful employment practice occurred" (300 days if the charge is first filed with a state or local agency). The EEOC can sue just thirty days after that, provided the EEOC has been unable to settle. If the EEOC delays more than 180 days without filing suit, the EEOC shall notify the person aggrieved, who has just ninety days to sue. Once a lawsuit is filed, it is the duty of the court "to assign the case for hearing at the earliest practicable date and to cause the case to be in every way expedited." If these periods will not get the case to court quickly enough, the EEOC may seek immediate injunctions. This rapid process was once considered "the very backbone" of EEOC's effectiveness.[1]

What happened? It has been nearly forty-five years since the passage of Title VII, and the EEOC never gets a case to court in thirty days, seldom seeks immediate injunctions, and frequently issues its notices of the right to sue beyond 180 days. And now? Congress is considering a special rule to extend the deadline for filing some claims *forever*, by eliminating the deadline for many cases involving a wage payment.

Under the bill that has passed the House of Representatives, called the Lilly Ledbetter Fair Pay Act of 2007, an escalator principle would be applicable. A single pay decision that was made two, five, ten, or even twenty years earlier could be challenged *today* if the employee has not moved up to the pay level he would have had if the earlier decision been more favorable. To take a simple example, an employee awarded a four percent pay raise at the end of 1998 could challenge that pay raise in 2008 by making two assertions: (1) "but for" her gender, race, national origin, religion, age, or disability, she would have received a larger raise; and (2) the larger raise would have caused her paychecks to be larger *now*. It would make no difference whether the employee is currently paid more or less than her peers or how long ago the challenged decision occurred. The only issues would be whether a more favorable decision in 1998 would have resulted in higher current pay, and whether the earlier decision was influenced by a discriminatory intent.

Should the bill become law, it will put considerable pressure on employers to make gender- and race-conscious wage decisions as prophylactic protection against stale claims. This will result from the ease with which an employee can create a prima facie case of pay discrimination using circumstantial evidence. The plaintiff's burden, which establishes a refutable presumption of discrimination, has alternatively been referred to as "minimal,"[2] "not demanding,"[3] and "not onerous."[4] It requires only that the plaintiff prove she was subject to lower pay than a male employee in the same job classification. It is then the employer's burden to clearly set forth, with admissible evidence, the nondiscriminatory reasons for the pay disparity. Generally, the employer cannot rely upon an after-the-fact rationalization by someone who did not participate in the decisions but must offer reasons that were relied upon by actual decision makers. When the claims are stale, this is a near impossible task for an employer relying on supervisors to make thousands, or even hundreds of thousands, of pay decisions each year.

Consider, for example, the allegations in the case that gives the bill its name, *Ledbetter v. Goodyear Tire and Rubber Company.*[5] Lilly Ledbetter became an employee of Goodyear in Gadsden, Alabama, in 1979. She worked for Goodyear until 1998. She was subjected to annual merit-based pay decisions that each year gave her lower pay raises than her male counterparts. Finally, in March 1998, she filed a discrimination claim asserting that the cumulative effects of the annual pay decisions cost her in terms of current pay. This claim placed at issue every salary decision made during Ledbetter's nineteen-year career, and, as stated by the Eleventh Circuit Court of Appeals, "put the onus on Goodyear to provide a legitimate, non-discriminatory reason for every dollar of difference between her salary and her male co-workers' salaries."

It was undisputed that Ledbetter's claim was not entirely time-barred. Any decision affecting her pay that was made within Title VII's 180-day limitations period could be challenged. This challenge, observed the court of appeals, "[would be] identical in form to the raise-denial claims courts routinely consider."[6] But Ledbetter also sought to require Goodyear to defend the nineteen years' worth of decisions that resulted in the pay disparity that existed at the end of her career. Her argument, said the court of appeals, is "directly contrary to the central purposes of the time-filing requirement," which is to "encourage prompt resolution of employment disputes."[7]

The proponents of the Lilly Ledbetter Fair Pay Act of 2007 agree that if Ledbetter had been demoted or denied a promotion she would have had to file any sex discrimination claim within 180 days of the event, even though the demotion or promotion loss would have caused reduced paychecks over the course of her career. Or, had she then become disabled by a physical assault during this period, she would have had only two years to bring an intentional tort claim under Alabama law, regardless of whether the assault limited her ability to earn wages into the future. But Lilly Ledbetter was not assaulted, or demoted, or denied a promotion. Instead, she was subjected to annual merit-based pay decisions that each year gave her lower pay raises than her male counterparts.

We are told that the reason wage claims should have a special rule, not applicable to assaults, demotions, or lost promotions is the "reality of wage discrimination." The proponents of the bill say that pay disparities are significantly different from other adverse actions because they often occur

....

** Donald R. Livingston is a Partner with Akin Gump Strauss Hauer & Feld, LLP, in their Washington D.C. office. He served as General Counsel of the Equal Employment Opportunity Commission from 1990-1993.*

in small increments, develop over time, and must be shown by comparative pay information that is often hidden by the employer. But it is a distortion of how statutes of limitation are interpreted to suggest they allow the employer to game the judicial system to deprive a victim of recourse by hiding the discriminatory act. Courts have protected claimants from the harsh application of a statute of limitations where the claimant did not have notice of possible discrimination.[8] The existence of this "discovery rule" was recognized in the *Ledbetter* opinion, though the Court declined to address the issue because it was not argued by Ledbetter.[9] It seems that Ledbetter did have notice of possible discrimination. She testified that she knew at least three years before she filed her Title VII charge that her pay was lower than her peers.[10]

If the sponsors of the Lilly Ledbetter Fair Pay Act of 2007 were concerned about wage discrimination being hard to detect during the 180-day charge filing period, they could legislate to codify a discovery rule applicable to wage claims. Instead, they propose to open up decades-old employment decisions to current challenge even in situations where it is undisputed that the claimant knew years earlier that an adverse pay decision had occurred. In other words, this truthful statement in 1990, "Mary, we are giving everyone a raise this year but you," would be treated the same as this lie in 1991, "Mary, we are giving only you a raise this year." Both claims could be challenged in 2008 as having an adverse impact on 2008 pay. Under present law, the 1990 decision would need to be challenged within 180 or 300 days. The 1991 decision could be challenged within 180 or 300 days after Mary knew or should have known of discrimination.

There is a statement in the preamble of the Ledbetter bill that the intent of the bill is to return to pre-*Ledbetter* law. The statement is supported by ample lower court authority, but ignores Supreme Court precedent. In 1976 in *United Air Lines, Inc. v. Evans*, for example, the Supreme Court held that current application of a facially neutral seniority system is not a discriminatory act, even when the seniority system perpetuates the effects of past discrimination.[11] Subsequent opinions in *Delaware State College v. Ricks* (1980), *Lorance v. AT&T* (1989), *National Railroad Passengers Corp. v. Morgan* (2002), and *Ledbetter v. Goodyear Tire & Rubber Company* (2007) consistently hold that an employment decision that pre-dates the charge-filing period under Title VII cannot be challenged as a current violation of Title VII, even where the decision carries forward the effects of prior, uncharged discriminatory decisions.[12]

It is not contested that this precedent rules out Ledbetter's claims. It is asserted, however, that the Supreme Court's 1986 decision in *Bazemore v. Friday* carved out an exception for wage claims under which "[e]ach week's paycheck that delivers less to a black than to a similarly situated white is a wrong actionable under Title VII, regardless of the fact that this pattern was begun prior to the effective date of Title VII."[13] Numerous lower courts have interpreted *Bazemore* to hold that a claim based on a discriminatory pay decision could challenge the current effects of that decision, even if the decision was made years before the expiration of the limitations period. Each paycheck could be challenged separately regardless of whether the pay differential had its genesis in a discriminatory act years earlier. But the case law has not been entirely consistent on this point, as illustrated by the observation of the Seventh Circuit Court of Appeals that there is "a line of cases decided in this court… that are in tension with the rule that treats each check in a simple discriminatory pay claim as a new violation."[14]

This "tension" corresponds to questions about the holding of *Bazemore*. Did *Bazemore* refuse to insulate a pay scheme that is presently illegal on the basis that it was not illegal when adopted? Or did *Bazemore* hold that a past discriminatory pay decision is a present violation until it is corrected?

It has been described both ways, even by the EEOC. In *Cardenas v. Massey*, the EEOC described *Bazemore* as holding that plaintiffs can currently challenge an ancient discriminatory wage decision under Title VII, where the decision causes current unequal wages. The EEOC distinguished *Evans* on the ground the *Bazemore* is a wage case, subject to a special limitations rule ("the *Bazemore* Court specifically distinguished wage cases from cases like *Evans*").[15] In the EEOC's brief to the Sixth Circuit Court of Appeals in *EEOC v. Ameritech*, however, the EEOC described *Bazemore* as a challenge to a "discriminatory wage structure." Because the pay structure was facially discriminatory, it was immaterial that the pay structure had existed for years. "Each paycheck," said the EEOC, "was a new discriminatory act." The EEOC distinguished *Bazemore* from *Evans,* not on the basis that *Bazemore* creates a special carve-out for wage claims, but because *Bazemore*, unlike *Evans*, involved a facially discriminatory system that could be challenged at any time.[16] This is precisely the way the Supreme Court distinguishes *Bazemore* in its *Ledbetter* opinion. The Court states,

> *Bazemore* stands for the proposition that an employer violates Title VII and triggers a new EEOC charge filing period whenever the employer issues paychecks using a discriminatory pay structure. But a new Title VII violation does not occur and a new charging period is not triggered when an employer issues paychecks pursuant to a system that is "facially nondiscriminatory and neutrally applied."

Undoubtedly, a statute of limitations prejudices plaintiffs by cutting off the right to challenge an offense. As the *Ledbetter* facts attest, there are advantages to plaintiffs when there is no time limit to challenge an action. However, a time limit imposes discipline on the judicial process. It protects the parties from speculative claims that have to be tried with foggy memories and lost records, or, as in Goodyear's case, without a key witness who had died.

That wage discrimination may be hard to detect is not an adequate justification for removal of any limitation period. Other offenses are also hard to detect. Courts deal with this. They employ equitable principles to mitigate harsh, unfair results that would flow from a strict application of a time limit in circumstances where a claimant could not have known to file a claim. Because these equitable principles have not been shown inadequate, there is no good justification for Congress to remove Title VII's limitations period for a claim that specific pay decisions, long in the past, were motivated by discriminatory intent. Justice is not served by extending a deadline for a claimant who has neglected timely to assert her right, when by the exercise of reasonable diligence she could have asserted the claim timely.

Endnotes

1 *In re EEOC*, 709 F.2d 392 (5th Cir. 1983).

2 Young v. Warner-Jenkinson Co., 152 F.3d 1018, 1022 (8th Cir. 1998); Cordova v. State Farm, Ins. Co., 124 F.3d 1145, 1148 (9th Cir. 1997).

3 Greenway v. Buffalo Hilton Hotel, 143 F.3d 47, 52 (2d Cir. 1998).

4 Brennen v. GTE Gov't Sys. Corp., 150 F.3d 21, 26 (1st Cir. 1998).

5 127 S.Ct. 2162 (2007).

6 421 F.3d 1169, 1180 (11th Cir. 2005).

7 *Id.* at 1181.

8 Numerous decisions on this point are cited in Lindemann & Grossman, Employment Discrimination Law, 4th ed., (BNA 2007), p. 1796, n. 108.

9 Ledbetter v. Goodyear Tire & Rubber Co., 127 S.Ct. 2162, 2177 n. 10.

10 *Ledbetter,* Joint Appendix at 231-33.

11 431 U.S. 553 (1977).

12 Delaware State College v. Ricks, 449 U.S. 250 (1980), Lorance v. AT&T Technologies, Inc., 490 U.S. 900 (1989), National Railroad Passengers Corp. v. Morgan, 536 U.S. 101 (2002), and Ledbetter v. Goodyear Tire & Rubber Company, 127 S.Ct. 2162 (2007).

13 478 U.S. 385, 395 (1986).

14 Reese v. Ice Cream Specialties, Inc., 347 F.3d 1007 (7th Cir. 2003).

15 No. 00-522, Brief of EEOC as Amicus Curiae, http://www.eeoc.gov/briefs/cardenas.txt.

16 No. 04-3496, Brief of EEOC as Appellant, http://www.eeoc.gov/briefs/ameritech1.txt.

Other People's Money...

*By Rob McKenna & Geoffrey William Hymans**

Davenport v. Washington Educational Association had all the ingredients of a blockbuster: a campaign finance case with First Amendment speech and association claims, important federalism implications, and major players with significant resources (a large labor union and the State of Washington) on each side.[1] Of the thirteen judges who examined the case before it reached the United States Supreme Court, eight thought the statute unconstitutional, while five held the opposite view.[2] In the Washington State Supreme Court, six justices voted to strike down the campaign finance law, which had the approval of the electorate and protected non-union members. Their opinion rested heavily on a short phrase drawn from a U.S. Supreme Court precedent: "dissent is not to be presumed."[3] Yet when the opinion was handed down from the United States Supreme Court, it was a 9-0 decision, with three justices stating that they would not even have reached all the union's claims.[4] Placing the lower court's misapplication of the phrase in proper context, the case resulted in a decisive victory for the rights of public employees to dissent and protected the ability of states to experiment with varied worker-protection and campaign finance regulations.[5]

The Statute

As Justice Scalia noted in the unanimous opinion he authored, "the National Labor Relations Act leaves States free to regulate their labor relationships with their public employees."[6] States have adopted varied approaches to their regulation of, and relationship with, public sector labor unions. In a "closed shop" state, all employees are required to be members of a labor union.[7] In an "agency shop" state, public employees do not have to belong to a union but they must still pay a fee, known as an "agency shop fee," to the union to support its collective bargaining activities.[8] In a "right to work" state, public employees typically are not required to belong to a union or to pay agency shop fees.[9]

Washington is an "agency shop" state.[10] Under Washington statutes and collective bargaining agreements in place at the time *Davenport* arose, public sector employees were required to pay an agency shop fee to labor unions equivalent to union dues, and those fees were deducted from public employees' paychecks and deposited in the union's accounts.[11] The acknowledged purpose of requiring payment of such agency shop fees as a condition of employment is "to prevent nonmembers from free-riding on the union's efforts, sharing the employment benefits obtained by the union's collective bargaining without sharing the costs incurred."[12]

In 1992, the citizens of the Washington State approved an initiative known as the Fair Campaign Practices Act.[13] That act was approved by an overwhelming majority of Washington voters, with 73% voting in favor.[14] The Act contained the provision at issue in this case, which stated:

> A labor organization may not use agency shop fees paid by an individual who is not a member of the organization to make contributions or expenditures to influence an election or to operate a political committee, unless affirmatively authorized by the individual.[15]

This statute provided a deceptively simple limitation on union election spending. A union could not spend agency shop fees—the fees paid by non-members of the union to support collective bargaining—"to influence an election" unless the non-member "affirmatively authorized" such expenditures. The provision did not specify how such affirmative authorization was to be obtained and did not set in place any documentation or record-keeping requirements. The provision did not regulate the spending of union *member* dues (presuming that all union members agreed with the unions' election-related expenditures). The restriction was limited to expenditures to influence an election or operate a political action committee (PAC). There were no statutory restrictions on a union spending member dues, or non-member agency shop fees, on other "political" activities, such as lobbying or "public education" campaigns.

The Precedent

The U.S. Supreme Court has, in a series of cases, established limits on the ability of unions to use their members' dues for political purposes without their members' consent.[16] In *Abood v. Detroit Board of Education*,[17] the Court held that "public-sector unions are constitutionally prohibited from using the fees of objecting nonmembers for ideological purposes that are not germane to the union's collective-bargaining duties."[18] In current parlance, the amount of union dues that is "germane" to collective bargaining activities is "chargeable," and the remaining portion of union dues (or agency shop fees) is "nonchargeable." *Abood* grounded the dissenter's rights in the First Amendment, both in the freedom from compelled speech and the freedom from forced association.[19]

In *Teachers v. Hudson*,[20] the Court "set forth various procedural requirements that public-sector unions collecting agency fees must observe in order to ensure that an objecting non-member can prevent the use of his fees for impermissible purposes."[21] These constitutionally mandated "*Hudson* procedures" require a public-sector union to notify non-members of their right to object to paying fees for non-chargeable expenses and to provide the non-members with "an adequate explanation of the basis for the fee, a reasonably prompt opportunity to challenge the amount of the fee before an impartial decisionmaker, and an escrow for all amounts reasonable in dispute while such challenges are pending."[22] *Hudson* thus held that the ability to "opt-out" of paying non-chargeable expenses was constitutionally required, and legislative schemes which reflect such procedures are referred to as "opt-out" statutes.[23] A statute such as Washington's, requiring affirmative assent before a union could spend a subset on non-chargeable expenses, is an "opt-in" legislative scheme.

....

** Rob McKenna is the Attorney General for Washington State and personally argued Davenport at the United States Supreme Court. Geoffrey William Hymans is an Assistant Attorney General in Washington State and was the law clerk for Justice Richard Sanders when Davenport was before the Washington State Supreme Court.*

The Case Below

In 2001, the Washington Education Association (WEA) was sued by both the State of Washington and a group of non-members who had been required to pay agency shop fees. The WEA is the exclusive collective bargaining agent for approximately 70,000 public educational employees in Washington.

One case began when the Evergreen Freedom Foundation, a Washington nonprofit public policy group, filed a complaint with the Washington State Public Disclosure Commission (PDC). This state agency has general statutory authority to enforce state campaign finance laws. The complaint alleged violations of section 760, and the WEA entered into a stipulation with the PDC agreeing that the WEA had violated section 760 during the 1999-2000 fiscal year. The PDC referred the case to the Office of the Attorney General, and the Attorney General filed an action alleging that the WEA had violated section 760 from 1996-2000. The trial court judge granted the state's motion for partial summary judgment holding section 760 constitutional, and proceeded to a bench trial to determine whether the union had in fact violated the statute. After the trial the judge issued a letter opinion in which he held that the WEA had violated section 760 and assessed a fine of $400,000. With statutory costs and attorney's fees the total judgment against the union was $590,000.[24]

The second action also began in 2001, when five public educational employees who were not WEA members, including the named plaintiff, Gary Davenport, filed a class action against WEA seeking a refund of the portion of agency shop fees used for political expenditures.[25] The trial court stayed proceedings pending interlocutory appeal, and the cases were consolidated for review in the Washington State Court of Appeals.[26] The court of appeals split 2-1, with the majority holding that section 760's:

> [A]ffirmative authorization requirement... would unduly require a union to protect nonmembers who disagree with a union's political expenditures but are unwilling to voice their objections. The procedures imposed on unions by federal law fully protect nonmembers' First Amendment rights. Further restrictions, such as an opt-in procedure, upset the balance between nonmembers rights and the rights of the union and the majority.[27]

The case thus arrived at the state supreme court, which affirmed the state court of appeals, holding in a 6-3 decision that section 760 was unconstitutional. Before the state supreme court, the WEA initially argued that "the *Hudson* process satisfies the requirement of affirmative authorization because it provides each individual non-member the opportunity to object, to obtain a refund, and to prevent fees from being used by WEA[.]"[28] Thus, the union actually argued that a teacher's failure to respond equaled *affirmative* authorization. The state supreme court majority did not expressly reject this linguistic oxymoron, but ultimately accepted what the plain language "seemed" to state:

> Because § 760 does not define "affirmative authorization," it is unclear whether the *Hudson* process satisfies the authorization requirement. The plain language seems to indicate a nonmember must provide an expression of positive authorization. Failure to respond to the Hudson packet may be considered acquiescence, but it would not fulfill the affirmative authorization requirement. The difference is that affirmative authorization seems to indicate that the member must say "yes," instead of failing to say "no."[29]

The majority then turned to the union's constitutional challenges to the statute. Examining the U.S. Supreme Court's *Street, Abood, Ellis* and *Hudson* cases, the majority determined that these cases not only provided a constitutional "floor" setting forth requirements to protect dissenting non-member's rights, they also imposed a constitutional "ceiling" limiting the amount of protection for dissenters a state could provide. The majority plucked a phrase from *Street*: "[D]issent is not to be presumed—it must affirmatively be made known to the union by the dissenting employee,"[30] then examined *Abood* and concluded that the U.S. Supreme Court "affirmed that the burden is on the employee to make his objection known."[31]

The majority stated that "Section 760 impermissibly shifts to the union the burden of the non-members' rights. This has the practical effect of inhibiting one group's political speech (the union and supporting non-members) for the improper purpose of increasing the speech of another group (the dissenting non-members)."[32] It went on to accept the union's assertion that the administrative procedures required to procure affirmative authorization would be "extremely costly and would have a significant impact on the union's political activities."[33] Finally, the majority held that "a presumption of dissent violates the First Amendment rights of nonmembers as well" because "[i]t assumes that because an employee has not joined the union, he or she disagrees with the union's political expenditures…. For those non-members who agree with the union's political expenditures, section 760's presumption of dissent presents an unconstitutional burden on the right to associate themselves with the union on political issues."[34]

Three justices dissented. The dissent's author, Justice Richard Sanders, began with a quote from Thomas Jefferson "[t]hat to compel a man to furnish contributions of money for the propagation of opinions which he disbelieves and abhors is sinful and tyrannical." The dissent noted that:

> The majority turns the First Amendment on its head. Unions have a statutory, not constitutional, right to cause employers not only to withhold and remit membership dues but also to withhold and remit fees from nonmembers in an equivalent amount. Absent this statutory mechanism for the withholding and remission of agency fees (or membership fees for that matter), there is no right, constitutional or otherwise, for the union to require it.[35]

Justice Sanders stated that since the legislature (or the people, acting through the initiative process) could eliminate the compelled payment of an agency fee altogether, it was "nearly beyond comprehension" that the compelled collection of such fees could not be qualified to protect dissenters even above the constitutional "floor." Finally, the dissent expressed skepticism that non-union employees who elected not to associate with the union would nevertheless want to have their fees used to support the union's political activities, and the dissent concluded that simply requiring those rare individuals to "check a box" would constitute no constitutionally cognizable burden on their association or speech rights.

The Case Arrives

Davenport thus arrived at the U.S. Supreme Court with the potential to redefine accepted First Amendment interests of union members and non-members. It also raised significant federalism issues, as noted in the amicus brief of the States of Colorado, Alabama, Idaho, Ohio, Utah, and Virginia. As set forth in that brief, states have adopted a wide variety of legislation regulating the collection of public employee union dues, from "right to work" states that ban compulsory fees to requirements that unions maintain separate political accounts funded only by voluntary contributions to legislative implementation of the *Hudson* procedures.[36]

The case attracted a significant number of amicus briefs (thirteen) filed in support of the State of Washington, but only a single amicus brief filed in support of the WEA.[37] Several of those briefs are worthy of mention. The Solicitor General filed a brief on behalf of the United States which essentially tracked the arguments made in the dissent below. The *Davenport* plaintiffs and several amici filed briefs arguing that dissent on the part of non-members should be presumed, and thus the constitution should be interpreted to require an opt-in approach. Some amici argued that even union members should be presumed to dissent from the political spending of a union unless they affirmatively authorize the use of their dues for such purposes.[38]

Oral argument took place on January 10, 2007.[39] Washington State Attorney General Rob McKenna appeared on behalf of the State of Washington and United States Solicitor General Paul Clement appeared on behalf of the United States in support of the state. John M. West appeared on behalf of the WEA. The bench was active, initially focusing on the apparent intent of section 760 as a campaign regulation rather than as a worker protection measure and whether the Court was bound by the state supreme court's determination of legislative intent with respect to the statute. Solicitor General Clement discussed the federalism aspects of the case. Perhaps his most interesting colloquy came when Justice Alito echoed several amici by asking "why should the First Amendment permit anything other than an opt-in scheme?" General Clement replied that the First Amendment interest distilled from the *Abood* line of cases was "in not having a compelled extraction, and as part and parcel of the constitutional violation, it seems to have assumed there's a need for a stated objection." This response likely unnerved those hoping the Court would replace "opt-out" with "opt-in" as the constitutional floor.

Justice Alito, in questioning the WEA's counsel, expressed skepticism that non-members who want their fees used for the union's election spending even exist. Chief Justice Roberts questioned the extent of the administrative burden an opt-in requirement would actually place on the union, and Justice Scalia's questioning focused on what would become the crux of the opinion—the legislature's underlying authorization for the union to compel payment of agency shop fees as a condition of public employment.

The Finale

The Court's unanimous opinion was issued on June 14, 2007. Following a short recitation of the facts and a brief history of the union compelled-speech cases, Justice Scalia noted that "it is undeniably unusual for a government agency to give a private entity the power, in essence, to tax government employees." Justice Scalia then characterized section 760 as "simply a condition on the union's exercise of this extraordinary power[.]"[40]

The Justice then turned to the lower court's opinion. His first target was the lower court's focus on *Street's* oft-quoted phrase, "dissent is not to be presumed." Scalia wrote that the lower court's view of a constitutionally mandated "balance" (making *Hudson* procedures both a floor and a ceiling) extended the "agency-fee cases... well beyond their proper ambit. Those cases were not balancing constitutional rights in the manner respondent suggests, for the simple reason that unions have no constitutional entitlement to the fees of non-member-employees."[41] Scalia went on: "The mere fact that Washington required more than the *Hudson* minimum does not trigger First Amendment scrutiny. The constitutional floor for unions' collection and spending of agency shop fees is not also a constitutional ceiling for state-imposed restrictions."[42]

Justice Scalia then set the "dissent is not to be presumed" maxim in its proper context, stating that:

> We meant only that it would be improper for a court to enjoin the expenditure of the agency fees of all employees, including those who had not objected, when the statutory or constitutional limitations established in those cases could be satisfied by a narrower remedy.... [A]s the dissenting justices below correctly recognized, our repeated affirmation that *courts* have an obligation to interfere with a union's statutory entitlement no more than is necessary to vindicate the rights of nonmembers does not imply that legislatures (or voters) themselves cannot limit the scope of that entitlement.[43]

In a section joined by five other justices,[44] Justice Scalia dismissed the union's claims that the Court's campaign finance cases mandated strict scrutiny of section 760's election-related expenditure limitation, rejecting the union's contention that the expenditure limit was an unconstitutional content-based restriction on speech:

> "We do not believe that the voters of Washington impermissibly distorted the marketplace of ideas when they placed a reasonable, viewpoint-neutral limitation on the State's general authorization allowing public-sector unions to acquire and spend the money of government employees.... The restriction on the state-bestowed entitlement was thus limited to the state-created harm that the voters sought to remedy."[45]

Justice Scalia ended on a cautionary note, emphasizing that the Court was upholding the opt-in scheme only as to public sector unions. The distinction in the mechanism of coercion—government mandate as compared to contract—might produce a different outcome. Justice Scalia even qualified his qualification in a conspicuous footnote ("We do not suggest that the answer must be different."[46]).

Postscript

A decisive victory for dissenting worker's compelled speech and association rights, *Davenport* actually raised concerns for one commentator.[47] And the opportunity to determine whether an opt-in mechanism is constitutionally required, as Justice Alito appeared inclined to accept at oral argument, awaits another day. But the case decisively established the *Abood* line of cases as protective of dissenting workers' rights, and put to rest the notion that those cases serve as a straightjacket on states, limiting their ability to experiment with greater worker protections. *Davenport* is therefore best seen as perhaps one of the more understated, yet important, *federalism* decisions of the 2007 term. The decision fulfilled its promise more subtly than originally anticipated, but perhaps with greater long-term effect.

Endnotes

1 Davenport v. Washington Ed. Assoc., 127 S. Ct. 2372 (2007).

2 One of the two trial court judges in the consolidated cases issued no rulings on the merits of the constitutional claims. *Davenport*, 127 S. Ct. at 2378.

3 Teachers v. Hudson, 475 U.S. 292, 306 n. 16 (1986) (quoting Machinists v. Street, 367 U.S. 740, 774 (1961).

4 *Id*. at 2382 (concurring opinion of Justice Breyer, joined by Chief Justice Roberts and Justice Alito).

5 "One of federalism's chief virtues, of course, is that it promotes innovation by allowing for the possibility that 'a single courageous State may, if its citizens choose, serve as a laboratory; and try novel social and economic experiments without risk to the rest of the country.'" Gonzales v. Raich, 545 U.S. 1, 42 (2005) (quoting New State Ice Co. v. Liebmann, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting)).

6 *Davenport*, 127 S. Ct. at 2376 (citing 29 U.S.C. §152(2)).

7 *See* Pattern Makers' League of North America, AFL-CIO v. N.L.R.B., 473 U.S. 95, 105-106 (1985) (citing R. Gorman, 106 Labor Law 639 (1976)).

8 *Davenport*, 127 S. Ct. at 2376.

9 *See* Amicus Brief of States of Colorado, Alabama, Idaho, Ohio, Utah, and Virginia at 11.

10 Wash. Rev. Code § (RCW) 41.56.122(1); 41.59.100.

11 RCW 41.56.10; 41.59.060.

12 *Davenport*, 127 S.Ct. at 2377 (citing *Street,* 367 U.S. at 760-764).

13 *See* Wash. State Pub. Disclosure Comm'n v. Washington Ed. Ass'n., 156 Wn. 2d 543, 553, 130 P.3d 352 (2006) (citing Initiative 134, set forth at Laws of 1993, ch.2, §1-36). The State of Washington included in its State Constitution an initiative process by which citizens can collect a set number of signatures and place a legislative proposal before the state legislature. If the legislature fails to take action on the proposal during the next legislative session, the proposal is placed on a statewide ballot, and upon approval such initiatives become enacted state statutes. *See* Wash. Const. art. II, § 1.

14 *See* http://www.secstate.wa.gov/elections/initiatives/statistics_initleg.aspx.

15 Initiative 134, codified at RCW 42.17.760 (hereinafter §760).

16 Machinists v. Street, 367 U.S. 740 (1961) interpreted a federal statute (the Railway Labor Act) as restricting expenditure of exacted money for political purposes against the express wishes of a dissenting employee. The Supreme Court expressly stated that it so interpreted the statute in order to avoid reaching the constitutional objections also raised by the dissenting employees. *Id*. at 750.

17 431 US 209, 97 S.Ct. 1782 (1977).

18 *Davenport*, 127 S. Ct. at 2377 (citing *Abood*, 431 U.S. at 235-36).

19 *Abood*, 431 U.S. at 233-235.

20 475 U.S. 292, 106 S. Ct. 1066 (1986).

21 *Id.* In *Ellis v. Railway Clerks*, 466 U.S. 435, 104 S. Ct .1883 (1984) the Court struck down a "pure rebate approach" as inadequate under the First Amendment, but the Court did not set constitutionally-required procedural parameters. *See also Hudson*, 475 U. S. at 305.

22 *Hudson*, 475 U. S. at 310.

23 Such a procedure would be constitutionally required under *Hudson* even in the absence of any specific state statutory regulation of public sector union dues collection procedures.

24 All these facts are drawn from the Washington State Supreme Court's opinion at 156 Wn.2d 543, 551-53, 130 P.3d 352 (2006).

25 The non-union teacher plaintiffs were represented by attorneys provided by the National Right to Work Foundation.

26 The court of appeals only issued in an opinion in the consolidated PDC case, but after holding §760 unconstitutional in that case the court of appeals remanded the *Davenport* case to the trial court for dismissal. *Id.* at 553.

27 State ex. rel. Washington State Public Disclosure Com'n v. Washington Educ. Ass'n, 117 Wn. App. 625, 640, 71 P.3d 244 (2003).

28 Wash. State Pub. Disclosure Comm'n v. Wash. Ed. Ass'n., 156 Wn. 2d at 553-54.

29 *Id*. at 554-55 (emphasis added).

30 *Id*. at 559 (quoting *Street*, 367 U.S. at 774).

31 *Id.* The majority made the same claim regarding *Hudson* and *Ellis* in the next paragraph.

32 *Id.* at 560.

33 *Id.* at 561, 569-70. The majority engaged in no discussion of how the burden would be greater on the union than the existing *Hudson* packet, or how the "burden" would be greater on a non-members who actually wanted the union to use their fees for the union's election spending than returning a post-card with the "spend my money" box checked.

34 *Id.* at 561-62. Again, the majority never discussed how that "burden" was greater than perhaps returning a prepaid post-card. The dissent quoted a Michigan "opt-in" case stating "the suggestion that asking people to check a box once a year unduly interferes with the speech rights of those contributors borders on the frivolous." *Id.* at 578 (quoting Mich. State AFL-CIO v. Miller, 103 F.3d 1240, 1253 (6th Cir. 1997)).

35 *Id.* at 571-72.

36 *See* Amicus Brief of States of Colorado, Alabama, Idaho, Ohio, Utah, and Virginia at 12-13. Federalism was also the focus of the amicus brief filed by the American Legislative Exchange Council.

37 *See* Amicus Brief of the American Federation of Labor and the Congress of Industrial Organizations (AFL-CIO) and Change to Win in Support of Respondent.

38 *See, e.g.,* Amicus briefs of the Association of American Educators, Mackinac Center for Public Policy, the Mountain States Legal Foundation, and the Pacific Legal Foundation.

39 The transcript of the oral argument is available at http://www.supremecourtus.gov/oral_arguments/argument_transcripts/05-1589.pdf.

40 *Davenport*, 127 S. Ct. at 2378.

41 *Id.* at 2379.

42 *Id.*

43 *Id*.

44 Justice Breyer, joined by Chief Justice Roberts and Justice Alito, wrote a concurrence in which he stated that these arguments by the union should not have been addressed because they were raised for the first time in the WEA's briefs to the United States Supreme Court. *Id*. at 2383.

45 *Id.* at 2382.

46 *Id.* at n. 4.

47 Erik S. Jaffe, *When Easy Cases Make Bad Law: Davenport v. Washington Education Association and Washington v. Washington Education Association*, 2007 Cato Sup. Ct. Rev. 115.

Locke v. Karass: Should the Court Overrule *Lehnert*'s Test For Determining Whether Union Expenditures Are Related to Collective Bargaining?

*By Phineas E. Leahey**

In the fall 2008 term, the U.S. Supreme Court will hear argument in *Locke v. Karass*, a case of more potential significance than suggested by the narrow question presented: whether, consistent with the First Amendment, the State may compel non-member employees to fund litigation by the affiliate of a union certified as their exclusive bargaining agent. Certiorari was granted to resolve a circuit split over whether such "extra-unit" litigation expenses are "chargeable" to dissenting non-members, but *Locke* presents a possible opportunity for the Court to revisit the prevailing constitutional standard for determining when public sector unions may compel financial support for their activities from non-members.

In *Locke*, both the non-members and their exclusive bargaining agents under Maine law urge the Court to rule in their favor based on *Lehnert v. Ferris Faculty Association*,[1] where the majority opinion of a splintered court led by Justice Blackmun—joined, in relevant part, by Justices Rehnquist, White, Stevens, and Marshall—announced a three-part test under which non-members are responsible for costs that: (1) are "germane" to collective bargaining; (2) are justified by the state's interest in "labor peace" and avoiding "free riders"; and (3) do not significantly add to the burdening of free speech inherent in laws permitting extraction of service fees from non-members.[2] In a concurring and dissenting opinion, Justice Scalia—joined, in relevant part, by Justices O'Connor, Kennedy, and Souter—would have limited compelled contributions to "the costs of performing the union's statutory duties as exclusive bargaining agent," warning that Justice Blackmun's broad approach would engender further "confusion."[3] Neither approach was wholly unsupported, but rather were distillations of statements in prior case law.[4]

Applying *Ellis v. Railway Clerks,*[5] a private sector union case decided under the Railway Labor Act, Justice Blackmun's plurality opinion held that litigation "that does not concern the dissenting employees' bargaining unit" would be not be "germane" under the First Amendment either.[6] Lower courts, including the Third and the Sixth Circuits, have ruled otherwise by reasoning that only Chief Justice Rehnquist and Justices White and Stevens joined in that portion of the opinion; and, in any event, the holding in *Ellis* is limited to "direct contribution" of local union monies to litigation efforts by an affiliate "without expectations of reciprocal contributions."[7] Finding that, in *Locke*, local monies were "pooled" as part of a cost-sharing agreement, the First Circuit likewise distinguished *Ellis* and, under *Lehnert*'s three-part test, determined that the extra-unit litigation expenses at issue were chargeable.[8]

* *Phineas E. Leahey is an Associate in the Issues & Appeals practice in the New York office of Jones Day. He successfully represented a dissenting non-member employee in Seidemann v. Bowen. The opinions expressed here are solely those of the author and do not necessarily reflect those of Jones Day or any of that firm's clients, past, present, or future.*

Since *Lehnert* was decided, the composition of the Supreme Court has changed significantly,[9] which raises the possibility of the Court overruling the three-part test.

The narrow question presented in *Locke* concerns the ambiguity as to whether unions may charge non-members the costs of extra-unit litigation when there is a pooling arrangement between the exclusive bargaining agent and the affiliate.[10] Although Justice Blackmun considered extra-unit litigation "akin to lobbying"[11]—which is not chargeable unless related to contract "ratification or implementation"[12]—litigation may vary from the partisan political to the "germane."[13] Even Justice Scalia's discussion of "on demand" services for the *direct* benefit of the bargaining unit might be read to implicitly reject a per se ban on charging non-members for pooling extra-unit litigation expenses.[14] The current "case-by-case" approach of *Lehnert* supports the conclusion of the First Circuit that "litigation is not susceptible to a single label," but may be, on a particular set of facts, "expressive" or "central to the negotiation and administration of a collective bargaining agreement."[15]

But *Lehnert* is the product of a badly fractured court, which arrived at divergent views on the chargeability of six expenditures.[16] In its immediate aftermath, one commentator opined that the Court should "reexamine" chargeability jurisprudence in part because "unions will be lost in a destructive morass of judicial busy work such as the kind *Lehnert* fosters."[17] The same concern applies equally, if not more so, to non-members who, as the *Locke* petitioners note, are placed in the "untenable position of litigating for years or decades seeking refunds of money that should never have been collected from them."[18] Both non-members and unions are more inclined, under a fluid test, to what Justice Scalia referred to in *Lehnert* as "give it a try litigation."[19] Even though a five-justice agreement on the three-part test in 1991 is stare decisis, "when governing decisions are unworkable or are badly reasoned," the Court "has never felt constrained to follow precedent," particularly in constitutional cases.[20] Justice Scalia's approach, if clarified (or modified), would permit incursions on the First Amendment rights of non-members *only* when necessary to achieve the objective of the compulsory agency shop or to otherwise provide direct benefits to a local bargaining unit.[21]

Last year, in *Davenport v. Washington Educational Association*, Justice Scalia delivered a unanimous opinion for the Court upholding a Washington law requiring unions to obtain affirmative consent of non-members before using their agency fees for political causes.[22] As he noted there, a union's entitlement to any monies from non-members, even fees for collective bargaining, is a creature of federal or state law, not a constitutional mandate.[23] That does not mean a state legislature should prohibit all compelled contributions, nor that such laws are constitutionally infirm. Each legislature, to the outer limits of the First Amendment, may decide whether to prohibit compulsory dues, condition them on affirmative consent, or extract them under fair procedures that require the union to return a pro rata portion of them spent on activities unrelated to effectuation of collective bargaining agreements. Justice Scalia's approach in *Lehnert* would limit compelled contributions, where authorized by state or federal law, to expenses reasonably necessary to performing the duties of an exclusive bargaining agent; which, in turn, allows unions a fair measure of compensation for protecting the interests of non-members on a par with their own.

Justice Blackmun's critique of Justice Scalia's "statutory duties" approach highlights the generality of state laws authorizing agency shops, which he considered a "poor criteri[a]" for determining which charges violate the First Amendment because the obligations of the union "extend beyond those delineated in skeletal state labor law statutes."[24] But Justice Scalia would concede that his approach limits compelled contributions to a narrower set of circumstances commensurate with the union's state-mandated duties to represent non-members.[25] He does, however, incorporate state common law, including judicial construction of labor statutes in suits alleging breach of the duty of fair representation,[26] and he appears to allow risk pooling between local bargaining units and their affiliates.[27]

If the Court adopted Justice Scalia's test, then states would determine in the first instance the scope of the exclusive bargaining agents' duties. Even if state statutes and their common law varied in material respects, as noted, many do not even permit extraction of compulsory dues from non-members. That others would authorize exclusive bargaining agents to collect such fees and define their duties with varying degrees of specificity seems to be little cause for concern. Indeed, *Davenport* illustrates the manageability of allowing states to define the duties of exclusive bargaining agents and the corresponding obligations of non-members.[28]

Defining First Amendment limits by reference to state law might be a concern where legislatures or courts significantly broaden the duties of exclusive bargaining agents. But statutes expanding their duties outside the "financial core" of "collective bargaining, contract administration, and grievance adjustment"—a trio approved by the Court in *Communication Workers v. Beck*[29]—would be constitutionally suspect.[30] States should be permitted to establish, for public sector unions, a duty of lobbying for the narrow purpose of ratifying or implementing the non-member's collective bargaining agreement as well as pooling, but only for *reasonably anticipated* extra-unit expenses.[31] As such, the "statutory duties" test need not be so narrowly construed but would limit forced compensation to unions from non-members to expenses reasonably necessary for them to fulfill their legal, yet constitutional, mandates and, where appropriate, to provide the local unit direct benefits. Even assuming that "scant" guidance in state labor law rendered Justice Scalia's approach problematic, an alternative would be to limit chargeability to "financial core" duties, in addition to pooling for direct benefit to the local unit as a reasonable means of performing them.

While activities beyond this "whittled-down" core might benefit non-members, "private speech often furthers the interests of nonspeakers, and that does not alone empower the state to compel the speech to be paid for."[32] As the *Ellis* Court explained, "free riding" on "the union's organizing efforts outside the bargaining unit" is "not the type of free riding that [the Railway Labor Act] seeks to prevent."[33] Nor are non-representational activities the type of "free riding" that the First Amendment should permit a state to remedy. The Court has twice observed—implying that such limitations on chargeability would be appropriate—"*by allowing the union shop at all*, we have already countenanced a significant impingement on First Amendment rights."[34] Accordingly, the benefits from the union's performance of activities outside the scope of the union's duties as bargaining agent do not support more impingement on non-member rights than is reasonably necessary to effectuate the purpose of the agency shop and, in the case of pooled expenses, to provide for the direct benefit of the local bargaining unit.[35]

Endnotes

1 500 U.S. 507 (1991).

2 *Id.* at 519.

3 *Id.* at 551 (Scalia, J., concurring in part and dissenting in part). Justice Scalia disagreed with the principal opinion's expansive permission to charge dissenting non-members for the costs of the union's magazine relating to "teaching and education generally" and similar matters, sending delegates to an affiliate's convention—at least to the extent that the local unit's bargaining activities are not discussed—and preparing for an illegal strike. *Id.* at 559-562. With respect to extra-unit expenses, he rejected Justice Blackmun's view that there need not be "'a direct relationship between the expense at issue and some tangible benefit to the dissenter's bargaining unit.'" *Id.* at 562, quoting 500 U.S. at 522.

4 Both Justices relied on agency fee cases decided under federal labor statutes. Railway Employees v. Hanson, 351 U.S. 225 (1956); Machinists v. Street, 367 U.S. 740 (1941); Railway Clerks v. Allen, 373 U.S. 113 (1963); Ellis v. Railway Clerks, 466 U.S. 435 (1984); Communication Workers v. Beck, 487 U.S. 735 (1988). Because the Court has construed federal labor laws to be consistent with the First Amendment in order to remove doubts about their constitutionality, each Justice considered these cases relevant to public sector unions. *Lehnert*, 500 U.S. at 516; *id.* at 555 (Scalia, J., concurring in part and dissenting in part). Before *Lehnert*, cases decided under the First Amendment focused on the procedural requirements for public sector unions to exact service fees from non-members. Abood v. Detroit Bd. Of Educ., 431 U.S. 209 (1977); Chicago Teachers Union, Local No. 1 v. Hudson, 475 U.S. 292 (1986).

5 466 U.S. 435 (1984).

6 *Lehnert*, 500 U.S. at 528.

7 Otto v. Penn. Educ. Ass'n, 330 F.3d 125, 135-39 (3d Cir. 2003); Reese v. City of Columbus, 71 F.3d 619, 624 (6th Cir. 1995). Earlier decisions construed *Lehnert* to have adopted *Ellis*'s prohibition on compelling non-members to contribute to extra-unit litigation. *See, e.g.*, Pilots Against Illegal Dues v. Airline Pilots Ass'n ("PAID"), 938 F.2d 1123, 1129-31 (10th Cir. 1991).

8 Locke v. Karass, 498 F.3d 49, 58-59, 63 (1st Cir. 2007). In *Lehnert*, Justice Scalia commented that *Ellis* should reflect "the constitutional rule" applicable to public sector union cases. *Lehnert*, 500 U.S. at 555 (Scalia, J. concurring in part and dissenting in part). As the First Circuit noted, however, Justice Scalia did not "specifically" address extra-unit litigation and he approved of charges for services provided directly by affiliates to the bargaining unit. *Locke*, 498 F.3d at 59-60.

9 Chief Justice Rehnquist and Justices White, Marshall, Blackmun, and O'Connor are no longer on the Court and Chief Justice Roberts and Justices Thomas, Ginsburg, Breyer, and Alito have since been appointed. Even before these appointments, one commentator suggested that "with the departure of Justice Marshall from the Court, the next dues objector case to reach the

Court may well overturn the analytical structure of *Lehnert.*" Joseph A. Ciucci, *Defining the Permissible Uses of Objecting Members' Agency Dues: Is the Solution Any Clearer After Lehnert v. Ferris Faculty Ass'n*, 70 U. Det. Mercy L. Rev. 89, 122 (Fall 1992).

10 Conceivably, the discussion of extra-unit litigation in *Lehnert* is dicta because the expenditures at issue in the case concerned *reporting* on extra-unit litigation. 507 U.S. at 544-45 (Marshall, J. concurring in part and dissenting in part).

11 *Lehnert,* 500 U.S. at 528. For this proposition, Justice Blackmun cited *NAACP v. Button*, 371 U.S. 415 (1963), which pertained not to compulsory contributions to litigation, but a Virginia statute that curtailed access to the courts for the purpose of redressing racial discrimination.

12 *Id*. at 520.

13 *See Otto*, 330 F.3d at 139 (citing example of local contributions that "ensure the availability of resources for collective bargaining litigation").

14 *Lehnert*, 507 U.S. at 562 (Scalia, J. concurring in part and dissenting in part) ("expert consulting services on call, even in the years when they are not used," constitute a "tangible benefit" to the local union). Such pooled expenses may be "reasonably necessary to effective performance of the statutory duty of bargaining." *Id*. at n.4. Whether Justice Scalia would distinguish litigation as inherently ideological remains an open question.

15 *Locke*, 498 F.3d at 65.

16 Courts have often commented on the difficulty of applying the three-part test. *See, e.g.,* Bd. of Regents of Univ. of Wisconsin System v. Southworth, 529 U.S. 217, 232 (2000) (noting that in *Lehnert* "different Members of the Court reached varying conclusions regarding what expressive activity was or was not germane to the mission of the association"); Beckett v. Airline Pilots, 59 F.3d 1276, 1281 (D.C. Cir. 1995) (Silberman, J. concurring dubitante) ("it is impossible to detect in the Supreme Court cases—particularly *Lehnert*—a principled basis for distinguishing expenditures that are 'germane' from those that are not").

17 Calvin Siemer, *Lehnert v. Ferris Faculty Ass'n: Accounting to Financial Core Members: Much A-dues About Nothing?,* 60 Fordham L Rev. 1057, 1082 (Apr. 1992). Chargeability disputes may be resolved in court or by arbitration under a union's agency fee procedure. *Hudson*, 475 U.S. at 310 (requiring union to provide an "impartial decisionmaker"). After *Lehnert*, the Supreme Court decided *Airline Pilots Ass'n v. Miller*, 523 U.S. 866 (1998), which requires courts to adjudicate chargeability claims in the first instance, if requested by the non-member. *Id*. at 879-880. The burden has often fell on reluctant courts to apply *Lehnert*'s three-part test. *See, e.g.,* Weaver v. Univ. of Cincinnati, 970 F.2d 1523, 1536 (6th Cir. 1992) ("Courts should not involve themselves in the factual inquiries involved in making a chargeability determination").

18 Pet. Br. at 40-41, citing, *inter alia*, *Street*, 367 U.S. at 795-96 (Black, J. dissenting) (predicting that adjudicating agency fee claims "may prove very lucrative to special masters, accountants and lawyers," but "this formula, with its attendant trial burdens, promises little hope for financial recompense to the individual workers whose First Amendment freedoms have been flagrantly violated").

19 *Id*. at 551 (Scalia, J. concurring in part and dissenting in part).

20 Payne v. Tennessee, 501 U.S. 808, 827-30 (1991) (noting overruling of cases "decided by the narrowest of margins, over spirited dissents challenging the basic underpinnings of those decisions," or where they "have been questioned by Members of the Court in later decisions, and have defied consistent application by the lower courts").

21 Ordinarily, Supreme Court Rule 14.1(a) limits review to the issue raised in the petition for certiorari, *Yee v. City of Escondido,* 503 U.S. 519, 535 (1992), but "the Court has not always confined itself to the set of issues addressed by the parties," *Kolstad v. American Dental Ass'n*, 527 U.S. 526, 540 (1999) and, on occasion, has even overruled cases, despite the parties' consensus that existing precedent was controlling. Blonder Tongue Lab., Inc. v. Univ. of Ill. Foundation, 402 U.S. 313, 319-21, 320, n.6 (1971).

22 -- U.S. --, 127 S. Ct. 2372, 2383 (2007).

23 *Id*. at 2378. The legislatures in "right to work" states have prohibited exclusive bargaining agents from collecting any agency fees from non-members. *See, e.g.*, Ark. Const. amend. 34 § 1 ("nor shall any person against his will be compelled to pay dues to any labor organization as a prerequisite to or condition of employment"). These laws may be grounded in *Davenport's* view that agency shops "give a private entity, in essence, the power to tax government employees" which constitutes "an extraordinary *state* entitlement to acquire and spend *other people's* money." *Davenport*, 127 S.Ct. at 2378 & 2380 (emphasis in the original).

24 *Lehnert*, 500 U.S. at 526.

25 *Id*. at 562 (Scalia, J. concurring in part and dissenting in part).

26 *Id*. at 558, 559, n.3. The "agency shop provision" of the Michigan statute at issue in *Lehnert*, like that of many other states, advanced "much the same" government interests "as those promoted by similar provisions in federal labor law." *Abood,* 431 U. S. at 225. In *Locke*, the unions were certified as exclusive collective bargaining agents under the Maine law, 26 Me. Rev. Stat. Ann. 979, *et seq*. which permits them to charge non-members for "representational and bargaining services." *Opinion of the Justices*, 401 A.2d 135, 147 (Me. 1979).

27 *See supra* note 14. A literal or parsimonious "statutory duties" approach that did not allow unions to charge for expenses reasonably necessary to carry out their general directives would be inconsistent with the conclusions reached in Justice Scalia's concurring opinion. *Lehnert*, 500 U.S. at 532 n.6 (Blackmun, J.) (criticizing Justice Scalia's approach).

28 In upholding the Washington law, Justice Scalia observed that the prior decisions of the Court set a constitutional "floor" for non-member rights, not a "ceiling." *Davenport*, 127 S.Ct. at 2379.

29 487 U.S. 735, 745 (1988); *accord* Marquez v. Screen Actors Guild, Inc., 525 U.S. 33, 37-38 (1998) ("financial core" approach requires "only that employees pay the fees and dues necessary to support the union's activities as the employees' exclusive bargaining representative"). These were private sector cases decided under the National Labor Relations Act. Justice Blackmun noted the "'somewhat hazier' line between bargaining-related activities and purely ideological activities in the public sector," where unions devote considerable resources to secure legislative ratification of collective bargaining agreements. *Lehnert*, 500 U.S. at 520, quoting *Abood*, 431 U.S. at 236.

30 In *Abood*, 431 U.S. at 225, the Court upheld the constitutionality of the Michigan statute authorizing the agency shop because it limited the union's duties to "collective bargaining, contract administration, and grievance adjustment."

31 In contrast to ordinary administrative expenses, litigation, like lobbying, tends to be expressive. Therefore, extra vigilance is needed to protect against abuse of pooled expenditures for such purposes. In a lawsuit or arbitration, the burden should continue to be on unions to prove that their service charges are permissible. *Lehnert*, 500 U.S. at 524 (citations omitted). More should be required, however, at the pre-litigation *notice* stage. Although the union's notice to non-members listing chargeable and non-chargeable expenditures need not be"exhaustive," it should provide "sufficient information to gauge the propriety of the union's fee" relating to pooled expenses, *Hudson*, 475 U.S. at 306, without the need for non-members to file an action and request discovery only to find themselves improperly charged months, if not years, later.

32 *Lehnert*, 500 U.S. at 556 (Scalia, J. concurring in part and dissenting in part).

33 466 U.S. at 452-53.

34 *Hudson*, 475 U.S. at 302 n.8 (1986), quoting *Ellis*, 466 U.S. at 455 (emphasis added).

35 Under almost all agency shop provisions, only *objecting* non-members are entitled to a pro rata reduction of their agency fees for expenses unrelated to collective bargaining. If they choose to do so, unions may solicit non-members with their own funds to either rejoin the union or to waive recoupment of agency fees devoted to extra-statutory activities that benefit non-members.

Litigation

Stoneridge Investment Partners : The Supreme Court Rejects "Scheme Liability"

*By Larry Obhof**

On January 15, 2008, the Supreme Court issued its decision in *Stoneridge Investment Partners LLC v. Scientific-Atlanta, Inc.*, a case heralded by commentators as the "most important securities case in decades."[1] The *Stoneridge* decision rejects a theory of scheme liability that would have greatly expanded the universe of potential securities class action defendants.

What makes *Stoneridge* so important? In simple terms, the plaintiff sought to expand the scope of Section 10(b) actions beyond the securities markets and into the realm of ordinary business operations. The defendants were customers and suppliers to Charter Communications, Inc., the company that issued the securities in question. They did not directly mislead investors, "but were business partners with those who did."[2] If accepted by the Court, the plaintiff's theory of scheme liability could have extended the Section 10(b) private right of action to cover any transactions involving publicly-traded companies, so long as those transactions are later incorporated into the public company's financial statements. Such a "sweeping expansion" of the right of action would have exposed customers, suppliers, and other secondary actors to billions of dollars in liability when *other parties* make misstatements to the market.[3]

The Supreme Court prudently declined to extend the private right of action. It is well established that a plaintiff seeking to impose primary liability for securities fraud must prove reliance on *the defendant's* deceptive conduct, not the conduct of other parties. This requirement ensures that there is a causal connection between the defendant's misrepresentation and the plaintiff's injury. The *Stoneridge* plaintiff, however, did not rely on the defendants' alleged acts when purchasing or selling securities. Congress has repeatedly declined to extend the private right of action to cover such circumstances. The Court's decision in *Stoneridge* respects that choice and sends a strong signal that policymaking, including the decision to create or expand a cause of action, is properly left to Congress.

Section 10(b) of the Securities and Exchange Act makes it unlawful "[t]o use or employ, in connection with the purchase or sale of any security… any manipulative device or contrivance in contravention of such rules and regulations as the [Securities and Exchange] Commission may prescribe."[4] Pursuant to this section, the SEC promulgated Rule 10b-5, which makes it unlawful "[t]o employ any device, scheme, or artifice to defraud… [or] engage in any act, practice, or course of business which operates or would operate as a fraud or deceit upon any person… in connection with the purchase or sale of any security."[5] Rule 10b-5 encompasses only conduct already prohibited by Section 10(b).[6]

Although the text of the Securities and Exchange Act does not provide for a private cause of action for Section 10(b) violations, the Supreme Court has found an implied private right of action in the statute and Rule 10b-5.[7] A plaintiff bringing a Section 10(b) private action must prove "(1) a material misrepresentation or omission by the defendant; (2) scienter; (3) a connection between the misrepresentation or omission and the purchase or sale of a security; (4) reliance upon the misrepresentation or omission; (5) economic loss; and (6) loss causation."[8]

The Supreme Court has made clear that the implied private right of action does not extend to aiders and abettors of securities fraud. In *Central Bank of Denver, N.A., v. First Interstate Bank of Denver, N.A.*, the Court held that "a private plaintiff may not maintain an aiding and abetting suit under §10(b)."[9] The lack of a private action for aiding and abetting is not an oversight—Congress imposed other forms of secondary liability as part of the 1934 Act. Thus, *Central Bank* points to the "deliberate congressional choice" against imposing secondary liability in private securities fraud actions.[10]

This does not mean that secondary actors are always free from liability. Any person or entity that "employs a manipulative device or makes a material misstatement (or omission) on which a purchaser or seller of securities relies" may still be liable as a *primary* violator under Rule 10b-5, as long as all of the usual requirements for liability are met.[11] For example, primary liability could attach where the secondary actor himself disseminates or transmits false information to investors, such as when an accountant knowingly certifies false financial statements or an attorney knowingly prepares false opinion letters.[12] Aiding and abetting, however, falls short of the mark. A plaintiff "must show reliance on *the defendant's* misstatement or omission to recover under 10b-5."[13] By its very nature, a claim for aiding and abetting seeks to impose liability on a secondary actor for facilitating the primary actor's misstatements or omissions. Investors rely upon *those* misstatements or omissions—which are made only by the primary actor—when purchasing or selling securities. Investors are not aware of, and thus do not rely on, the conduct of the secondary actor. A plaintiff's reliance on representations made by someone other than the defendant cannot form the basis of liability.[14]

Congress explicitly addressed the issue of secondary liability in the aftermath of *Central Bank*. Rather than restoring private aiding-and-abetting liability, it left enforcement to the government and effectively removed private plaintiffs from the equation.[15] Congress enacted Section 20(e), giving the SEC, but not private litigants, the authority to prosecute parties who provide "substantial assistance" to those engaged in securities fraud.[16] "Congress decided, both when it enacted Section 20(e) in 1995 and again when it enacted Sarbanes-Oxley in 2002—*not* to extend the right to enforce this liability to private plaintiffs."[17] Thus, Congress has consistently rejected the idea

....

** Larry Obhof is an associate at Kirkland & Ellis LLP. The views expressed in this article are his own.*

of secondary liability in private securities fraud actions, both before and after *Central Bank*.

The *Stoneridge* complaint alleged that Charter Communications, Inc., engaged in a pervasive fraudulent scheme intended to artificially boost its reported financial results.[18] Among other things, Charter overstated its operating cash flow by hundreds of millions of dollars for both 2000 and 2001.[19] The market price of Charter's securities fell substantially when its financials were eventually restated to reflect economic reality.[20] Stoneridge Investment Partners subsequently brought a securities fraud class action on behalf of Charter's shareholders. In addition to Charter and its executives, the plaintiff named as defendants Arthur Anderson, LLP, which had served as Charter's independent auditor during the class period, and two equipment vendors, Scientific-Atlanta, Inc. and Motorola, Inc. (the "Vendors").

How could the plaintiff sue the Vendors for *Charter's* misstatements? Stoneridge Investment Partners attempted to circumvent the limitations of *Central Bank* by pleading a theory of scheme liability. The plaintiff alleged that the Vendors entered into "wash" transactions with Charter—transactions that had no economic substance, but enabled Charter's overstatement of its revenue and operating cash flow. Charter agreed to pay the Vendors excessive amounts for the set-top cable boxes they provided, with the understanding that the Vendors would then use the additional funds to purchase advertising from Charter.[21] The companies drafted documents to make it appear as though the transactions were unrelated. For example, "Scientific-Atlanta sent documents to Charter stating—falsely—that it had increased production costs."[22] The set-top box agreements were backdated to make it appear as though they were negotiated a month before the advertising agreements.[23]

According to Stoneridge Investment Partners, the Vendors' actions had the purpose and effect of furthering Charter's scheme to overstate its revenue and cash flow.[24] Charter improperly capitalized its increased equipment expenses, but treated the returned advertising fees as immediate revenue.[25] This allowed Charter to inflate its revenue and operating cash flow by approximately $17 million in the fourth quarter of 2000.[26] Stoneridge Investment Partners argued that the Vendors were more than aiders and abettors of Charter's fraud—they were *primary* violators because "they engaged in classic fraudulent behavior themselves."[27]

Although Stoneridge Investment Partners labeled its theory "scheme liability," the allegations set out a model example of the type of secondary liability already prohibited by *Central Bank*.[28] The plaintiff alleged "fraudulent practices engaged in *by Charter*… to present a false picture of financial growth and success."[29] The Vendors' deceptive acts did not relate to the purchase or sale of securities—they involved the sale of goods and the purchase of advertising. The Vendors played no role in preparing Charter's misleading financial statements. [30] The Vendors "did not themselves disseminate the false information to the securities market."[31]

The plaintiff's claims closely resembled the statutory definition of aiding and abetting. Section 20(e) defines aiding and abetting as "knowingly provid[ing] substantial assistance" to one who commits securities fraud. Stoneridge Investment Partners used very similar terms to define its allegations against Scientific-Atlanta and Motorola. The question presented by the petitioner's brief alleged that the Vendors "enabled" the publication of inflated financial statements by Charter.[32] Whether one uses the words "providing assistance" or "enabling," the crux of the claim is the same: Stoneridge and its lawyers sought to impose liability against the Vendors because they engaged in business transactions with Charter, and *Charter* later accounted for those transactions improperly.[33]

The *Stoneridge* decision makes clear that this chain of events is too remote to impose liability on the Vendors. Secondary actors can be held liable for securities fraud where *all* of the requirements for primary liability are met. The *Stoneridge* complaint, however, is deficient in at least one regard: it does not allege that Stoneridge Investment Partners (or any other investors) relied upon the Vendors' statements when purchasing or selling Charter's stock.[34] Reliance is an essential element of the Section 10(b) cause of action. The requirement ensures that there is a causal connection between the defendant's misrepresentation and the plaintiff's injury.[35]

While courts will often presume reliance on the part of shareholders, neither of the reasons for that presumption apply to the facts of *Stoneridge*.[36] The Vendors had no duty to disclose facts to Charter's shareholders.[37] Because the Vendors' deceptive acts were not communicated to the public, the fraud-on-the-market doctrine does not apply.[38] Thus, the only possible reliance in *Stoneridge* is indirect. It was Charter, not the Vendors, that filed the fraudulent financial statements. Investors relied only on *Charter's* deceptive acts when purchasing or selling its stock. Stoneridge Investment Partners tried to side-step this problem by arguing that in an efficient market investors rely not only upon the public documents relating to a security but also the transactions that those statements reflect.[39] Under this theory, the cause of action could reach any company with which the issuer does business, because all transactions with the issuer are ultimately incorporated into its financial statements. The *Stoneridge* decision rejects this expansive theory of indirect reliance, bluntly stating that "there is no authority" for such a rule.[40]

Like the Court in *Central Bank*, the *Stoneridge* majority emphasizes that Congress has considered the issue of secondary liability and made a deliberate choice not to extend the private right of action. "Petitioner's theory," Justice Kennedy writes, "would put an unsupportable interpretation on Congress' specific response to *Central Bank*."[41] "Were we to adopt this construction… we would undermine Congress' determination that this class of defendants should be pursued by the SEC and not by private litigants."[42] The majority also sends a strong signal that courts should not be in the business of creating or expanding causes of action. The Court will not find an implied cause of action unless the underlying statute demonstrates the intent to create one.[43] Where courts have already created a cause of action—such as the implied private right of action found in Section 10(b) and Rule 10b-5—the decision to extend the cause of action must be made by Congress.[44]

The majority's opinion suggests that courts, moving forward, must be more respectful of Congress' role as the creator of federal statutory claims. This emphasis on the separation of powers is notable. *Stoneridge* represents a significant shift in the way the Court views its role relative to that of Congress. The Section 10(b) private cause of action is, after all, a judicial construct. The statutory text does not provide for private claims.[45] The Court has acknowledged that "there is no indication that Congress, or the Commission when adopting Rule 10b-5, contemplated such a remedy."[46] Nevertheless, in the 1970s the Supreme Court "acquiesced" in the lower courts' "acceptance" of a private right of action.[47]

In the intervening decades the Supreme Court has reigned in the impulse to create new causes of action. It is now well settled that private rights of action to enforce federal law must be created by Congress, rather than the courts.[48] The judicial task is not to determine whether particular remedies are *desirable*. Rather, a court's duty is to "interpret the statute Congress has passed" to determine whether there is a Congressional intent to create a particular remedy.[49] If Congress does not intend to create a private right and a private remedy, "a cause of action does not exist and courts may not create one."[50]

Stoneridge directly addresses the separation of powers and puts to rest any doubts about the relative roles of the courts and Congress. "In the absence of congressional intent the Judiciary's recognition of an implied private right of action 'necessarily extends its authority to embrace a dispute that Congress has not assigned itself to resolve.'"[51] It is Congress, not the courts, which must decide whether to extend the Section 10(b) to secondary actors like the *Stoneridge* defendants. Congress chose not to extend the private right of action to cover this type of liability. The *Stoneridge* decision respects that choice and properly defers to the legislative branch.[52]

In retrospect, of course, the claim that *Stoneridge* is the "most important securities case in decades" may seem a bit hyperbolic. That is only true because we know the outcome. Adopting the plaintiff's theory of scheme liability would have been a significant departure from settled law. The Section 10(b) cause of action would have extended beyond the securities markets into the realm of ordinary business operations.[53] As the Court aptly states, "the federal power would be used to invite litigation beyond the immediate sphere of securities litigation and in areas already governed by functioning and effective state-law guarantees."[54]

The practical results of this change would have been significant. If securities class actions were untethered from the element of reliance, there would be little limitation on the number of potential class action defendants or the scope of their potential liability. Any transaction ultimately accounted for in a public company's financial statements could become the subject of a claim for securities fraud. Section 10(b)'s implied cause of action would effectively reach the entire marketplace in which publicly traded companies do business.[55] The consequences of such an expansive rule are not lost on the Court. The *Stoneridge* majority emphasizes that scheme liability would "expose a new class of defendants," including innocent parties, to increased "uncertainty and disruption."[56] According to the Court, this would effectively raise the cost of doing business in the United States, thereby deterring foreign investment and shifting securities offerings away from domestic capital markets.[57]

Of course, whether scheme liability would cause unintended harm is a separate question from whether the plaintiff's theory properly fits within Section 10(b). Even where *Stoneridge* discusses the practical consequences of the plaintiff's theory, it is clear that the Court bases its decision on law, rather than on policy. For example, although the majority worries aloud that scheme liability would "reach the whole marketplace," the Court does not rely on that fact. The majority rejects the plaintiff's theory because "there is no authority" for such a broad expansion of the implied right of action.[58] "Congress rather than the courts controls the availability of remedies for violations of statutes."[59]

It is worth mentioning what *Stoneridge* does *not* do. The Court does not absolve secondary actors from all liability. Parties engaging in (or facilitating) securities fraud can (and should) be punished. Secondary actors are still subject to criminal penalties and civil enforcement by the SEC.[60] The SEC may obtain injunctive relief, issue administrative orders, and impose large civil penalties on any companies engaged in aiding and abetting fraud.[61] These enforcement mechanisms are not toothless. In fiscal year 2006 alone, the Commission initiated 914 investigations, 218 civil proceedings, and 356 administrative proceedings.[62] That same year, the Commission recouped over $3.3 billion in disgorgement and other penalties.[63] Similarly, the Department of Justice's Corporate Fraud Task Force has obtained more than 1,200 corporate fraud convictions in the past five years.[64] Some states' securities laws also permit state authorities to seek fines and restitution from aiders and abettors.[65]

Nor are secondary actors immune from private suit. *Stoneridge* does not affect shareholders' ability to pursue actions against secondary actors who commit primary violations.[66] As before, a plaintiff may allege primary liability where all of the usual requirements, including reliance, are met. The securities statutes also provide an express private right of action against accountants and underwriters in certain circumstances.[67] Where a party's fraud involves transactions unrelated to the purchase or sale of securities—such as with the sale of goods or purchase of advertising—plaintiffs will have causes of action for fraud. They just will not have claims for *securities* fraud. That limitation is consistent with the statutory scheme, which was designed to provide remedies for securities-related misconduct, not to serve as a catchall federal remedy for fraud.[68]

Although *Stoneridge* had the potential to be the most important securities case in decades, the Court's decision is perhaps best viewed as an affirmation of the status quo. The Court dutifully applies Central Bank and respects Congress' decision not to extend the private right of action to cover this type of liability. *Stoneridge* shows the Supreme Court's reluctance to find new implied causes of action or to expand existing ones. It is a strong reaffirmation of the principle that Congress, not the courts, determines the remedies for violations of federal statutes.

Endnotes

1 Stoneridge Investment Partners LLC v. Scientific-Atlanta, Inc., 128 S. Ct. 761 (2008); *see* Richard A. Epstein, *Primary and Secondary Liability Under Securities Law: The Stoneridge Investment Saga*, Oct. 9, 2007, *at* http://www.pointoflaw.com/columns/archives/004373.php (describing *Stoneridge* as "one of the most important securities law cases in decades"); Edward Iwata, *Fewer Lawsuits Charge Securities Fraud*, ABC News, Oct. 7, 2007, *available at* http://abcnews.go.com/Business/story?id=3700843&page=1 (describing *Stoneridge* as "the most important securities-fraud case in decades").

2 *See* Lyle Denniston, "Court Limits Securities Fraud Lawsuits," Jan. 15, 2008, *at* http://www.scotusblog.com/wp/uncategorized/court-limits-securities-fraud-law ("The Supreme Court, in one of the most important securities law rulings in years, decided Tuesday that fraud claims are not allowed against third parties that did not directly mislead investors but were business partners with those who did.").

3 Brief for the United States as Amicus Curiae at 9, Stoneridge Investment Partners LLC v. Scientific-Atlanta, Inc., 128 S. Ct. 761 (2008) (Case No. 06-43)(hereinafter "Brief for the United States").

4 15 U.S.C. § 78j(b).

5 17 C.F.R. § 240.10b-5.

6 *Stoneridge*, 128 S. Ct. at 768; *see also* United States v. O'Hagan, 521 U.S. 642, 651 (1997) ("Liability under Rule 10b-5… does not extend beyond conduct encompassed by § 10(b)'s prohibition.").

7 *Stoneridge*, 128 S. Ct. at 768 (citing Superintendent of Ins. of N.Y. v. Bankers Life & Casualty Co., 404 U.S. 6, 13 & n.9 (1971)).

8 *Id.* (citing Dura Pharm., Inc. v. Broudo, 544 U.S. 336, 341-42 (2005)).

9 511 U.S. 164, 191 (1994).

10 *See id.* at 184 ("The fact that Congress chose to impose some forms of secondary liability, but not others, indicates a deliberate congressional choice with which the courts should not interfere.").

11 *Id.* at 191.

12 *See* Brief for Richard I. Beattie, et al., as Amici Curiae at 14, Stoneridge Inv. Partners LLC v. Scientific-Atlanta, Inc., 128 S. Ct. 761 (2008) (Case No. 06-43) ("[T]he sort of liability-producing conduct by a secondary actor that the [*Central Bank*] Court had in mind… was the dissemination or transmission of materially inaccurate information to investors by the secondary actor itself, such as an accountant's act of certifying misleading financial statements or an attorney's preparation of false opinion letters.") (quotations and alterations omitted).

13 *Central Bank*, 511 U.S. at 180 (emphasis added).

14 *See* Anixter v. Home-Stake Production Co., 77 F.3d 1215, 1225 (10th Cir. 1996) (quoting *Central Bank*, 511 U.S. at 177).

15 *See* Brief of the Washington Legal Foundation as Amicus Curiae at 20, Stoneridge Inv. Partners LLC v. Scientific-Atlanta, Inc., 128 S. Ct. 761 (2008) (Case No. 06-43) ("Instead of heeding calls for the restoration of private aiding-and-abetting liability, Congress sought to 'remov[e] the plaintiffs' class action bar from the equation.'") (quoting 4 Bromberg & Lowenfels, Securities Fraud & Commodities Fraud § 7:308, at 7-506 (2d ed. 2006)).

16 *See id.*; *see also* 15 U.S.C. § 78t(e) ("For purposes of *any action brought by the Commission*… any person that knowingly provides substantial assistance to another person in violation of a provision of this chapter… shall be deemed to be in violation of such provision….") (emphasis added).

17 *See* Brief for Former SEC Commissioners and Officials and Law and Finance Professors as Amici Curiae at 12, Stoneridge Inv. Partners LLC v. Scientific-Atlanta, Inc., 128 S. Ct. 761 (2008) (Case No. 06-43)[hereinafter "Brief for Former SEC Commissioners"] (citing S. Rep. No. 104-98, at 48 (1995); H.R. Rep. No. 107-414, at 54 (2002); 148 Cong. Rec. S6584 (daily ed. July 10, 2002)). Congress has been explicit in rejecting private aiding and abetting liability. For example, according to the Senate Committee Report accompanying the 1995 bill, "provid[ing] for private aiding and abetting liability actions under Section 10(b) would be contrary to [the bill's] goal of reducing meritless securities litigation." S. Rep. No. 104-98, at 48 (1995).

18 *In re* Charter Communications, Inc., Securities Litigation, 443 F.3d 987, 989 (8th Cir. 2006).

19 *See, e.g.*, Brief for the United States at 23 (noting that Charter's operating cash flow was overstated by $195 million in 2000 and by $292 million in 2001).

20 *See* Brief for Petitioner at 9, Stoneridge Inv. Partners LLC v. Scientific-Atlanta, Inc., 128 S. Ct. 761 (2008) (Case No. 06-43)(hereinafter "Brief for Petitioner").

21 *Stoneridge*, 128 S. Ct. at 766.

22 *Id.* at 767.

23 *Id.*

24 *See, e.g.*, Brief for Petitioner at 14 ("Respondents' deceptive acts had the purpose and effect of furthering the fraudulent scheme."); *id.* at 38 ("the scheme in which Respondents engaged had the purpose and effect of artificially increasing Charter's revenue and cash flow reflected in Charter's financial statements").

25 It is worth noting that Charter could have accounted for these transactions in a way that would have rendered its financial statements accurate. Indeed, the Vendors properly accounted for the transactions in their financial statements. They "booked the transactions as a wash, under generally accepted accounting principles." *Stoneridge*, 128 S. Ct. at 767.

26 *Id.*; *see also In re Charter Communications, Inc.*, 443 F.3d at 989-90.

27 Brief for Petitioner at 26; *see id.* ("They participated in transactions that they knew to be shams, and they falsified records about those transactions…. This is just the kind of conduct that the language of Section 10(b) and Rule 10b-5 forbids.").

28 This point was made succinctly in a brief filed by a group of former SEC officials and prominent law and finance professors: "Petitioner's 'scheme liability' theory is simply a semantic ploy designed to recast secondary conduct as a primary violation." Brief for Former SEC Commissioners at 5; *id.* at 10 (same).

29 Brief for Petitioner at 5 (emphasis added).

30 *Stoneridge*, 128 S. Ct. at 767.

31 Brief for Petitioner at 38.

32 *Id.* at i (emphasis added).

33 *See In re Charter Communications, Inc.*, 443 F.3d at 991 ("plaintiffs contend that the Vendors are liable to Charter's investors on the basis that they engaged in a business transaction that Charter improperly accounted for") (citation and alteration omitted).

34 *See Stoneridge*, 128 S. Ct at 769 ("[R]espondents' acts or statements were not relied upon by the investors… as a result, liability cannot be imposed upon respondents.").

35 *Id.*

36 Courts will presume reliance in two circumstances. First, if there is an omission of material fact by one who has a duty to disclose, an investor to whom the duty was owed need not prove reliance. *Id.* (citing Affiliated Ute Citizens of Utah v. United States, 406 U.S. 128, 153-54 (1972)). Second, under the fraud-on-the-market doctrine, reliance is presumed when a defendant's misstatements become public because that information is reflected in the market price of the security. *See id.* (citing Basic Inc. v. Levinson, 485 U.S. 224, 247 (1988)).

37 *Id.*

38 *See id.* ("their deceptive acts were not communicated to the public"); *id.* at 770 ("respondents' deceptive acts… were not disclosed to the investing public").

39 *See id.* at 770.

40 *Id.*

41 *Id.* at 771.

42 *Id.*

43 *See id.* at 772 ("Though the rule once may have been otherwise … it is settled that there is an implied cause of action only if the underlying statute can be interpreted to disclose the intent to create one.") (citations omitted).

44 *Id.* at 773 ("Concerns with the judicial creation of a private cause of action caution against its expansion. The decision to extend the cause of action is for Congress, not for us.").

45 *See* Lampf, Pleva, Lipkind, Prupis & Petigrow v. Gilbertson, 501 U.S. 350, 358 (1991) ("The text of § 10(b) does not provide for private claims."); Ernst & Ernst v. Hochfelder, 425 U.S. 185, 196 (1976) (noting that "§ 10(b) does not by its terms create an express civil remedy for its violation").

46 *Ernst & Ernst*, 425 U.S. at 196 (footnotes omitted); *see also Lampf*, 501 U.S. at 358-59 ("Such claims are of judicial creation…. [W]e have made no pretense that it was Congress' design to provide the remedy afforded.").

47 Touche Ross & Co. v. Redington, 442 U.S. 560, 577 (1979) ("[T]his Court simply explicitly acquiesced in the twenty-five-year-old acceptance by the lower federal courts of an implied action under § 10(b)."); Cannon v. Univ. of Chicago, 441 U.S. 677, 693 n.13 (1979) ("the Court explicitly acquiesced in the 25-year-old acceptance by the lower federal courts of a Rule 10b-5 cause of action").

48 Alexander v. Sandoval, 532 U.S. 275, 286 (2001); *see Touche Ross & Co.*, 442 U.S. at 578 (remedies available are those "that Congress enacted into law").

49 *Alexander*, 532 U.S. at 286; *see also Touche Ross & Co.*, 442 U.S. at 575 ("The central inquiry remains whether Congress intended to create, either expressly or by implication, a private cause of action.").

50 *Alexander*, 532 U.S. at 286-87.

51 *Stoneridge*, 128 S. Ct. at 772 (quoting Cannon v. Univ. of Chicago, 441 U.S. 677, 746 (1979) (Powell, J., dissenting)).

52 *Id.* at 773 ("The decision to extend the cause of action is for Congress, not for us.")

53 *See id.* at 770 ("The petitioner invokes the private cause of action under § 10(b) and seeks to apply it beyond the securities markets—the realm of financing business—to purchase and supply contracts—the realm of ordinary business operations.").

54 *Id.* at 771.

55 *Id.* at 770.

56 *Id.* at 772.

57 *Id.*

58 *Id.* at 770.

59 *Id.* at 773 (quoting Wilder v. Virginia Hosp. Ass'n, 496 U.S. 498, 509 n.9 (1990)).

60 *See id.* (citing 15 U.S.C. § 78ff (providing for criminal penalties) and 15 U.S.C. § 78t(e) (providing or civil enforcement by the SEC)).

61 *See, e.g.*, 15 U.S.C. §§ 78u, 78u-3.

62 U.S. Securities and Exchange Commission, 2006 Performance and Accountability Report 8 (2006).

63 *Id.*

64 Department of Justice, Fact Sheet: President's Corporate Fraud Task Force Marks Five Years of Ensuring Corporate Integrity (July 17, 2007).

65 *Stoneridge*, 128 S. Ct. at 773 (citing Del. Code Ann., Tit. 6, § 7325 (2005)).

66 *Id.* at 773-74 (citing *Central Bank*, 511 U.S. at 191).

67 *Id.* at 773 (citing 15 U.S.C. § 77k).

68 *See* Marine Bank v. Weaver, 455 U.S. 551, 556 (1982) ("Congress, in enacting the securities laws, did not intend to provide a broad federal remedy for all fraud.")

Canadian Class Action Law: A Flawed Model for European Class Actions

*By John Beisner, Karl Thompson & Allison Orr Larsen**

Activists urging European Union nations to adopt the class action device have recently begun citing Canada as a model. Like the United States, Canada has adopted formal class action rules that permit plaintiffs to bring class proceedings. And there is a perception that, to date, Canada has been spared the sort of rampant, U.S.-style class action litigation that has been widely criticized for imposing "huge, avoidable, and unnecessary cost[s]" on the economy.[1] Some European class action advocates have therefore suggested that Canada provides a guide to creating a class action regime without opening a "Pandora's box" leading to the American experience.[2]

On closer inspection, however, Canadian class action law turns out not to be the ideal model it is sometimes claimed to be. European policy makers should think twice before importing it. Perhaps the most striking feature of Canadian class action law is its exceptionally permissive standard for class certification, which makes it markedly easier to certify a class in Canada than in the United States. This feature of the Canadian system stems not simply from permissiveness among Canadian judges or any other similarly contingent cause but rather from a core *structural* feature of Canadian class action law itself: the lack of a formal "predominance" requirement prohibiting class certification unless common issues predominate over individualized issues in a proposed class.

The permissiveness of Canadian class certification procedures makes Canada a dangerous model for Europe, because it opens the door to the kind of class action abuse found in the United States. As experience in the U.S. and Canada shows, class certification is the key battleground in class action litigation: although class action statutes typically contemplate that certification will only be the first step in a class proceeding, the reality is that in nearly all instances, it is the last step as well. This is so because once a class is certified the risk of trial loss to a company's financial stability can be so great, even as to extremely weak claims, that corporations have no choice but to settle. Permissive class certification rules thus create a powerful and dangerous incentive that encourages plaintiffs to file weak or even frivolous claims, simply to take advantage of their potential settlement value. This value is enhanced—and the incentive to file such suits heightened—because permissive class certification rules also lead to inherently unfair and abusive class procedures, such as "perfect plaintiff" classes, "common issues" trials, and "trials by statistics," that (as discussed below) tilt class proceedings sharply in plaintiffs' favor.

These kinds of abusive procedures have already appeared in Canada, apparently as a direct result of its permissive certification rules. However, to date, other features of the Canadian system, such as the "loser pays" rule and the implicit bar on contingency fee arrangements, appear to have kept the volume of class action litigation below that of the U.S. But class actions in Canada are a relatively recent innovation, and the full character of the Canadian system probably has not fully emerged—in the United States, for example, pervasive class action abuse took decades to develop fully. European policymakers should therefore not simply assume that procedural rules like fee-shifting authorizations or contingency fees prohibitions will hold the line against class action abuse in the long run.

Notably, U.S. plaintiffs' lawyers have recently begun to use Canadian class actions as an integral part of cross-border litigation strategies, filing parallel class actions in the U.S. and Canada, attempting to use U.S. discovery to support Canadian claims, and generally taking advantage of efficiencies of scale to decrease the cost of bringing class actions in foreign countries. The ultimate impact of these kinds of strategies on Canadian class actions remains to be seen. But given this uncertainty, it would be foolish to assume that Canada is somehow immune from the kind of class action abuse found in the United States, simply because such abuse has not yet fully blossomed.

Instead of relying on this kind of rudimentary (and misleading) empirical correlation, European policymakers should look more closely at the substance of Canadian class action law itself. That law is characterized by liberal certification rules that create a powerful incentive to file class action claims—regardless of merit. It may be that, in some instances, this incentive can be counterbalanced (at least temporarily) by other features of a legal system. But rather than attempting to adjust these conflicting incentives to achieve a balance that avoids U.S.-style abuse, European policymakers would be far better served by rejecting the Canadian model and declining to import this flawed structure in the first place.

I. History of Class Actions in Canada

The permissiveness of Canadian class certification procedures is no accident. The history of Canadian class actions evidences a clear effort to create a procedure that is more amenable to certification than the U.S. model. Class actions (or "class proceedings") are a relatively recent innovation in Canada. Although Quebec became the first Canadian province to adopt class action legislation in 1978,[3] the move toward widespread class action legislation in Canada did not begin until 1982, when the Ontario Law Reform Commission issued a Report on Class Actions. That report set forth three objectives for future lawmakers to consider—(improving judicial efficiency, affording greater access to justice for individuals, and achieving behavior modification among manufacturers)—and recommended that, to meet these objectives, provinces adopt a class certification procedure modeled after U.S. Federal Rule of Civil Procedure 23.[4] The Commission's recommendation, however, departed from U.S. law in one crucial respect: rather than a *separate* requirement that common questions predominate over those that are specific to individual members of the class, the report recommended that courts consider predominance as just "one of the factors employed to gauge whether the class action is superior."[5]

** John Beisner, Karl Thompson, and Allison Orr Larsen are attorneys in the Washington office of O'Melveny & Myers LLP.*

The report had little impact at first, but in the mid-1990s two provinces passed significant legislation aimed at accomplishing the Reform Commission's missions: the Ontario Class Proceedings Act in 1992, and the British Columbia Class Proceeding Act of 1995.[6] Both laws "draw heavily on the experience under Rule 23 of the U.S. Federal Rules of Civil Procedure," but, like the Commission's report, are in certain respects substantially more permissive of class certification than the U.S. rule.[7] This was deliberate: the drafters of the Ontario Act "expressly adopted a regime thought to be more hospitable to the certification of class actions than had previously been exhibited for many years in the United States."[8] Under both the Ontario and the British Columbia statutes, courts "shall" or "must" grant certification of a class if seven conditions are met:

1. The pleadings invoke a cause of action;[9]
2. there is an identifiable class of two or more persons;
3. the claims of the class members raise common issues;
4. the class proceeding is the "preferable procedure" for resolving the common issues;
5. the class representative fairly and adequately represents the interests of the class;
6. the class representative produces a plan for the proceeding; and
7. the class representative does not have a conflict with the interests of the class members.[10]

These laws have set the legal standard under which most class actions in Canada are litigated. Ontario and British Columbia are two of the most populated Canadian provinces; further, courts in those jurisdictions are willing to approve national class actions, certifying classes even when some (or even most) members of the class reside in other provinces.[11] In addition, with the exception of Quebec—whose certification laws are even more liberal—other Canadian provinces and territories that have enacted comprehensive class action legislation have followed the lead of the Ontario and British Columbian legislatures and adopted substantially similar certification requirements.[12]

Today, Canadians in every province have the option of bringing class actions, even if the province in which they reside has not passed class action legislation. In 2001, the Canadian Supreme Court decided *Western Canadian Shopping Centres, Inc. v. Dutton*, which extended class actions to all Canadian jurisdictions, holding that "[a]bsent comprehensive legislation, the courts must fill the void [using] their inherent power."[13] After surveying the class action legislation enacted throughout the country, the Court pointed to the Ontario and British Columbia statutes as models that should "provide guidance" to courts throughout Canada. It then distilled four minimum criteria for the certification of a class:

1. The class must be "capable of clear definition;"
2. there "must be issues of fact or law common to all class members;"
3. there cannot be conflicting interests among class members—victory for one must mean victory for all; and
4. the class representative must adequately represent the class.[14]

The Court also articulated several discretionary factors that could be considered in a class certification proceeding, including whether "[c]lass members... raise important issues not shared by all members of the class."[15] Accordingly, in these provinces as well, predominance is at best a discretionary factor, rather than a required element for class certification.

II. Canada's Liberal Class Certification Test

The enactment of class proceedings legislation in Canada has, unsurprisingly, led Canadian courts to grant certification "in many cases that would not traditionally have qualified as representative actions" under the common law.[16] More significantly—particularly for those who would look to Canada as a model of class-action restraint—the enactment of class action legislation in Canada has produced a system that is even more liberal in its certification policy than the U.S. class action regime.[17]

In the United States, Rule 23 of the Federal Rules of Civil Procedure provides that a class may not be certified unless "[common] questions of law or fact... predominate over any questions affecting only individual members."[18] As a result, in U.S. class action proceedings, federal and state courts "devote a great deal of attention to whether there is sufficient identity of interest or cohesiveness among the class members" to warrant certification, "particularly when separate transactions, acts, or omissions are involved."[19] For example, putative class actions in the U.S. are generally "doomed to fail" if part of the proposed lawsuit involves evaluating injuries and damages on an individual basis.[20] The rationale behind the predominance requirement is, as the U.S. Supreme Court has explained, to ensure that "proposed classes are sufficiently cohesive to warrant adjudication by representation"[21]—i.e., to ensure that class action litigation does not devolve into multiple trials of issues and facts relevant to only some of the class members.[22]

In practical terms, the predominance requirement serves an important screening function: it is "an attempt to achieve a balance between the value of allowing separate actions to be instituted so that individuals can protect their own interests and the economy that can be achieved by allowing a multiple party dispute to be resolved on a class-action basis."[23] Put differently, the predominance requirement exists to ensure that the class action device is not unmoored from its fundamental purpose—namely, "to facilitate the adjudication of disputes involving *common* questions and multiple parties in a single action."[24]

In Canada, "the threshold for class certification" is "generally considered to be lower than in the United States," mainly because there is no formal predominance requirement.[25] Legislatures in Saskatchewan, British Columbia, Alberta, Newfoundland, and Manitoba have purposely and expressly disclaimed this requirement in the text of their laws, directing courts to certify classes if the class members raise common issues, regardless of "whether... those common issues predominate over issues affecting only individual members."[26] Laws in other provinces (e.g., Ontario) are silent about the predominance requirement, but at the direction of the Supreme Court, courts there, as elsewhere, have consistently required only that class members share "a substantial common ingredient;" they have not found it "necessary that common issues predominate over non-common" ones.[27] Laws in one province—Quebec—have an even *lower* commonality requirement than the rest of Canada,[28] permitting certification if the class members' claims raise "identical, similar, or *related* questions of law or fact."[29]

The permissiveness of class certification procedures in Canada is (at least in some instances) exacerbated by other procedural and evidentiary rules that make it even easier for plaintiffs to certify classes. In Quebec, for example, a decision certifying a class cannot be appealed by the defendant, but if a class is *not* certified, the would-be plaintiff *is* allowed to appeal.[30] Moreover, "recent changes to Quebec's procedural rules make it very difficult for defendants to challenge the veracity of the plaintiffs' factual submissions at the certification hearing."[31] It is actually a matter of the court's discretion whether a defendant has a right to present evidence at the certification stage at all.[32]

These permissive standards have led Canadian courts to certify classes that would almost certainly have been rejected in the United States. One court in Quebec, for example, authorized a class proceeding for 2,400 members of a community who had suffered damage to their homes due to industrial air pollution.[33] The court acknowledged vast differences among the class members: only some owned their homes (others were tenants), only some were long-term residents, and only some had houses constructed with material vulnerable to damage from the relevant form of pollution.[34] In fact, the court observed that it was "doubtless" that "the damages and inconvenience caused by the air pollution varied from house to house, depending on the nature and value of the house and the distance from the port installations and its location with reference to the prevailing wind."[35] But despite these differences among class members, the court certified the class anyway, concluding that the law did not require that all—or even the majority—of the questions in the case be related.[36] It was enough, the court held, "if the claims of the members raise *some* questions of law or fact that are sufficiently similar or sufficiently related to justify a class action."[37]

This kind of liberal authorization of class actions has been widespread. In *Boulanger v. Johnson & Johnson*, in the aftermath of alleged cardiac reactions connected to Propulsid (a drug designed to treat stomach disorders), a Canadian court in Ontario certified a class of all persons in Canada outside of Quebec who ingested the drug.[38] The court's opinion acknowledged that "not all class members stand to recover damages at the same level," and that "some class members may not be able to demonstrate that they have sustained injuries and losses" at all.[39] Nevertheless, the court continued, Canadian class action law imposes only a "low bar, and the fact that many individual issues may remain after resolution of the common issue... is not a bar to certification."[40]

To be sure, Canadian courts do give limited consideration to whether a proposed class action turns on common issues as opposed to individual ones. They address this factor, however, only when evaluating the catch-all test in their certification process—whether the class action is the "preferable procedure" to resolve the dispute.[41] Statutes in some provinces explicitly list predominance as one of the "preferable procedure" factors;[42] courts in other jurisdictions have considered it without explicit legislative direction to do so. Consideration of predominance in this way, however, has generally been treated as a matter of discretion that lies solely with the judge: if individual issues are present in a lawsuit, these do not bar the certification of the class.[43] Thus, to the extent courts apply a predominance restriction in Canada at all, it is a diluted version of its U.S. counterpart.

The upshot of these structural features of Canada's class action law is that a large proportion of proposed classes in Canada are certified. There is a general scholarly consensus that Canadian class certification is more liberal than that in the United States.[44] For example, in the medical products liability field, there was "no successful opposition to certification... in Ontario [from] 1994" to at least 2003.[45]

III. Why it Matters: The Consequences of A Liberal Certification Rule

Canada's low bar for class certification, driven by the conspicuous absence of a formal predominance requirement, has important practical consequences that should alarm European policymakers contemplating Canada as a model for class action form. *Any* class action system has the potential to increase the pressure on defendants to settle cases, simply because a verdict in a class action will result in liability to numerous class members, rather than to a single plaintiff. This pressure is increased by liberal certification rules. That is true in part for the straightforward reason that liberal rules make it easier for class plaintiffs to subject defendants to this class-based pressure to settle. But it is also true because liberal certification rules can lead to a series of class procedures that are tilted severely in plaintiffs' favor. For example, absent a strict requirement that the named plaintiffs prove the elements of each individual class member's claim, courts may authorize "perfect plaintiff" classes, "common" issues trials, and "trial by statistics." These procedures prevent defendants from defending themselves fully. The use of such procedures—which skew the class action process in favor of plaintiffs—forces defendants to choose between two unpalatable options: either face an unfair trial they are likely to lose, or settle plaintiffs' claims regardless of their merit.

A. The Prejudicial Procedures Resulting From Liberal Certification Rules

The loose class certification requirements currently applied in Canada make it possible to certify a broad class that includes plaintiffs who are more different than they are alike, making a fair trial of all class members' claims virtually impossible. Because such classes involve dissimilar plaintiffs alleging highly individualized claims, it is not possible for a court to determine whether each class member can establish the elements of his or her claims based on common evidence alone. Thus, courts must instead determine a class action defendant's liability to the class "generally." This often occurs in one of three different ways: (1) the determination of liability based on a "perfect plaintiff's" claims; (2) the certification of "issues" classes; or (3) the use of "trial by statistics." These generalized methods of adjudicating class claims essentially stack the deck in plaintiffs' favor, allowing plaintiffs to prevail without ever having to demonstrate that individual class members can establish the elements of their claims through common proof. As a result, defendants are often forced to settle claims regardless of merit, due to the risk inherent in trying claims on a class-wide basis where plaintiffs are not required to prove key elements of their claims.

1. Defending Against a Fictional, Composite Plaintiff

In a class action, representative plaintiffs are appointed to litigate their claims on behalf of the other members of the proposed class: if the representative plaintiff's claims succeed, so do the claims of every class member. As a result, class actions function correctly only if the named plaintiff is so similarly situated to the class members that a finding in the representative's case applies fairly and equally to every other class member's case as well.

If, however, a court certifies a class in which the plaintiffs' claims vary, plaintiffs can "stitch together" the strongest portions of their various cases—for example, highlighting evidence that the defendant made serious misrepresentations to one plaintiff, that a second plaintiff relied on defendants' statements, and that a third plaintiff suffered damages. An action based on such "a fictional composite" is often "much stronger than any plaintiff's individual actions would be."[46] Some courts have referred to this as the "perfect plaintiff" problem,[47] because plaintiffs enjoy "the practical advantage of being able to litigate not on behalf of themselves but on behalf of a 'perfect plaintiff' pieced together for litigation."[48]

Allowing class members to prevail based on the fictional case of a "perfect plaintiff" is problematic for several reasons. First, it raises serious fairness concerns, as it allows the fact-finder to ignore the weaknesses in one plaintiff's case in light of the strength of another plaintiff's case. As such, it forces the defendant to defend against a much stronger collection of claims than it would have to face absent class certification. But, perhaps more importantly, it often leads to class recovery in cases where the individual members of the class would be unable to establish the elements of their claims standing alone. If two individual plaintiffs were alleging separate tort claims against a single defendant, each would have to prove that the defendant owed them a duty, breached that duty, and as a result caused them damages. If each individual plaintiff could not prove *both* elements—if, for example, one plaintiff could not prove that the defendant owed her a duty and the other plaintiff could not prove that she suffered injury as a result of defendant's conduct—neither could prevail. However, in a "perfect plaintiff" scenario, those two plaintiffs can effectively join forces, with one plaintiff providing evidence of a duty and the other proving causation and damages. This kind of patently unfair result is exactly what occurs in class actions involving plaintiffs alleging individualized claims, as class members with weak claims are allowed to combine to present an airtight case against the defendant.

2. "Common Issues" Trials

Some Canadian courts have attempted to avoid the problems inherent in class trials of dissimilar claims by certifying classes for the limited purpose of addressing one or more "common issues" that the court has determined apply to each class member's claims, and then leaving the individualized aspects of the claims to be resolved in subsequent individual trials.

For example, in *Heward v. Eli Lilly & Co.*, the court certified an issues class brought by users of the anti-psychotic drug, Zyprexa, who claimed personal injury as a result of using the drug.[49] In that case, the court determined that seven "common" questions would be answered as to the entire class based on a single trial, including whether "defendants knowingly, recklessly, or negligently breach[ed] a duty to warn or materially misrepresent[ed] any of the risks of harm from Zyprexa" and whether "class members [are] entitled to special damages for medical costs incurred in the screening, diagnosis and treatment of diseases related to Zyprexa."[50] Similarly, in *Boulanger v. Johnson & Johnson Corp.*, the court certified a class of users of the acid reflux drug Propulsid who claimed that the drug caused cardiac injury.[51] There, the court determined that issues such as "whether [Propulsid causes] adverse cardiac events" and whether defendants engaged in the "negligent design, manufacture, marketing, and sale of [Propulsid]" were common to the entire class and therefore could be resolved in a single proceeding.[52]

The idea behind these types of issues classes is to streamline the resolution of common questions applicable to all plaintiffs and to try individual issues separately. In reality, however, it is impossible to resolve these issues in a vacuum, divorced from the facts of each plaintiff's case. For example, in the product liability suits described above, the Canadian courts certified questions such as whether the product at issue *could* cause the injury plaintiffs allege or whether the defendant *generally* failed to warn consumers about the product's risks. However, answering these questions in the abstract, without reference to the facts of any particular plaintiff's case, does nothing to advance the class members' cases. For instance, an issues trial on "general causation" asks the court (or jury) to answer the highly prejudicial question of whether plaintiffs have shown that the product is *capable* of causing harm. This is an undeniably easy standard for plaintiffs to meet since they do not have to prove that the product *actually did* cause injury to class members—a showing that would require evidence of each class member's medical history, susceptibility to injury, and risk of injury from other sources. Instead, plaintiffs must merely prove that such an injury is generally *possible*.

The same is true with regard to an issues trial on the question of whether a class action defendant generally "failed to warn" consumers about the risks associated with a product. In such a trial, plaintiffs would only be required to prove that the defendant's statements would *likely* have misled consumers, not that the statement *did* mislead each individual consumer—an inquiry which would require highly individualized evidence related to, inter alia, each consumer's exposure to the defendant's statements, the changing nature of the defendant's own knowledge of the risks of its product, and the availability of risk information from other sources.

The supposed purpose of an issues trial is to resolve common issues in one trial so that a second trial on individualized issues can be conducted quickly for each plaintiff. However, because "general" findings as to causation or failure to warn made in an issues trial do not relate to the facts of any individual class member's claim, they offer no real efficiency benefit. Instead, issues trials simply stack the deck against defendants because: (1) the "common issues" are so generalized that they are almost certain to result in a plaintiff verdict; and (2) the second phase of the trial—*i.e.*, the actual trials of individual plaintiffs' claims—is inevitably infected by the verdict from the first phase.

Plaintiffs in an issues trial are not required to present any case-specific evidence; they merely need to "establish" general liability—whether on the issue of general causation, failure to warn, or some other allegedly "common" issue. Thus, it is almost impossible for a defendant to prevail on the issues phase of such cases. And even though each class member would have to prove in a subsequent trial that the defendant actually caused his or her injury specifically, that second trial would inevitably be affected by the initial finding that the defendant's actions could have caused harm to the class. In essence, the defendant is forced to defend the individualized second-phase trials with a red mark next to its name.

Moreover, even though it does not result in even a cent of actual liability, an issues trial verdict against a class action defendant can have serious public relations and investment consequences. Even before a single class member has proved that he or she was actually injured as a result of the defendant's conduct, or legally entitled to any relief, plaintiffs are nevertheless able to claim publicly that they have "established" general liability. The resulting media coverage can have a serious negative impact on a class action defendant's business and financial well-being. As a result, defendants are effectively subject to economic punishment despite the fact that no plaintiff has ever proved his or her claim.

3. Class Trial by "Statistics"

Permissive class certification requirements that allow certification of dissimilar, highly individualized claims can also lead to the use of generalized "statistics" as proof of plaintiffs' claims. Simply put, if class members' claims vary, it may not be possible to resolve all of their claims based on the facts of a representative plaintiff's case. As a result, class action plaintiffs often attempt to prove highly individualized elements of their claims—such as causation and damages—on a class-wide basis using over-generalized statistical evidence that fails to address the merits of any one class member's claims.

The problems inherent in such a "trial by statistics" are obvious. Like a "common" issues trial, a trial based on statistical evidence and expert testimony does not establish the defendant's liability as to individual class members. Instead, statistical evidence merely indicates that it is *probable* that the defendant's actions caused harm to the group as a whole. Thus, assessing liability based solely on this type of generalized proof essentially allows individual class members to recover without ever having to prove the basic elements of their claims. As a result, defendants are forced to compensate an entire group of plaintiffs without any one of those plaintiffs having to prove that he or she was actually injured or that his or her injury occurred as a result of the defendant's conduct.

In addition, assessing a defendant's liability as to an entire class based on statistical evidence robs the class action defendant of the ability to adequately defend itself against plaintiffs' claims. For example, in a product liability failure-to-warn case, the defendant will likely have a variety of defenses against an individual plaintiff's claims based on the individual circumstances of that plaintiff's exposure to warnings, history of heeding warnings, and knowledge of the alleged risks from other sources. Some plaintiffs will have a history of ignoring warnings and using dangerous products. Other plaintiffs may have heard from another source about the risks that the defendant allegedly withheld. Thus, those plaintiffs will have a much weaker failure-to-warn case than a plaintiff who had no knowledge of the alleged risks and is adamant that she would never have used the product if she had been aware of the risk. A trial by statistics makes no allowance for such distinctions—even though common sense suggests that a more or less random sample of consumers, with highly varied habits and histories, will have very different claims. As a result, even though some plaintiffs' cases may be fatally flawed (including plaintiffs who cannot establish all of the elements of their cause of action or whose claims are susceptible to individualized defenses, such as the statute of limitations), those flaws will never be uncovered during a "trial by statistics."

In short, a trial by statistics is more akin to a theoretical "trial by average" where it is determined that the defendant's actions would probably have harmed the "average consumer," but never established that the defendant actually harmed any real consumer. Like a trial of putatively "common" issues, it is an example of "rough justice" at its most stark: an acceptance of prejudice to defendants and unproven recovery to plaintiffs, simply to achieve the expediency of disposing of multiple claims at once.

B. The Pressure to Settle

The risks to a defendant from "perfect plaintiff" trials, "issues" trials and "trials by statistics" all have the same effect: they exacerbate the already-existing pressure on defendants to settle class actions regardless of the merits of the plaintiffs' claims. In *any* class action system, there is a risk that opportunistic plaintiffs, with an eye towards settlement, can seek to identify potential classes, knowing that once the potential liability for a claim is multiplied by the number of class members, the defendant will face tremendous pressure to settle, regardless of the merits of the underlying claims.[53] As noted earlier, the mere fact that a defendant faces a class, rather than an individual plaintiff, dramatically increase the downside risk of proceeding to trial. The bulk of this pressure to settle, as one American court recently explained, is felt at the moment a class action is certified: "class certification may be the backbreaking decision that places 'insurmountable pressure' on a defendant to settle, even where the defendant has a good chance of succeeding on the merits."[54] Thus, in the U.S., the class certification proceeding is the "major battleground" of the litigation.[55] Once a class action has been certified, in the eyes of many defendants, the game is over.

The weight of settlement pressure is magnified where class certification standards are relaxed. In Canada, for example, it is relatively rare for class action lawsuits to proceed all the way to a trial. By one commentator's count, only two class action cases in Canada were tried between 1993 and 2001, and, although some class actions were resolved by summary judgment, most concluded in settlement.[56] This tendency toward settlement is likely linked not simply to the permissiveness of Canadian class certification requirements but also (and more specifically) to Canadian courts' general acceptance of "issues classes." Put simply, once a class is certified, class action defendants are faced with a lose-lose proposition: either spend substantial sums litigating a case that begins with a series of "common" issues trials they are destined to lose, or agree to pay out substantial settlement costs to class members whose claims have never been tested and are likely without merit.

For this reason, the "relatively undemanding" criteria for class certification in Canada are a cause for deep concern among potential class action defendants in that country—and should likewise concern policymakers tempted to use Canada as a model for class action legislation.[57] If classes are even easier to certify in Canada than in the United States, plaintiffs' lawyers will have an even greater incentive to file unmeritorious claims, simply to exploit their potential settlement value. This can, in turn, cause a variety of disruptive effects common to American class action litigation, such as causing companies to minimize or avoid investment in particular jurisdictions, or causing them to divert resources from innovation and other productive activities to the defense of class action litigation.

IV. Avoiding Unintended Consequences: Learning Lessons from American Mistakes

Even if Canada has, to date, avoided some of the excesses (particularly the volume of litigation) that characterize the U.S. class action regime, the structure of the Canadian system, with its lack of a predominance requirement and permissive class certification requirements, render it even *more* inherently prone to abuse than the U.S. system. Indeed, as discussed, the kinds of inherently unfair procedures that generally accompany liberal certification rules (such as "common" issue trials) are already visible in Canada. The fact that Canada has avoided all forms of U.S.-style abuse to date does not mean it will remain immune from such problems.

In this regard, it is instructive to recall that the kind of class action abuse that permeates the U.S. system did not appear overnight. In 1966, when U.S. rule-makers amended Federal Rule of Civil Procedure 23 to alter certain features of U.S. class action law, they "apparently believed that they were simply making rule 23 a more effective procedural tool."[58] According to its accompanying note, the 1966 amendment, which reformulated the text of the rule, was intended:

> (1) to redefine the cases that could proceed under rule 23, by adopting more functional definitions of class actions, (2) to clarify the effect of a class action judgment on members of the class, (3) to codify some of the better class action practices that federal judges had developed, (4) to provide district court judges more guidance regarding their procedural powers and responsibilities, and (5) to deal explicitly with the notice that should be provided to absent class members.[59]

This seemingly benign adjustment to the law's structure was not driven by "revolutionary notions," but rather by far more modest and benign intentions: to create a more efficient legal "mechanism for securing private remedies."[60] As U.S. civil procedure scholar Arthur Miller has explained, "the class action onslaught caught everyone, including the draftsmen [of Rule 23] by surprise."[61]

The development of rampant class action abuse resulted in large part from the unforeseen interaction between these apparently modest changes in class action law and other developments in U.S. law that were still under way when Rule 23 was revised. Specifically, in the second half of the twentieth century, Congress and the American federal courts began recognizing new substantive rights in the areas of (*inter alia*) civil rights, antitrust, securities litigation, and consumer protection laws.[62] American product liability law, for instance, has undergone massive changes since the 1960s, evolving from "a rule that held that manufacturers were not liable unless they had been negligent in the manufacture of their products to a (mostly judge-imposed) rule that manufacturers were liable (even if there had been no negligence) if products left the factory in an 'unreasonably dangerous' condition."[63] Though these changes were not initially enacted with the class action device in mind, entrepreneurial American class action lawyers were quick to see the advantageous marriage of the class action device and these broad new substantive rights.

Meanwhile, as the potential uses for the American class action device began to grow, there were no safeguards in place to prevent abusive litigation. Lawyers who bring class actions in the United States provide their services in exchange for an (often large) percentage of the ultimate settlement or award. This contingency-fee arrangement, in combination with the lack of a "loser pays" rule in the U.S., makes it relatively easy and inexpensive for U.S. plaintiffs to institute class actions. Moreover, plaintiffs' lawyers have ample incentives to bring such suits almost regardless of their merits, due to the potentially enormous awards (which can include a large punitive damage component) and the fact that the percentage devoted to the lawyer's fee is typically quite high.[64] Further adding to the potential for abuse is the fact that liberal American discovery rules impose extraordinarily high costs on defendants from the start of litigation, a fact that plaintiffs can leverage into an early settlement.[65]

In short, although Rule 23 was crafted with modest intentions, unforeseen interactions with evolving substantive law, together with the peculiarities of U.S. procedural rules, led to the widespread abuse that now characterizes the American class action regime.

CONCLUSION

The fact that Canada has not yet been overrun with the same volume of class actions as the United States is no reason to recommend it as a model. Indeed, any limitation on the zeal of class action lawyers in Canada comes not from its liberal class action rules but from other features of Canadian law. Canadian plaintiffs' lawyers are not allowed to take cases on pure contingency arrangements. Instead, courts determine how much compensation a Canadian class action lawyer is entitled to collect by considering the number of hours worked and the size of the ultimate award collected.[66] Further, these fees are generally not as large as the ones collected in the U.S., due to caps on pain-and-suffering damages and restrictions on punitive damages.[67] Moreover, under Canadian law, when the defendant prevails in a class action, the class representative himself is responsible for the defendant's legal fees. This, too, serves as a disincentive to mounting a frivolous suit.[68]

But European policymakers should *not* conclude from this Canadian experience that they can simply import the Canadian system into Europe with impunity, thereby avoiding some of the abuses that have plagued the United States. Canada's class action regime is less than thirty years old and is still developing. As demonstrated above, U.S. class action abuse problems did not appear overnight, and were not anticipated by the framers of the U.S. class action rules. It would be naive to assume that the same harms can be easily and casually avoided by others now. The fact that Canada has thus far managed to avoid some of the excesses of the U.S. system does not mean the structure of Canadian class action law is the secret of its success. To the contrary: as discussed above, the structure of Canadian class action law, and in particular its lack of a formal predominance requirement, has already led to inherently unfair class procedures, and leaves the Canadian regime even *more* inherently vulnerable to manipulation than the law in the United States.

Moreover, as class actions continue to grow in Canada, political pressure will likely mount to shed some of the safeguards (such as contingency fee restrictions and "loser pays" rules) that until now appear to have protected the country from pervasive class action abuse. Several Canadian provinces, for example, have already adopted strategies to alleviate the harsh consequences of the traditional loser pays rule for the named plaintiff: Quebec has made class representatives liable for only nominal costs; British Columbia excuses class representatives from paying defendants' legal costs unless the failed suit was "frivolous or vexatious;" and Ontario legislation provides that the named plaintiff has to pay the defendant's costs only if the action was a "test case, raised a novel point of law or involved a matter of public interest."[69]

The character of Canadian class action practice is also under pressure from the recent increase in cross-border litigation. As other commentators have noted, the U.S. plaintiffs' bar is making aggressive moves into Canada, and has increasingly engaged in concerted cross-border strategies, such as attempting to use U.S. discovery and expert testimony in Canadian courts, or using parallel litigation in the United States and Canada to increase the overall pressure on defendants to engage in a settlement.[70] These developments both reduce the costs of litigating class actions in Canada, and sweep Canada into a multi-national dynamic that includes, and is driven by, class action litigation in the United States.

The ultimate effect of these developments is still uncertain, but there is good reason to believe that class action abuses in Canada will continue to mount. Accordingly, European policymakers should look past the simple, unreflective empirical conclusion that Canada has fewer class actions than the United States and instead look more closely at how exactly Canadian class action law is structured. And on such examination, it becomes clear that Canada provides an extremely liberal certification regime, with a firmly embedded structural bias in favor of permissive certification. This bias, in turn, creates both inherently biased class procedures, and a powerful incentive toward bringing class actions—even class actions based on frivolous claims.

Endnotes

1 Christopher Hodges, *Multi-Party Actions: A European Approach,* 11 Duke J. Comp. & Int'l L. 321, 346-47 (2001). *See also* Edward Sherman, *Group Litigation Under Foreign Legal Systems: Variations and Alternatives to American Class Actions,* 52 DePaul L. Rev. 401, 431 (2002) ("Canadian class action settlements have not drawn the same level of criticism as have settlements in the United States."); *id.* at 402 (noting that Canada has not "experienced the same mushrooming of class action practice as in the United States").

2 Sherman, *supra* note 2, at 403 ("[M]ost other countries view American class actions as a Pandora's box that they want to avoid opening.").

3 *See* Quebec Civ. Code, R.S.Q. 1978, ch. C-25, §§ 999-1051; *see also* Edward Sherman, *American Class Actions: Significant Features and Developing Alternatives in Foreign Legal Systems,* 215 F.R.D. 130, 153-54 (2003).

4 W.A. Bogart, *Questioning Litigation's Role -- Courts and Class Actions in Canada,* 62 Ind. L. J. 665, 692 (1987).

5 *Id.*

6 Class Proceedings Act, S.O., ch. 6 (1992) (Ont.); Class Proceedings Act, R.S.B.C. 1996, ch. 50 (1995) (B.C.).

7 Michael McGowan, *Certification of Class Actions in Ontario: A Comparison of Rule 23 of the U.S. Federal Rules of Civil Procedure,* 16 Carswell Practice Cases 172 (1993).

8 Paul J. Martin et al., *International Legal Development in Review: 2003 Comparative Law – Canadian Law*, 38 Int'l Law. 553, 553 (2004).

9 This does not mean, however, that courts will inquire into the merits. Thus, the fact that a cause of action may be barred by a particular defense (a limitations period for example), does not bar certification of the class action. *See* McGowan, *supra* note 7, at 172.

10 *See* R.S.B.C. 1996, c. 50, s.4 (B.C.); S.O. 1992, c.6, s.5 (Ont.).

11 S. Gordon McKee, *Why the Development of Mass Torts in Canada is Important to Corporate America,* 71 Def. Couns. J. 32, 34 & n.2 (2004) ("Often there will be a national class certified in Ontario, comprised of either all Canadian residents or Canadian residents other than those in British Columbia and Quebec, with companion actions in those provinces. Trial level courts in Ontario and B.C. and an appellate court in B.C. have been willing to certify national classes.").

12 Steven Penney, *Mass Torts, Mass Culture: Canadian Mass Tort Law and Hollywood Narrative Film*, 30 Queen's L.J. 205, 217 (2004); Alex Kotkas & Casey Smith, *Class Actions in Canada: Another Western Province (Alberta) Enacts Legislation*, 28 No.1 Class Action Reports 4 (Feb. 2007) (noting that with Alberta's 2004 legislation, there are now six Canadian provinces who have enacted comprehensive class action legislation).

13 [2001] 2 S.C.R. 534 ¶ 34.

14 *Id.* ¶¶ 38-41.

15 *Id.* ¶ 42.

16 Penney, *supra* note 12, at 218.

17 Garry D. Watson, *Class Actions: The Canadian Experience,* 11 Duke J. Comp. & Int'l L. 269, 272 (2001) ("in certain respects the Canadian legislation is more liberal in facilitating class actions than its American counterpart"). *See also* Andrew Borrell, *The Evolving Evidentiary Standard for Certification in Canada,* 26 No. 6 Class Action Reports 3 (2005); Caroline Davidson, *Tort au Canadien: A Proposal for Canadian Tort Legislation on Gross Violations of International Human Rights and Humanitarian Law,* 38 Vand. J. Transnat'l L. 1403, 1442 (2005) ("[i]n many respects, particularly certification, the

Ontario regime is more liberal.") (quoting Michael McGowan, *Certification of Class Actions in Ontario: A Comparison of Rule 23 of the U.S. Federal Rules of Civil Procedure,* 16 Carswell Practice Cases 172 (1993)).

18 *See* Fed. R. Civ. P. 23(b)(3).

19 Edward Sherman, *American Class Actions: Significant Features and Developing Alternatives in Foreign Legal Systems,* 215 F.R.D. 130, 140 (2003).

20 Gary Will & Paul Miller, Association of Trial Lawyers of America, *Participation of United States Counsel in Canadian Class Actions and Mass Torts,* 1 Ann. 2006 ATLA-CLE 483 (2006).

21 Amchem Prods., Inc. v. Windsor, 521 U.S. 591, 623 (1997).

22 Sherman, *supra* note 19, at 140-41.

23 7AA Charles Alan Wright et. al, Federal Practice and Procedure §1777 (1998).

24 *Id.* §1751 (emphasis added).

25 S. Gordon McKee & Martha Cook, *Class Actions in Canada: 2005 State of the Union – Growing, Expanding, Developing,* 73 Def. Couns. J. 31, 32 (2006).

26 *See, e.g.,* Class Actions Act, S.S. 2001, c.-12.01, §6 (Sask.); Class Proceedings Act, S.B.C., ch. 21 (1995) (B.C.); Class Proceedings Act, S.A. 2003, c. C-16.5, §5 (Alta.); Class Actions Act, S.N. 2001, c. C-18.1, §5 (Nfld.); Class Proceedings Act, S.M. 2002, c. 14, § 4 (Man.).

27 W. Canadian Shopping Ctrs., Inc. v. Dutton, [2001] 2 S.C.R. 534, ¶ 39.

28 Gordon McKee & Robin Linley, *The Evolving Landscape for Pharmaceutical Product Litigation in Canada,* 73 Def. Couns. J. 242, 243 (2006) ("In Quebec, the threshold for certification is even lower than in the other Canadian provinces.").

29 *W. Canadian Shopping Ctrs.* [2001] 2 S.C.R. 534, ¶ 37 (citing Quebec Code of Civil Procedure, art. 1003).

30 Jean Saint-Onge & Glenn M. Zakaib, *Cross-Border Class Actions—A Canadian Perspective,* 744 Practicing Law Institute, Litigation and Administrative Practice Course Handbook Series 87, 93 (2006); John C. Kleefeld, *Class Actions as Alternative Dispute Resolutions,* 39 Osgoode Hall L.J. 817, 827-28 (2001) ("in Quebec, where a refusal to certify a proceeding as a class action can be appealed, while a decision to certify cannot").

31 McKee & Cook, *supra* note 26, at 32.

32 Saint-Onge & Zakaib, *supra* note 30, at 93.

33 Comité d'environement de la Baie, Inc. v. Société d'électrolyse et de chimie Alcan Ltée, [1990] 29 Q.A.C. 251.

34 *Id.* ¶ 21.

35 *Id.*

36 *Id.* ¶ 22 (emphasis in original).

37 *Id.* (emphasis in original).

38 Boulanger v. Johnson & Johnson Corp., [2007] 40 C.P.C. (6th) 170 (Ont. Sup. Ct.).

39 *Id.* ¶ 21.

40 *Id.* ¶ 28.

41 *See, e.g.,* Hollick v. Metro. Toronto, [2001] 13 C.P.C. (5th) 1, ¶ 30 (Sup. Ct.) ("The question of preferability, then, must take into account the importance of the common issues in relation to the claims as a whole... I cannot conclude... that the drafters intended the preferability analysis to take place in a vacuum. There must be a consideration of the common issues in context."); Knight v. Imperial Tobacco Canada Ltd., [2006] 54 B.C.L.R. (4th) 204, ¶ 12 (Ct. App).

42 The Alberta legislation provides such an example. *See* Class Proceedings Act, S.A. 2003, c. C-16.5 (Alta).

43 Campbell v. Flexwatt Corp., [1997] 44 B.C.L.R. (3d) 343 ¶ 61 (Ct. App.) ("nowhere does the Act mandate that if an individual issue should predominate, an action must not be certified. Instead, the Act sets out a variety of factors to be considered. The existence of an individual issue is not necessarily determinative."); Cloud v. Canada, [2004] 73 O.R. (3d) 401 (Ct. App.).

44 *See* Watson, *supra* note 17, at 4 ("in certain respects the Canadian legislation is more liberal in facilitating class actions than its American counterpart."); Davidson, *supra* note 17, at 1442 ("[i]n many respects, particularly certification, the Ontario regime is more liberal."); McKee & Cook, *supra* note 25, at 32 ("The threshold for class certification in Canadian provinces is generally considered to be lower than in the United States."); Charles M. Wright & Matthew D. Baer, *Price Fixing Class Actions: A Canadian Perspective*, 16 Loy. Consumer. L. Rev. 463, 467 (2004) ("Canadian courts may be somewhat more inclined to certify classes").

45 McKee & Cook, *supra* note 25, at 32.

46 *See* Broussard v. Meineke Disc. Muffler Shops, Inc., 155 F.3d 331, 344-45 (4th Cir. 1998).

47 *Id.*

48 *Id.*

49 [2007] 39 C.P.C. (6th) 153 (Sup. Ct.)

50 *Id.* ¶ 77

51 [2007] 40 C.P.C. (6th) 170 (Ont. Sup. Ct.)

52 *Id.* ¶¶ 28, 34-36.

53 Martin Redish & Adriana Kastaneck, *Settlement Class Actions, The Case-Or-Controversy Requirement, and the Nature of the Adjudicatory Process,* 73 U. Chi. L. Rev. 545, 557 (2006) (noting that the use of the class action device has enabled plaintiffs' lawyers to "extract a settlement far in excess of the individual claims' actual worth") (quotation omitted).

54 Regents of the Univ. of Calif. v. Credit Suisse First Boston (USA), Inc., 482 F.3d 372, 379 (5th Cir. 2007).

55 *Watson*, *supra* note 18, at 279; John C. Kleefeld, *Class Actions as Alternative Dispute Resolutions,* 39 Osgoode Hall L.J. 817, 827-28 (2001) ("In the class action context, [ ] certification is everything.").

56 *See* Watson, *supra* note 17, at 278.

57 Watson, *supra* note 17, at 273 ("The Canadian criteria for certification of a proceeding as a class action are relatively undemanding."); Wright, *supra* note 44, at 467 ("Canadian courts are somewhat more inclined to certify classes").

58 Arthur Miller, *Of Frankenstein Monsters and Shining Knights: Myth, Reality, and the "Class Action Problem,"* 92 Harv. L. Rev. 664, 670 (1979).

59 *Id.* at 669 (citing the1966 amendments to the Federal Rules and their accompanying literature which can be found at 39 F.R.D. 69 (1966)).

60 *Id.* ("The Advisory Committee's objectives in rewriting the [federal class action rule] were rather clear. It had few, if any revolutionary notions about its work product. Although it was expected that the revision would operate to assist small claimants, the draftsmen conceived the procedure's primary function to be providing a mechanism for securing private remedies, rather than deterring public wrongs or enforcing broad social policies.").

61 *Id.* at 670.

62 *Id.* at 672-74 (discussing, among others, the Civil Rights Act of 1964, the availability of court-awarded attorneys fees in antitrust cases, the judicial expansion of the concept of "security," the passage of the Fair Credit Reporting Act).

63 Stephen B. Presser, *How Should The Law of Products Liability Be Harmonized? What Americans Can Learn From Europeans,* 2 Global Liability Issues (1-2) Feb. 2002.

64 *Id.* at 4 ("Counsel's fees are then taken from the settlement or judgment, and will generally be a substantial portion of the recovery, as much as 30-40%, or more").

65 Stephen N. Subrin, *Fishing Expeditions Allowed: The Historical Background of the 1938 Federal Discovery Rules,* 39 B.C. L. Rev. 691, 739-740 (1998); Richard M. Dunn & Raquel M. Gonzalez, *The Thing About Non-U.S. Discovery for U.S. Litigation: It's Expensive and Complex*, 67 Def. Couns. J.

342, 342, 346-47 (2000) ("American lawyers are accustomed to the broad, sometimes intrusive, discovery available to litigants in the United States under federal and state rules of procedure.").

66 *See* Watson, *supra* note 17, at 277.

67 *See* Sherman, *supra* note 19, at 156; S. Gordon McKee, *Why Development of Mass Torts in Canada is Important to Corporate America*, 71 Def. Couns. J. 32, 36 (2004).

68 *See* Watson, *supra* note 17, at 274.

69 *Id.*

70 McKee & Cook, *supra* note 25, at 42-43 (2006) ("Given the potential for alleged wrongs to occur across international boundaries, plaintiffs can be expected to rely more and more on cross-border discovery to gather the evidence necessary to support their motions for certification in Canada.").

An example of this emerging cross-border coordination can be seen in the Vitapharm litigation. *See* Vitapharm v. F. Hoffman-LaRoche, Ltd, [2001] 11 C.P.R. (4th) 230 (Ont. Super. Ct.) (affirmed, [2002] 18 C.P.R. (4th) 267 (Ont. Div. Ct). There, vitamin manufacturers were accused of fixing prices in both Canada and the United States. Lawsuits were pending in both countries, and the defendant sought to prevent the Canadian plaintiffs from gaining access to U.S. discovery. The court denied the defendant's request, noting that "as a result of the inexorable globalization and expanding international free trade and open markets, there will be an ever-increasing inter-jurisdictional presence of corporate enterprises. This is seen particularly in respect of American and Canadian business activity, given the extent of cross-border trade." *Id*. ¶ 27. Because access to the U.S. discovery could save time and allow plaintiffs to "more easily determine and discard what is not relevant for the purpose of the Canadian proceeding," *id*. ¶ 48, and because allowing the access did not cause unfairness to the defendants in the Canadian courts, *id*. ¶ 50, the Court permitted the sharing of information and denied the defendant's motion.

Professional Responsibility & Legal Education

Electing State Judges: Unpleasant, But Not Unconstitutional

*By Ed Haden & Conrad Anderson, IV**

"States, we don't like your method of judicial selection. In fact, it is downright terrible. But we strike down only unconstitutional laws, not stupid ones. Change it if you want, but it's not our problem." This seems to be the message of the U.S. Supreme Court in its recent decision, *New York State Board of Elections v. Lopez Torres*.[1]

Election of New York's Judiciary

Under the New York State Constitution, justices on the supreme court, that state's trial court of general jurisdiction, are elected to fourteen-year terms in one of the twelve judicial districts. The constitution provides that the candidate selected by each party automatically appear on the ballot with the party's endorsement. While the method for the parties' selection has seen changes, for the past eighty years candidates have been selected by a convention composed of delegates elected by party members.

Under this "convention" selection method, each party holds a delegate primary in which party members elect delegates from each of the state's 150 assembly districts. These delegates then attend the nominating convention in one of the twelve judicial districts where they nominate the party's supreme court candidate to run at large in that district. The selected nominees automatically appear on the general election ballot and may be joined by independent candidates and candidates of smaller political organizations who gain access by obtaining a set number or percentage of signatures.

This system has been widely criticized as leading to corruption and turning New York's trial courts into puppet shows for party bosses. Apparently, party members often do not know enough or simply care little about the delegates who attend the nominating convention. The party bosses organize a slate of delegates in each district who either run unopposed or against poorly funded opponents. A candidate seeking the party's nomination without the support of party leadership would have to organize his or her own slate of delegates, get them elected at the delegate primary, and do this in numerous assembly districts to have meaningful support at the nominating convention. Thus, the candidates favored by party leadership inevitably win the delegate primaries. Once elected, these delegates attend the convention and follow the instructions of the party boss in choosing the district's nominees. Under this arrangement, both the delegates and the nominees tend to be allies of the party bosses, and upsetting the boss can quickly end judicial aspirations, as the respondent before the U.S. Supreme Court found out.

** Ed Haden is an attorney with Balch & Bingham LLP in Birmingham, Alabama, where he is a member of the firm's litigation section and leads its Appellate Practice Focus Group. Conrad Anderson IV also practices in the litigation section of Balch & Bingham LLP in Birmingham, Alabama. Mr. Anderson's practice focuses on intellectual property and business litigation.*

The System Is Unfair If I Can't Win

In 1992, Margarita Lopez Torres, a nominee of the Democratic Party, was elected to one of New York's county courts, which are of more limited jurisdiction than the supreme court. She unsuccessfully sought the party's supreme court nomination in 1997, 2002, and 2003. In 2004, she and a group of voters and other failed candidates filed suit against the New York Board of Elections, the governmental agency in charge of the state's election laws. She claimed that party leaders unfairly used their influence to block her attempts to gain the supreme court nomination because she refused to make patronage hires. She alleged that the state's election method burdened the rights of candidates who were not favored by party leadership because it made it virtually impossible for them to get elected (even though she could still run as an independent, she claimed that there was no realistic opportunity of winning the election as an independent). She further asserted that the law deprived voters and candidates of their right to ballot access and the right to associate in choosing the party's candidates, all in violation of the First Amendment.

The district court enjoined the state's use of the convention system. The Second Circuit Court of Appeals affirmed, holding that Torres had a First Amendment right to a "realistic opportunity to participate in [a political party's] nominating process, and to do so free from burdens that are both severe and unnecessary."[2] According to the court of appeals, the reality was that the judicial districts were subject to "one-party rule" and a candidate had no legitimate shot of winning as an independent. Candidates therefore have a constitutional right to access to the party's convention in the court's opinion.

The Supreme Court Weighs In

The U.S. Supreme Court unanimously reversed the Second Circuit. Writing for the Court, Justice Scalia explained that there is no precedent suggesting a constitutional guarantee of a "fair chance" at winning a party's nomination. While the Court has acknowledged a right to vote in the party's primary, and has invalidated state laws that unduly burden that right, it has not acknowledged any right to run in the primary. Furthermore, even if the Court were to find such a right, New York's signature and deadline requirements are reasonable and survive constitutional scrutiny.

The real concern for Torres and the other respondents, the Court stated, is not that they cannot vote or run in the election, but that the state's convention process does not give them a realistic shot at winning the party's nomination. This outcome, however, is not a result of the state's election laws themselves but, rather, the simple fact that the party leadership is able to get more support for the candidates it favors than an unsupported candidate can gather. The Court explained: "Our cases invalidating ballot-access requirements have focused on the requirements themselves, and not on the manner in which political actors function under those requirements."[3] Furthermore, although the Court has permitted states to enact laws which level the playing field for "insurgent" candidates, the Court has never required it—such questions as to what is a "fair shot" are up to the legislature, not the Court.

Finally, respondents pointed out that "one-party rule" prevailed throughout several of the judicial districts and argued that the First Amendment must be used to create competition. According to the Court, "This is a novel and implausible reading of the First Amendment."[4] One-party rule, the Court explained, is typically a result of the voters' preference for the party's candidates. Although states can discourage this result, such as by removing party affiliations from the ballot, the First Amendment does not require such competition.

The Court summed up its opinion by stating, "If [New York] wishes to return to the primary system that it discarded in 1921, it is free to do so; but the First Amendment does not compel that."[5]

Cautionary Concurrences

Justice Stevens, joined by Justice Souter, concurred only to make clear that the Court's determination that a law is not unconstitutional does not mean that the Court believes the law is a good one. Quoting former Justice Thurgood Marshall, he stated "The Constitution does not prohibit legislatures from enacting stupid laws."[6]

Justice Kennedy, joined by Justice Breyer, concurred in the judgment, noting that the respondents would have a much stronger First Amendment claim if the parties' convention method were the only way to get on the ballot; but the fact is that it is not—the state also permits candidates to appear on the ballot, albeit without party endorsement, by obtaining a reasonable amount of signatures to a petition.

Justice Kennedy also saw fit to close with commentary on the propriety of having an elected judiciary. He noted that it is difficult to reconcile having elected judges with the goal of having an independent judiciary, and while states are not prohibited from electing judges, they should strive to find methods of selecting a qualified judiciary and should demonstrate an actual concern for such independence. Although conceding that the laws were not unconstitutional, Justice Kennedy proclaimed, "If New York Statutes for nominating and electing judges do not produce both the perception and the reality of a system committed to the highest ideals of the law, they ought to be changed and to be changed now."[7]

To Elect or To Appoint?

The Supreme Court's opinion seems to suggest, in part, that while allowing people to elect judges has distasteful effects, the Court is not going to deny the people that right. Moreover, while States' methods of judicial election may not be unconstitutional, they are not necessarily something to be proud of. The Court's decision raises, once again, an oft-debated question: Are elected judges a good thing? And if electing judges is a good thing, what is the best way to do that?

Those in favor of election suggest that it produces accountability—judges must make decisions that are in line with the expectations of the public at large if they want to be re-elected. Elected judges, they suggest, are not beholden to the relative few that put them behind the bench through legislative or gubernatorial appointment, but rather are responsive to the community in which they preside.

Opponents of electing judges urge that the selection of these few and powerful persons with such significant responsibility is a decision that should be left to the "experienced elite," rather than the masses. Appointment, they suggest, leads to independence and allows a judge to apply the law in a neutral manner without concern for any special interests that may have funded the judicial campaign. Of course, this ignores the reality that those who do the appointing have selected a particular judge for a reason, and while a judge may not have any political affiliation or aspiration, those who do the appointing often do. It is no secret that appointments are often a result of the perceived political views of the appointee. If the judge is periodically reviewed, he or she may be beholden, as one who is elected, only to a smaller and more powerful group of people.

Providing life tenure may prevent a judge from becoming beholden to the electorate or to those who appointed him or her. With no review and no re-election, the judge would seem to have the independence to decide the law without any outside pressure. Of course, providing such independence can leave unchecked the ambition of judges to enhance their notoriety by solving social problems, running executive agencies, and striking down legislation they personally find distasteful. While deferring to the text and plain meaning of constitutions and statutes exemplifies professional integrity and the judicial function, it does, after all, allow the political branches to exercise most of the power and be more important. Because history has shown that men and women are not angels, a check can be more useful than professional integrity in protecting the rule of law. Sometimes judges whose job is not in jeopardy have a desire to "leave their mark" on the law, such that they become "legal creators" rather than "legal interpreters."

State Judicial Selection Methods

Left with the choice, most states have chosen to use some form of an elected judiciary. The methods for selection are varied.

Twenty-two states provide for the direct election of judges, seven of them in partisan contests. In those states, a party's nominee is selected either through a primary or convention and appears on the general election ballot with the party's endorsement. In the other fifteen states that have direct elections, party affiliation is not noted on the general election ballot, even though, in most instances, the nominee was selected through a partisan primary or convention. "Nonpartisan" does not necessarily mean non-political, however. In the recent non-partisan election for a seat on the Wisconsin Supreme Court, the Associated Press reported, "Democrats and labor groups… along with 220 judges and groups representing more than 18,000 law enforcement officers" supported the losing incumbent, while "Republicans and the majority of the state's district attorneys and sheriffs" supported the challenger.[8]

Other states employ a so-called merit selection system, a non-partisan method with a goal of selecting judges based on qualifications rather than political affiliation. This system developed from what is known as "The Missouri Plan," and although there is some variation throughout the states employing it, the general procedure is the same. Whenever there is a vacancy, those interested in filling the position submit applications and interview with a non-partisan judicial commission, usually made up of attorneys, citizens, and the chief justice. The commission submits the names of a set number of qualified candidates to the governor for his or her selection. In some states, senate confirmation of the governor's choice is also required. After at least one year in office, the judge's name is placed on a judicial ballot (without party affiliation) of the next general election and voters decide whether the judge continues to serve. The judge must receive a majority of votes in favor of being retained in order to serve the next full term.

In theory, such a procedure combines some of the best aspects of both the appointed and election systems. At least some of the "experienced elite" get to choose the candidate based on qualifications; once elected, however, the judge is accountable to the community and not to those relative few who made the initial selection. In practice, however, there appears to be little accountability because the judges run unopposed in an unpublicized "campaign." The voting public is not as informed about the judge's qualifications or judicial record as they would be if there was an opponent to raise these issues. As such, judges "elected" through merit selection by and large serve as long as they desire. Moreover, studies have shown that judges selected by this method are no better qualified than those elected by the public.[9]

The lack of accountability is evidenced by the fact that in Missouri's seventy-year history, not a single supreme or appellate court judge has lost a retention election. In 2006, only 27.5% of the 192 lawyers participating in a survey recommended the retention of a circuit judge in St. Louis.[10] She was retained. And when a vacancy on the Missouri Supreme Court arose, the Appellate Judicial Commission sent the Republican Governor three nominees who did not appear to fit his desire for a judge who would not legislate from the bench.[11] This has led some of the state's leadership to push for constitutional amendments that would give some control to the legislature. Under "The Accountable Commission Plan," four seats on the nominating commission currently filled by the Missouri Bar Association and the chief justice would be filled by lawyers selected by the state house and senate. Also, rather than the commission nominating a panel of candidates for the governor to choose from, the governor would submit his selected candidate to the commission for approval. Under "The Federal Model for Appointment," the governor would nominate the candidate to be confirmed by the senate. Under a third proposal, "Effective Retention and Removal," the retention vote currently held by the public would be taken and given to elected representatives; judges would be reviewed each decade and must obtain a simple majority of votes to keep their seats.

While Missouri is seeing efforts to get away from this plan, at least one state is trying to join. Last year, the Nevada legislature passed a proposal to amend the state's constitution and adopt a merit based election system. The measure must pass again in the 2009 legislative session to be placed on the general election ballot.

CONCLUSION

After *Torres*, the debate on judicial selection methods will continue, and ultimately be decided by the voters and their state representatives. Until men and women who serve on the bench are angels, democracy will continue to be "the worst form of government, except for all those other forms that have been tried from time to time."[12]

Endnotes

1 128 S. Ct. 791 (2008).

2 462 F.3d 161, 187 (2006).

3 *Lopez Torres*, 128 S. Ct. at 799.

4 *Id.* at 800.

5 *Id.* at 801.

6 *Id.*

7 *Id.* at 803.

8 Scott Bauer, *Wisconsin Justice Ousted in Nasty Race*, Associated Press, Apr. 2, 2008, *available at* http://ap.google.com/article/ALeqM5jqB2r3wafaaNQ-LgAyyPJnP0z7JwD8VP ST702 (last viewed April 28, 2008).

9 Michael DeBow et al., *The Case for Partisan Judicial Elections*, The Federalist Society, Jan. 1, 2003, *available at* http://www.fed-soc.org/publications/PubID.90/pub_detail.asp.

10 Constance Hately, *Judging the Judges*, Am. Spec., *available at* http:// www.spectator.org/util/print.asp? art_id13096 (last viewed April 28, 2008).

11 Gubernatorial dissatisfaction with the nominees that have been submitted is bipartisan. The Democratic Governor of Tennessee has also objected to the nominees that have been submitted to him.

12 Winston Churchill, Address to the House of Commons (Nov. 11, 1947).

Religious Liberties

Ministers, Minimum Wages, And Church Autonomy

*By Thomas C. Berg**

When Judge Richard Posner expounds on an area of law in one of his opinions for the Seventh Circuit, the result is almost always thought-provoking and fun to read. So it is with his recent panel opinion in *Schleicher v. Salvation Army*,[1] which applied the "ministerial exception" to employment laws to dismiss a suit for minimum wages brought by two former Salvation Army ministers.

Steve and Lori Schleicher were captains in the Salvation Army, ordained by the group to act as clergy and serving as administrators of its Adult Rehabilitation Center in Indianapolis. Such centers operate as "self-contained religious communities," in Posner's words, for alcoholics, drug addicts, and others who reside there and "whom the Salvation Army is attempting to redeem."[2] The complex includes not only living and dining areas but a chapel where the Schleichers preached, led worship and singing, and taught classes in Bible study and Christian living. The Schleichers' duties also included administering five thrift shops, staffed by the center's residents, which sell donated goods to the public.

Salvation Army ministers receive no wages, only "an allowance... sufficient for basic needs."[3] The Schleichers each received $150 a week, which fell below the federal minimum wage, given the number of hours, including overtime, they each worked. They brought suit alleging violations of the minimum-wage and overtime provisions of the Fair Labor Standards Act (FLSA).[4] The Army expelled them for bringing suit. The district court dismissed the suit for lack of subject-matter jurisdiction on the basis of the so-called ministerial exception, and the Schleichers appealed.[5]

The Ministerial Exception

The ministerial exception is a cornerstone of constitutional protection for religious organizations. It was first recognized thirty-five years ago in another employment suit by a Salvation Army officer (minister).[6] Billie McClure brought a Title VII claim alleging discrimination in pay and benefits because of her sex, but the Fifth Circuit affirmed a dismissal for lack of jurisdiction. Recognizing the Army as a church and McClure as one of its ministers, the court held that applying Title VII to that relationship would "cause the State to intrude upon matters of church administration and government," areas "of religious freedom which it is forbidden to enter by the principles of the free exercise clause of the First Amendment."[7] Title VII contains no textual exception from race or sex discrimination claims, the court noted—in contrast to the provision allowing religious organizations to hire and fire based on religion[8]—but following the principle that a statute should be construed so as to avoid "'a serious doubt of [its] constitutionality,'" the court held that Congress in Title VII "did not inten[d] to regulate the employment relationship between church and minister" for any claims of discrimination.[9]

Since *McClure*, dozens of lower court decisions have dismissed claims by clergy against their religious employers alleging violations of Title VII's sex, race, or age discrimination provisions;[10] the Americans with Disabilities Act;[11] and various state statutes or common law doctrines governing the employment relationship.[12] The exception extends to non-Christian clergy and houses of worship; to clergy in other religious organizations such as hospitals;[13] and to non-ordained employees, such as music or education directors, whose primary duties involve teaching the faith, church governance, or supervising ritual or worship.[14] The exception has even grown since the Supreme Court held in *Employment Division v. Smith* (1990) that the Free Exercise Clause usually does not require any exemption for religious conduct from a "neutral law of general applicability."[15] Every circuit to consider the issue has held that the ministerial exception survives *Smith*'s shrinking of free exercise rights.[16]

The ministerial exception is also more absolute than even the most protective free exercise standards before *Smith*. Courts do not ask whether any compelling governmental interest justifies the burden on religious institutions from clergy suits; for example, race discrimination claims are dismissed even though the Supreme Court has found a compelling interest in eliminating race discrimination in other contexts.[17] Moreover, unlike most other free exercise exemptions, the ministerial exception is not limited to cases where discrimination is motivated by religious beliefs, such as the male-only rule for Catholic priests or Orthodox rabbis. A race or sex discrimination suit is barred even if the church's doctrine strongly condemns such discrimination and its agent acted on the basis of pure prejudice.[18]

The ministerial exception remains vigorous because it has deep and historic roots, not only in principles of religious freedom, but in the related doctrine of church-state separation. In *Schleicher*, before applying the exception to the minimum-wage case, Posner touches on these foundations:

> [T]hough [the exception] is derived from policies that animate the First Amendment, the relevant policies come from the establishment clause rather than from the free-exercise clause.... The assumption behind the rule—for it is an interpretive rule—is that Congress does not want courts to interfere in the internal management of churches, as they sometimes do in the management of prisons or school systems. In a religious nation that wants to maintain some degree of separation between church and state, legislators do not want the courts to tell a church whom to ordain (or retain as an ordained minister), how to allocate authority over the affairs of the church, or which rituals and observances are authentic.[19]

This passage recognizes, importantly, that the purposes for the exception are not limited to claims by ministers; indeed, Posner says, it "is better termed the 'internal affairs' doctrine."[20]

* *Thomas C. Berg is the St. Ives Professor of Law at the University of St. Thomas.*

The freedom of religious organizations to govern their internal affairs extends to a number of other situations, including at least the handling of members and congregants—the terms and conditions for their admission, discipline, or expulsion—and the promulgation of the organization's teaching to its members, employees, and volunteers. An influential judge's use of the umbrella term "internal affairs" should encourage courts to connect these other situations to the well-established ministerial exception. (The countervailing risk is that the term "internal" might incorrectly suggest to judges that church autonomy is wholly irrelevant whenever a church's conduct has any external effects—as a great many forms of conduct do.)

The precise constitutional source of the exception has been disputed. Posner locates it solely in the Establishment Clause and its principle of church-state separation. The non-establishment rationale certainly makes sense, and some judges, like Posner, have been drawn to it in order to sidestep *Smith*'s disapproval of mandated exemptions under the Free Exercise Clause.[21] A prime feature of the English established church was that the government decided clerical matters such as the appointment of bishops (technically it still does so today) and other ecclesiastical matters such as the content of the Book of Common Prayer. As a result, several decisions have rested the ministerial exception on the ground that lawsuits would create "excessive entanglement" between church and state in violation of the third prong of the *Lemon v. Kurtzman* test for establishment issues.[22]

More broadly, as several academic commentators have argued, the Establishment Clause reflects a model of "dual jurisdictions," state and church—in the words of Carl Esbeck, two "spheres of competence" covering temporal matters (the state) and spiritual matters (the churches).[23] Non-establishment of religion means that the state has no control over religious affairs. In Ira Lupu and Robert Tuttle's words, the state disclaims "jurisdiction over [the] ultimate truths" that are the subject of religion.[24] This jurisdictional separation makes the state a "penultimate" institution with a limited horizon," forswearing any "comprehensive claim to undivided loyalty."[25] Religious institutions and associations have sovereignty in the sphere of spiritual matters, and this "places a powerful limit" on the ambitions of the state."[26] Indeed, "at crucial points in Western history," Esbeck emphasizes, religious institutions have "had a 'pivotal role in guarding against political absolutism'"[27]—from the medieval conflict between pope and emperor over the power of appointing clergy to the twentieth century's religiously inspired resistance movements against Communism. Church autonomy is crucial in our constitutional order, therefore, for the sake not only of religious freedom but of limited government.

The line between state and church jurisdictions is not always clear, and they may overlap; but as the medieval conflict suggests, a crucial feature of an independent religious sector is the ability of religious organizations to choose their leaders according to their own standards. For the state to interfere in such decisions would wrongly make it the "coauthor" of faith; thus, the ministerial exemption is among "the entailments of the jurisdictional limitations that the Establishment Clause places on the state's role."[28]

This non-establishment ground runs deep, but Posner is mistaken to write off the Free Exercise Clause as an alternative ground for churches' right of control over their internal affairs. The Supreme Court has specifically based this right in free exercise. Holding more than fifty years ago that a state could not throw its weight on one side of a schism in the Russian Orthodox Church, the Court endorsed "a spirit of freedom for religious organizations, ... in short, power to decide for themselves, free from state interference, matters of church government as well as those of faith and doctrine. Freedom to select the clergy, where no improper methods of choice are proven, [has] federal constitutional protection as a part of the free exercise of religion."[29] *McClure* relied on this passage to hold that the ministerial exception was necessary to protect free exercise.[30] Indeed, it may be most natural to think of a secular intrusion on clergy selection, by a state with no established church, as violating free exercise rather than non-establishment.

Employment Division v. Smith does not dispose of the free exercise argument; on the contrary, *Smith*'s language preserves it. Before announcing its general rule against exemptions, the Court made clear that the Free Exercise Clause still forbids the government from, among other things, "lend[ing] its power to one or the other side in controversies over religious authority or dogma."[31] For that proposition, *Smith* cited the very same case, *Serbian Eastern Orthodox Diocese v. Milivojevich*,[32] that Posner cites as authority for the "internal affairs" exception.[33]

In fact, both Religion Clauses work together in this context; institutional free exercise coincides with institutional separation from government. Thus, in several cases affirming churches' autonomy, the Court has simply cited "the First Amendment" without differentiating the clauses.[34]

Posner is also on shaky ground when he calls the ministerial exception merely "a rule of interpretation, not a constitutional rule" (albeit a rule "derived from the policies that animate the First Amendment").[35] Most cases treat it as a constitutional mandate, not just a statutory construction designed to avoid constitutional questions.[36] Even *McClure*, which spoke in terms of avoiding "serious doubts" about constitutionality, made clear that a Title VII suit would draw the court into areas "it is forbidden to enter by... the First Amendment."[37] The distinction between interpretive rule and constitutional mandate matters, not because Title VII is likely to be amended to authorize ministers' suits, but because state statutes might be interpreted to authorize them. That is exactly what happened in the union context; after the Supreme Court construed the National Labor Relations Act's coverage to exclude teachers at parochial schools because of potential constitutional difficulties,[38] lower courts applied state labor laws and brushed aside the constitutional questions.[39] Posner himself has criticized the canon "that statutes should be construed not only to save them from being invalidated but to avoid even raising serious constitutional questions," on the ground, among others, that it leaves constitutional boundaries too vague.[40]

As we will see, treating the ministerial exception as a rule of narrow construction may make somewhat more sense for the FLSA than for Title VII and other antidiscrimination laws.[41] But for antidiscrimination suits, *Schleicher*'s dictum is best read to say that they fall outside Title VII because they are unconstitutional, not because they merely raise constitutional questions.

Minimum-Wage Suits and the Exception's Scope

Most, although not all, ministerial exception cases have involved Title VII or other antidiscrimination claims. The Schleichers argued that their minimum wage suit was different, raising none of the evils that the exception aims to prevent. In most cases about clergy wages, this would be irrelevant, and the ministerial exception unnecessary: the FLSA applies only to "an enterprise engaged in commerce" (clarified by regulation to mean "ordinary commercial activities"), which most churches and religious organizations do not do.[42] But the Schleichers' case was complicated, and interesting, because part of their duties was to supervise the commercial thrift shops. The shops, as Posner pointed out, resembled those involved in the Supreme Court case of *Tony & Susan Alamo Foundation v. Secretary of Labor*[43]—commercial businesses used to finance a religious organization and staffed by the alcoholics and addicts whom the organization was trying to rehabilitate—and the Court had held the Alamo staffers covered by the FLSA. The question was whether the Schleichers as ministers were different. Did their claim implicate the evils the ministerial exception aims to avoid?

The narrowest evil from a minister's suit is that it may require the court to decide theological questions, a task plainly outside the power of civil authorities under the separate-jurisdictions, Establishment Clause framework described above.[44] As Posner noted in *Schleicher* (following several other decisions), a church will often answer a Title VII antidiscrimination suit by claiming that the plaintiff was fired or disfavored because she performed poorly as a minister, and "to evaluate such a defense—to determine, that is, whether it was sincere or pretextual—would require a court to weigh in on issues of [religious] doctrine and practice."[45] The court would inquire whether the minister did perform poorly or worse than others who were not fired. The impropriety of setting such standards is why courts have also uniformly rejected tort suits alleging so-called clergy malpractice in pastoral counseling relationships.[46]

But if inquiries into clergy performance and standards are the only wrong to avoid, a suit that required no such determination could proceed. That is what the Schleichers argued about their FLSA claim, since the duty to pay minimum wages does not depend on the employer's motive for withholding them. Indeed, even Title VII suits, the core of the ministerial exception, do not always or necessarily raise the issue of the minister's religious performance. The rationale for the exception must be broader.

Thus, *Schleicher* emphasizes, along with many other decisions, that the ministers' exception more expansively protects "a church's ability to determine who shall be its ministers."[47] The minister is (in *McClure*'s words) the church's "lifeblood," "the chief instrument by which [it] seeks to fulfill its purpose,"[48] and (in the Third Circuit's words) "the embodiment of [the church's] message," its "public representative, its ambassador, and its voice to the faithful."[49] As a result, as the Third Circuit puts it, the ministerial exception bars any claim "the resolution of which would limit a religious institution's right to select who will perform particular spiritual functions," "'even when such actions are not based on issues of church doctrine or ecclesiastical law.'"[50] The protected choice of ministers is implicated obviously when a Title VII plaintiff seeks reinstatement, but also when he seeks to impose liability for the church's hiring or firing decision.

But the Schleichers argued that their suit did not implicate this concern either. Prima facie, it seems, the Salvation Army could retain its right to choose its ministers while paying them above-minimum wages (at least when they supervised commercial activities).

Posner nonetheless applied the ministerial exception, articulating two arguments. First, he analogized the case of Salvation Army ministers receiving subsistence wages and supervising thrift shops to the example of monks who "take a vow of poverty" and produce wine "in order to finance the operation of the[ir] monastery."[51] The monks would not fall within the FLSA as employees of an organization "engaged in ordinary commercial activities": "[t]he vow of poverty is a hallowed religious observance," and "an intent to destroy it cannot reasonably be attributed to the [FLSA's] draftsmen," for "[n]o one would think the curious precapitalist economy of a monastery an ordinary commercial activity."[52] Similarly, while the Salvation Army's thrift shops might be commercial and their sales staff subject to the FLSA per *Alamo*, the Schleichers were employed not by the shops but by the overall Adult Rehabilitation Center, which is a church—with its worship services and Bible studies—and for which the Schleichers' duties including preaching, teaching, and counseling as well as supervision of the shops. The Schleichers therefore were like the bishop of a cathedral with a profit-generating gift shop; the bishop remains fundamentally an ecclesiastical rather than commercial administrator, outside the FLSA. "The commercial tail," Posner concluded, "must not be allowed to wag the ecclesiastical body."[53]

This reasoning may seem simply an interpretation of the statute: clergy working for an overall religious organization are not employees of an enterprise "engaged in ordinary commercial activities" simply because the organization has a commercial component. But the interpretation plainly was motivated by considerations related to the ministerial exception. *Schleicher* ultimately adopts "a presumption that clerical employees are not covered by the [FLSA]."[54] The presumption can be defeated by a showing that the "minister's function is entirely rather than incidentally commercial," for example if "a church received by inheritance a steel plant, and it happened to have among its ministers a former steel executive whom it assigned to manage the plant full time."[55] But this appears to leave the ministerial exception applicable to another category: ministers whose primary duties involve a commercial enterprise with a religious motivation like the Salvation Army thrift shops or the Alamo Foundation businesses. Suppose, for example, the Schleichers had done all their work in the thrift shops, not just a small share.

Ultimately, in treating nearly all ministers as outside the FLSA, Posner offers reasons grounded in the ministerial exception. One, already mentioned, is that the vow of poverty for clergy is a "hallowed religious observance." At least some minimum-wage cases, therefore, implicate a church's doctrinal belief in voluntary poverty for its leaders—just as a sex discrimination case against the Catholic Church would implicate the Church's doctrine of a male-only clergy. When such a doctrinal belief is involved, a ministerial exception need not rest on a broader right of church autonomy.

But Posner adds a second argument that does sound in church autonomy: that even minimum wage suits ultimately reduce to (impermissible) challenges to a church's ability to choose and evaluate its ministers. The Schleichers, for example, were dismissed from the Army when they filed suit, but they added no retaliation claim nor could they—because if the Army had answered that "their filing a suit seeking to enforce wage and overtime claims was inconsistent with their religious obligations as ministers and was thus an independent and adequate ground for firing them, the court would [improperly] have to explore the doctrines of the Salvation Army that define the role of its ministers."[56] Given that the Army had a policy against ministers filing suit, and given that its firing of ministers was unreviewable, "then however we rule no Salvation Army minister will ever receive the minimum wage. We are disinclined to take the first step on a path that leads so swiftly to so dead an end."[57]

Schleicher's Implications

By virtue of the church autonomy arguments in the preceding two paragraphs, *Schleicher* extends beyond a statutory interpretive rule that ministers normally fall outside the FLSA's coverage of "commercial" activities. The logic that a church can dismiss a minister for bringing a suit deemed inconsistent with the ministerial role, and that therefore it is pointless for a court to entertain the suit in the first place, creates a powerful constitutional shield covering a wide range of terms and conditions of clergy employment.[58] This coincides with the statement in *McClure*, the original ministerial-exception case, that the church's protected interests include not only the selection of a minister but "the determination of a minister's salary, his place of assignment, and the duty he is to perform in the furtherance of the religious mission of the church."[59] It also coincides with the idea that the church-minister relationship is, at least in the vast majority of situations, outside the civil courts' jurisdiction. If other courts follow this aspect of *Schleicher*, it would strengthen the ministerial exception.

On the other hand, *Schleicher* does not speak directly to the situation where courts have been most uneven in shielding churches' decisions concerning clergy: third-party suits alleging that the church was negligent in hiring or supervising a minister who sexually abused a child or adult parishioner. Some courts have held that such suits do not affect church autonomy because negligence is a wholly secular standard and the church's interest in selecting its ministers is not implicated in suits by third parties as opposed to suits by ministers themselves. The Ninth Circuit went so far as to hold that even a minister's suit—a Jesuit trainee's claim against the order for negligently failing to stop alleged sexual harassment by his superior—could proceed because the Jesuits still favored the plaintiff's ordination and therefore the order's "freedom to choose its representatives" would be unaffected.[60] But this ignores the order's interest in freedom to act concerning the clergy who are potential or alleged wrongdoers. A religious organization's ability to select clergy according to its preferred model can be severely affected by liability, imposed in hindsight, for failure to predict that a minister would become a wrongdoer, or to intervene; and that determinations of what a "reasonable bishop" would have done in such circumstances can easily require courts to impose their own vision of a proper ecclesiastical structure.[61]

At the same time, third-party suits do involve the interests of people—especially children—who, unlike clergy plaintiffs, have not voluntarily entered into a church-minister relationship. A sensible balance of these competing interests would hold churches liable for failures to supervise that are reckless or intentional—that is, knowing of a substantial risk of a minister's abusive tendencies—as opposed to merely negligent.[62] Courts will only draw this balance if they recognize that even third-party suits can affect fundamental interests in clergy selection. But suits by ministers themselves, such as *Schleicher*, do not call attention to that fact.

The final point in *Schleicher* is the court's conclusion that the dismissal of the plaintiffs' suit, although proper, should have been on the merits, through judgment on the pleadings, not for lack of subject matter jurisdiction.[63] This too is questionable. If the separation-of-jurisdictions model is accurate, then suits over clergy matters really do exceed the civil court's jurisdiction.[64] Numerous ministerial-exception decisions, beginning with *McClure v. Salvation Army* itself, have dismissed suits for lack of subject matter jurisdiction.[65] Posner argues that "[a] federal court could not entertain a suit to restore the Latin mass or to declare Christian Science a heresy. But it does have jurisdiction to decide cases brought to enforce the Fair Labor Standards Act.... Jurisdiction is determined by what the plaintiff claims rather than by what may come into the litigation by way of defense."[66] But the defense of sovereign immunity is generally treated as jurisdictional, and it resembles the ministerial exception. Both place fundamental limits on the powers of civil courts in order to preserve the sovereignty of another actor, whether government or church.[67]

Endnotes

1 518 F.3d 472 (7th Cir. 2008).

2 *Id.* at 476.

3 *Id.* at 474 (quotation in original).

4 29 U.S.C. § 201 *et seq.*

5 Schleicher v. Salvation Army, 2007 WL 129041 (S.D. Ind. Jan. 12, 2007). Posner calls it the "ministers' exception" (*Schleicher*, 518 F.3d at 474), presumably because "ministerial" misleadingly connotes "administrative" or "non-discretionary." But I will use the more familiar term "ministerial exception."

6 McClure v. Salvation Army, 460 F.2d 553 (5th Cir. 1972).

7 *Id.* at 560.

8 42 U.S.C. § 2000e-1.

9 *McClure*, 460 F.2d at 560-61 (quoting Ashwander v. TVA, 297 U.S. 288, 348 (1936) (Brandeis, J., concurring)).

10 *See, e.g.,* Rweyemamu v. Cote, 520 F.3d 198 (2d Cir. 2008) (race discrimination); Tomic v. Catholic Diocese of Peoria, 442 F.3d 1036 (7th Cir.

2006) (Posner, J.) (age discrimination); Rayburn v. Gen'l Conf. of Seventh-Day Adventists, 772 F.2d 1164 (4th Cir. 1985) (sex discrimination).

11 *See* Hollins v. Methodist Healthcare, Inc., 474 F.3d 223 (6th Cir. 2007).

12 *See, e.g.,* Natal v. Christian and Missionary Alliance, 878 F.2d 1575 (1st Cir. 1989) (state wrongful-termination claim).

13 *See, e.g.,* Shaliehsabou v. Hebrew Home of Greater Washington, Inc., 363 F.3d 299, 310 (4th Cir. 2004) (predominantly Jewish nursing home); Scharon v. St. Luke's Episcopal Presbyterian Hosp., 929 F.2d 360 (8th Cir. 1991) (hospital chaplain).

14 *See, e.g., Petruska*, 462 F.3d at 304 n.6 (quoting *Rayburn*, 772 F.2d at 1169); *Tomic*, 442 F.3d 1036 (choir director).

15 494 U.S. 872, 879 (1990).

16 *See, e.g.,* Gellington v. Christian Methodist Episcopal Church, 203 F.3d 1299, 1303 (11th Cir. 2000); Combs v. Central Tex. Ann. Conf., United Methodist Church, 173 F.3d 343, 349 (5th Cir. 1999); EEOC v. Catholic University, 83 F.3d 455, 462 (D.C. Cir. 1996).

17 Bob Jones University v. United States, 461 U.S. 574, 604 (1983).

18 *See, e.g.,* Simpson v. Wells-Lamont Corp., 494 F.2d 490, 493-94 (5th Cir. 1974)).

19 518 F.3d at 475.

20 *Id.*

21 *Id.* ("In reading into statutes of general applicability an exception favorable to religious organizations, the courts may seem to be flouting [*Smith*].").

22 Lemon v. Kurtzman, 403 U.S. 602, 613 (1971). *See, e.g., Catholic University*, 83 F.3d at 465-67; *Rayburn*, 772 F.2d at 1169-71.

23 Carl H. Esbeck, *The Establishment Clause as a Structural Restraint on Government Power*, 84 Iowa L. Rev. 1, 28, 10 (1998).

24 Ira C. Lupu & Robert W. Tuttle, *The Distinctive Place of Religious Entities in Our Constitutional Order*, 47 Vill. L. Rev. 37, 83-84 (2001).

25 *Id.*

26 Gregory A. Kalscheur, *Civil Procedure and the Establishment Clause: Exploring the Ministerial Exception, Subject Matter Jurisdiction, and the Freedom of the Church*, 17 Wm. & Mary Bill Rts. J. (forthcoming 2008), at 27, *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1108014.

27 Esbeck, *supra* note 23, at 67 (quotation omitted). *See also* Richard W. Garnett, *The Freedom of the Church*, 4 J. Cath. Soc. Thought 59 (2006).

28 Ira C. Lupu and Robert W. Tuttle, *Sexual Misconduct and Ecclesiastical Immunity*, 2004 B.Y.U. L. Rev. 1789, 1815.

29 Kedroff v. St. Nicholas Cathedral, 344 U.S. 94, 116 (1952). The classic exposition of free exercise church autonomy rights is Douglas Laycock, *Towards a General Theory of the Religion Clauses: The Case of Church Labor Relations and the Right to Church Autonomy*, 81 Colum. L. Rev. 1373 (1981).

30 460 F.2d at 559-60.

31 494 U.S. at 877.

32 426 U.S. 696 (1976)

33 *See Smith*, 494 U.S. at 877 (citing *Milivojevich*, 426 U.S. at 708-25); *Schleicher*, 518 F.3d at 475 (citing *Milivojevich*, 426 U.S. at 708-15).

34 Presbyterian Church v. Hull Church, 393 U.S. 440, 449-51 (1969); *see also Milivojevich*, 426 U.S. at 708-10, 725 ("the First and Fourteenth Amendments"); NLRB v. Catholic Bishop, 440 U.S. 490, at 499-504 (1979) ("the First Amendment" and "the Religion Clauses").

35 *Schleicher*, 518 F.3d at 475.

36 *See, e.g., Rweyemamu*, 520 F.3d at 207 (exemption is "constitutionally required"); *Petruska*, 462 F.3d at 303 n.4 ("the plain text of Title VII and its legislative history foreclose the possibility of imposing a limiting construction upon the statute," but allowing ministers' Title VII suits would violate First Amendment); *Rayburn*, 772 F.2d at 1165-67 (same).

37 460 F.2d at 560.

38 *Catholic Bishop*, 440 U.S. 490 (1979).

39 South Jersey Catholic Sch. Teachers Ass'n v. St. Teresa of the Infant Jesus Ch. Elementary School, 150 N.J. 575, 696 A.2d 709 (1997) (New Jersey collective-bargaining law could apply to wages, certain benefits, and other secular terms of employment); Catholic High School Ass'n v. Culvert, 753 F.2d 1161 (2d Cir. 1985), *cert. denied*, 484 U.S. 830 (1987) (New York law).

40 Richard A. Posner, *Statutory Interpretation in the Courtroom and in the Classroom*, 50 U. Chi. L. Rev. 800, 815-16 (1983).

41 *See infra* notes 51-53 and accompanying text.

42 *See* 29 U.S.C. §§ 206(a), 207(a)(1)); 29 C.F.R. § 779.214 (cited in *Schleicher*, 518 F.3d at 476). In addition, in most cases coverage of ministers might be limited because they are "supervisors" or "administrators," an exception "mysteriously not invoked" by the Salvation Army. *Schleicher*, 518 F.3d at 478.

43 471 U.S. 290 (1985).

44 *See supra* notes 23-28 and accompanying text.

45 *Schleicher*, 518 F.3d at 475. *See also, e.g., Tomic*, 442 F.3d at 1037-38.

46 *See, e.g.,* Nally v. Grace Comm. Church, 47 Cal. 3d 278, 763 P.2d 948, 253 Cal. Rptr. 97 (1988), *cert. denied,* 490 U.S. 1007 (1989); William W. Bassett, Religious Institutions and the Law § 8:19 n.9 (collecting cases following *Nally*).

47 *Schleicher*, 518 F.3d at 475.

48 460 F.2d at 558-59.

49 *Petruska*, 462 F.3d at 306.

50 *Id.* at 307; *id.* at 304 n.7 (quoting *Gellington*, 299 F.3d at 1303; *Combs*, 173 F.3d at 350).

51 *Schleicher*, 518 F.3d at 476 (citing "Wine-Tasting and Retreats at California Monastery," May 23, 2006, www.ajc.com/travel/content/travel/otherdesti nations/us_ stories/052406monastery.html)).

52 518 F.3d at 476.

53 *Id.* at 476-77.

54 *Id.* at 478.

55 *Id.*

56 *Id.* at 474. The same would hold if an organization paid clergy above-minimum wages but required them to return a portion and fired one for failing to do so. *Id.* at 477.

57 *Id.*

58 *See, e.g.,* Malichi v. Archdiocese of Miami, 945 So.2d 526, 531-32 (Fla. App. 1 Dist. 2006) (holding that priest's workers' compensation suit would violate church autonomy doctrine because of, among other things, possibility of anti-retaliation claim if diocese dismissed priest for filing compensation suit), *review denied,* 965 So.2d 1277 (Fla. 2007). *But see, e.g.,* St. John's Lutheran Church v. State Compensation Ins. Fund, 252 Mont. 516, 525-26, 830 P.2d 1271, 1277-78 (1992) (holding that workers' compensation coverage for pastor did not interfere with church-pastor relationship).

59 460 F.2d at 559.

60 Bollard v. Cal. Province, Society of Jesus, 196 F.3d 940, 947 (9th Cir. 1999), *reh'g denied,* 211 F.3d 1331 (9th Cir. 2000) (en banc).

61 Lupu & Tuttle, *supra* note 24, at 1854-56.

62 *See, e.g., id.* at 1860-67; Gibson v. Brewer, 952 S.W.2d 239, 247 (Mo. 1997) (adopting "intentional failure to supervise clergy" standard).

63 518 F.3d at 478.

64 For elaboration of this argument, see Kalscheur, *supra* note 26.

65 *McClure*, 460 F.2d at 561; *see also* Hollins v. Methodist Healthcare, 474 F.3d 223; Kalscheur, *supra* note 26, at 13-14 n.47 (collecting cases). But see, in addition to *Schleicher*, *Petruska*, 462 F.3d at 302-03 (exception reflects failure to state a claim, not lack of jurisdiction).

66 *Schleicher*, 518 F.3d at 478.

67 *See* Kalscheur, *supra* note 26, at 53-56 (citing sources).

The Application of RFRA to Override Employment Non-Discrimination Clauses Embedded in Federal Social Service Programs

*By Carl H. Esbeck**

The basic federal employment nondiscrimination law is Title VII of the Civil Rights Act of 1964.[1] The Act prohibits discrimination in employment on the basis of race, color, national origin, sex, or religion, and is binding when an employer has fifteen or more employees. However, section 702(a) of the act acknowledges the freedom of religious organizations to take religion into account in their employment decisions.[2] Moreover, the section 702(a) exemption is not forfeited when a faith-based organization accepts government grant funding,[3] nor does the exemption thereby become a religious preference in violation of the Establishment Clause.[4] There is more to federal civil rights compliance than Title VII, however, and strictly speaking the section 702(a) exemption is applicable only to claims brought under Title VII.[5]

I. Program-Specific Non-Discrimination Clauses in Federal Legislation

Almost all federal funding awards to independent-sector organizations to provide social services take the form of a "grant" or "cooperative agreement," rather than the form of a government "contract" for services.[6] Some federal social service programs have embedded in their authorizing legislation a nondiscrimination clause binding on the recipients of program grants. While this is true of only a minority of all federal welfare programs, still the number of programs with embedded clauses is not insubstantial. The principal thrust of these clauses is to prohibit discrimination against the ultimate beneficiaries of the social service programs. However, a few of the embedded clauses expressly prohibit discrimination by a service provider against its employees in addition to discrimination against the ultimate beneficiaries. In still other programs the embedded clauses prohibit discrimination against the "intended beneficiaries" of the funded social services, but "intended beneficiaries" has been interpreted broadly in judicial decisions to prohibit discrimination not only against the program's ultimate beneficiaries but also against the service provider's employees.[7]

These embedded employment nondiscrimination clauses are presumptively binding on all recipients of federal grants awarded under the programs in question. Examples are the Workforce Investment Act of 1998,[8] administered by the United States Department of Labor, and the first title of the Omnibus Crime Control and Safe Streets Act of 1968,[9] administered by the United States Department of Justice. An embedded clause, if it covers employees as well as ultimate beneficiaries, typically prohibits discrimination against a grantee's employees only while the employees are working in the government-funded program. Two programs, AmeriCorps VISTA and AmeriCorps State and National, both operated by the Corporation for National and Community Service, are unusual in this regard. These programs have an embedded clause restricting religious staffing, but the restriction is limited only to employees newly hired (if any) with the federal program funds in question.[10]

II. The Religious Freedom Restoration Act of 1993

The Religious Freedom Restoration Act of 1993 (RFRA)[11] gives relief to persons of faith and to faith-based organizations.[12] RFRA prohibits intentional discrimination on the basis of religion, but more importantly it also gives relief from substantial burdens on religion when the burden is the unintended impact of a generally applicable federal law.[13] For the religious claimant there are three elements to a prima facie case under RFRA: (1) that the professed religious practice is sincere; (2) that the burden on the practice is substantial; and, (3) that the practice is an exercise of religion. The government or other RFRA defendant has the affirmative defense (and thus the burden of proof to show) that application of the law is in furtherance of a compelling governmental interest and is the least restrictive means of achieving that governmental interest.[14]

When the aforementioned embedded employment nondiscrimination clauses apply to faith-based social service providers that staff on a religious basis, may these providers turn to RFRA for protection? The United States Department of Justice has answered in the affirmative. In October 2007 the Office of Justice Programs [OJP], which administers the social service programs within the Department of Justice, along with the Taskforce for Faith-Based and Community Initiatives within the Office of the Deputy Attorney General, posted on the Department of Justice webpage the determination that RFRA enables religious organizations to be eligible for federal grants while continuing to employ those of like-minded faith.[15] And faith-based organizations may do so "even if the statute that authorized the funding program generally forbids consideration of religion in employment decisions by grantees."[16] Successfully invoking RFRA is conditional on the sincerity of the faith-based grantee's professed religious motive for involvement in the program, and whether requiring the grantee to choose either religious staffing or federal funding would be a substantial burden on its religion.[17]

The October 2007 posting by OJP came about as a result of a formal request submitted by World Vision, a Christian world relief and community development organization. In early 2005, World Vision was awarded a $1.5 million grant by OJP to address an escalating gang presence and juvenile crime in Northern Virginia. The grant was awarded under the Juvenile Justice and Delinquency Prevention Act, which is subject to provisions of the Omnibus Crime Control and Safe Streets Act. The latter requires that grant applicants not discriminate in employment on the basis of religion when using grant monies.[18] As with many faith-based organizations, World Vision does consider religion when hiring and thus sought a determination that it could safely rely on RFRA and continue its hiring

* *Carl H. Esbeck is the R.B. Price Professor and Isabelle Wade & Paul C. Lyda Professor of Law at the University of Missouri.*

practices. After some delay, in June of 2007 the Office of Legal Counsel (OLC) at the Department of Justice provided OJP with a written legal opinion to the effect that RFRA did override conflicting federal employment nondiscrimination clauses, that World Vision was religiously motivated in its practice of staffing on a religious basis, and that World Vision would be substantially burdened if it could not continue to employ staff of like-minded faith while administering the grant. Although the OLC legal opinion provided to OJP is confidential under the attorney-client privilege, World Vision was soon advised as to the favorable conclusion. The aforementioned posting on the Department of Justice webpage in October 2007 made the ultimate determination available to other religious organizations awarded or applying for OJP grants. Because RFRA applies to social service grants issued by the Department of Justice, it necessarily follows that RFRA applies to grants awarded by other departments and agencies such as the Department of Labor and the Department of Housing and Urban Development.

RFRA protects religious practices from substantial burdens that are imposed by the federal government.[19] Religious charities have a strong interest in maintaining their religious character, and that character in turn is modeled to the poor and needy through its employees. The White House Office of Faith-Based and Community Initiatives published a booklet in June 2003 acknowledging that a faith-based organization's ability to select employees that share its religious values is vital to the group's self-identity and continued ministry.[20] It argues that nonreligious organizations receiving federal grant monies freely hire based on their core mission, just as Planned Parenthood requires that employees be pro-choice and Sierra Club screens applicants based on their view of global warming. Religious groups likewise cannot remain true to their founding creedal purposes unless employees are aligned with the energizing core of the mission.

It is true, of course, that when the aid is direct the government-funded social services must be delivered without prayer, proselytizing, or other inherently religious activities,[21] all as required by the separation of church and state. So congressional critics have argued that the delivery of government-funded services does not require an employee of a particular religion. The quip heard among the critics has been: "An evangelical homeless shelter doesn't need an evangelical employee if all she is doing is ladling soup in a feeding line." But the quip evidences an ignorance of religion. More to the point is Justice William Brennan's observation in his concurring opinion in *Corporation of the Presiding Bishop v. Amos*.[22] Justice Brennan notes that a religious organization is "an ongoing tradition of shared beliefs, an organic entity not reducible to a mere aggregation of individuals."[23] The organization's choice of whom to hire is an important means by which the group "defines itself."[24] The civil courts should be solicitous of those choices because "furtherance of the autonomy of religious organizations often furthers individual religious freedom as well."[25] Religious charities often regard the provision of social services as a means of fulfilling religious duty and as providing a ready example of the life of faith they seek to foster. Religious organizations like World Vision will tell you that their work in reducing gang violence and juvenile delinquency is successful among hard-to-reach adolescents because its employees credibly do what they do with the genuine care and sustained love that only their faith makes possible.

Not all employment discrimination is the same. Disapproving of a job applicant because of her race is senseless and invidious.[26] But one's religious beliefs speak to real and important differences about life's present purposes as well as the ultimate meaning of life, which in turn shape one's vocational objectives and job performance. While the Constitution ascribes no value to racial discrimination, a religious organization's employment discrimination on the basis of religion is often protected as a matter of free exercise. One who has never disagreed with others about religion is not thereby commendably tolerant, but is treating religious differences as trivial, as if religious beliefs do not matter. That is just a soft form of religious bigotry.

III. Additional Arguments by Critics of the Faith-Based Initiative

The most common response to a request such as that of World Vision is that if a faith-based organization does not want restrictions on its hiring then they should not take the money. But there is little doubt that a religious hiring restriction puts enormous pressure on faith-based organizations to recant on a cardinal religious tenet or lose the grant and with it the opportunity to help America's poor and needy. RFRA defines "exercise of religion" broadly to include "any" exercise, whether or not the exercise is "compelled by, or central to, a system of religious belief."[27] Every personnel decision by a religious organization has the potential for being an exercise of religion. And in an organization highly integrated in its faith and the delivery of social services, every personnel decision has the very real potential to be an "exercise of religion" as defined in RFRA.

Opponents continue to insist, however, that a religious organization can easily avoid the religious burden by simply forgoing the competition for federal grant monies. But requiring withdrawal from involvement in modern public life is hardly equitable treatment. Just as the government cannot justify restricting a particular form of speech (e.g., passing out handbills on a public street) merely by pointing to other opportunities that a person has to express herself (e.g., writing a letter to the editor of a newspaper), so the government cannot restrict a particular exercise of religion by pointing to some other course of action where the organization's religious practices are not penalized. And in any event, the question is free of serious doubt under RFRA. RFRA states that a "denial of government funding" on account of a social service provider's religion or religious practice can trigger RFRA's protection.[28] This is only logical. Congress enacted RFRA to "restore" the standard of protection for religious free exercise originally reflected in *Sherbert v. Verner*,[29] a case about a denial of government funding.[30] The Supreme Court held in *Sherbert* that an individual refusing to take a job entailing work on her Sabbath could not be put to the "cruel choice" of either forfeiting her claim for unemployment benefits or violating her religious Sabbath. Likewise, a faith-based organization cannot be put to the "cruel choice" of either forfeiting its ability to compete for valuable federal grant monies or violating its religious practice of employing those of like-minded faith.

As noted above, the term "religious exercise" is broadly defined in RFRA to include "any exercise of religion, whether or not compelled by, or central to, a system of religious belief."[31] Nonetheless, opponents of the faith-based initiative argue that for government to decline to facilitate the free exercise of religion is not a religious burden at all, whether substantial or de minimis. The argument will not stand close analysis. It is true, of course, that the Free Exercise Clause of the First Amendment is written in terms of what the government cannot do to a faith-based organization and not in terms of what a faith-based organization can exact from the government. But that line of argumentation does not describe what is occurring here. The government may indeed choose to deliver all social services by itself. In such circumstances, the fact that a faith-based provider cannot win a grant is not a free exercise burden because no one in the independent sector is eligible to win a grant.[32] The federal government, however, has not chosen such a path. Instead, almost all government social services are delivered by the independent sector. Having chosen to deliver welfare services through providers in the independent sector, the federal government cannot then pick and choose among the available providers using eligibility criteria that have a discriminatory impact on faith-based providers. A discriminatory impact on a religious practice as a result of an otherwise neutral law is the very type of occurrence that Congress sought to halt by enacting RFRA. RFRA states that "[g]overnment shall not substantially burden a person's exercise of religion even if the burden results from a rule of general applicability."[33]

Conceding, as they must, that by its terms a denial of grant funding can trigger RFRA protection, opponents of the faith-based initiative argue that RFRA cannot be invoked by a religious charity because the loss of grant monies is not a "substantial" religious burden.[34] This makes no sense. It is true that religious organizations making claims of increased financial burden, without more, have not been excused from compliance with general regulatory and tax legislation. [35] That is, it is not enough simply to show that a religiously neutral law increases a faith-based provider's cost of operation. But such cases have no resemblance to the claim of substantial burden here. Instead, an embedded restriction on religious staffing uniquely harms a faith-based organization by preventing it from sustaining its religious character by hiring those of like-minded faith. The harm is not financial or increased operating cost; the harm is uniquely religious.[36] A prohibition on religious staffing cuts the very soul out of a faith-based organization's ability to define and pursue its spiritual calling, as well as its ability to sustain its vision over generations.

RFRA itself can be overridden, of course, upon proof by the federal government of a "compelling governmental interest."[37] In the recent case of *Gonzales v. O Centro Espirita Beneficente Uniao do Vegetal*,[38] the Supreme Court held that under RFRA the government's showing of a compelling interest is limited to the particular exercise of religion by the claimant in the case.[39] In *Centro Espirita*, a religious group asked for an exemption from the Controlled Substances Act so that it could import a particular narcotic used only by adults during one of its religious ceremonies. The federal government opposed the importation request insisting that there was a "compelling governmental interest" in the uniform enforcement of the Controlled Substances Act.[40] In other words, the government claimed a compelling interest in no exceptions for anyone for any reason because to grant an exception for one would mean having to consider other requests for exemptions.[41] The Court rejected that interpretation of RFRA. The Court said that the proper statutory inquiry was more focused in that "RFRA requires the Government to demonstrate that the compelling interest test is satisfied through application of the challenged law 'to the person'—the particular claimant whose sincere exercise of religion is being substantially burdened."[42] Following the Court's rationale in *Centro Espirita*, the compelling-interest inquiry in a religious staffing case is not in uniformly preventing employment discrimination on a religious basis by all religious grantees. Rather, the government has to more narrowly show that it has a compelling interest in preventing the particular practice of religious staffing by the particular religious grantee in question.

It is a near impossibility for the government to meet such a focused evidentiary burden, and it is absurd to claim that the elimination of religious staffing by a particular faith-based organization is a compelling interest. Congress sought to achieve just the opposite when it provided in section 702(a) that Title VII's ban on religious discrimination should not apply to religious organizations. Section 702(a) is a parallel policy choice by Congress to accommodate the religious freedom of religious organizations. If anything, accommodating religious staffing expands religious freedom—and the expansion of religious freedom is a strong governmental interest, the leading example of which is the First Amendment. Lastly, it has been observed that protecting the religious character of faith-based organizations that participate in government programs expands the array of choices available to the poor and needy, some of whom desire to seek out assistance at robustly faith-centered providers.[43]

Permitting religious charities to staff on a religious basis does not undermine compelling social norms or enduring constitutional values. Just the opposite is true. The religious staffing freedom minimizes the influence of government actions on the religious choices of both religious providers and those wanting to receive services from a faith-based provider. Finally, safeguarding a faith-based organization's freedom of religious staffing advances the Establishment Clause value of noninterference by government in the religious affairs of religious communities.[44] Senator Sam Ervin (D-N.C.) said it more colorfully upon the revision of Title VII when he stated that the aim of the staffing freedom is to "take the political hands of Caesar off of the institutions of God, where they have no place to be." In *Corporation of the Presiding Bishop v. Amos*, the Supreme Court put its seal of approval on that congressional judgment concerning proper church-state relations.[45]

CONCLUSION

Freedom for religious staffing by faith-based grantees enhances our nation's religious pluralism and undeniable dynamism. Authentic pluralism[46] is rightly accommodated, not diminished, when the government affirms the equal treatment of these independent-sector providers to participate in social service programs. To do otherwise would privilege secularism, driving robust faith-based organizations underground and away from participation in modern public life. That would be more like hostility toward religion than neutrality toward religion. By the same token, the approach of government neutrality permits faith-based organizations to preserve their institutional character which is necessary to perpetuate their distinctive way of life. These are the social norms to be upheld and the enduring constitutional values to be reinforced. In the face of these realities, the opponents' bald assertions that a ban on religious staffing by federal grantees holds the moral high ground is little more than self-flattery.

Many religious organizations care deeply about retaining the ability to participate fully and equally in modern public life, while retaining their full character as religious organizations of integrity and vision. Not every religious grantee will care about the freedom to staff on a religious basis, of course, but many do.[47] And even this variance among religious groups goes to underline America's religious pluralism made possible only when America's religious freedom is extended to all. When RFRA overrides embedded employment nondiscrimination clauses, the rule of law chooses freedom over a crabbed notion of equality that acts to oppress the vital need of robust religious organizations to retain their institutional autonomy. This freedom, made possible by Congress in passing RFRA, is to be celebrated as in the best of our nation's legal traditions.

Endnotes

1 42 U.S.C. §§ 2000e to -17.

2 Section 702(a) provides, in relevant part, as follows:

> This title shall not apply to... a religious corporation, association, educational institution, or society with respect to the employment of individuals of a particular religion to perform work connected with the carrying on by such corporation, association, educational institution, or society of its activities.

42 U.S.C. § 2000e-1(a). The § 702(a) exemption is not limited to the employment of persons of the same denomination as the employer. Rather, a religious organization may seek employees that are like-minded with respect to the religious observance the employer deems relevant to the organization's religious nature whether or not the employee is a co-religionist or is a member of the same denomination. *Cf.* Killinger v. Samford University, 113 F.3d 196, 199-200 (11th Cir. 1997); Hall v. Baptist Memorial Health Care Corp., 27 F. Supp. 2d 1029, 1033, 1036-40 (W.D. Tenn. 1998), *aff'd*, 215 F.3d 618 (6th Cir. 2000).

3 Lown v. Salvation Army, Inc., 393 F. Supp. 2d 223, 246-52 (S.D.N.Y. 2005).

4 *Id.* at 249-52.

5 *See generally* Carl H. Esbeck, Stanley W. Carlson-Thies & Ronald J. Sider, The Freedom of Faith-Based Organizations to Staff on a Religious Basis (Ctr. for Pub. Justice 2004). This monograph may be downloaded at http://www.cpjustice.org/faith-based_initiative/equal_treatement/religious_staffing/key_resources. Portions of this extended essay are taken from the monograph.

6 The three terms are defined at 31 U.S.C. §§ 6301-08.

7 Programs such as the Community Development Block Grant Act and Head Start include civil rights language requiring that no one will be denied program benefits or be discriminated against on account of race, color, national origin, sex, or religion. Community Development Block Grant Act, 42 U.S.C. § 5309; Head Start Act, 42 U.S.C. § 9849(a). Although this embedded language does not mention employment, a White House bulletin notes that "some older U.S. Supreme Court cases indicate that these statutes may also apply to employment decisions." White House Office of Faith-Based and Community Initiatives, *Protecting the Civil Rights and Religious Liberty of Faith-Based Organizations: Why Religious Hiring Rights Must Be Preserved* 5-6 (June 23, 2003), http://www.whitehouse.gov/government/fbci/booklet.pdf.

8 29 U.S.C. §§ 2801-2945. § 2938(a)(2) prohibits employment discrimination on several bases, including religion.

9 42 U.S.C. §§ 3711 to 3797ee-1. § 3789d(c)(1) prohibits employment discrimination on several bases, including religion.

10 *Id.* § 12635(c)(1)-(2).

11 42 U.S.C. §§ 2000bb to -4. RFRA was amended in September of 2001 as part of the bill that enacted the Religious Land Use and Institutionalized Persons Act, 42 U.S.C. §§ 2000cc to -5. As amended, RFRA is no longer applicable to the operations of state or local governments, only to the operation of federal law and its implementation by federal officials. 42 U.S.C. § 2000bb-3(a). This amendment was made necessary because the Supreme Court held that RFRA's application to burdens on religion as a result of state or local law was an unconstitutional use of congressional power under section 5 of the Fourteenth Amendment. *See* City of Boerne v. Flores, 521 U.S. 507 (1997).

12 RFRA reads in terms of protecting the rights of "persons." 42 U.S.C. § 2000bb-1(a). However, in the United States Code the term "persons" includes organizations, thereby including protection for faith-based organizations. *See* 1 U.S.C. § 1.

13 42 U.S.C. § 2000bb-1(a).

14 *Id.* § 2000bb(b), 2000bb-1, 2000bb-2(3)-(4).

15 *See Effect of the Religious Freedom Restoration Act on Faith-Based Applicants for Grants*, http://www.usdoj.gov/fbci/effect-rfra.pdf. *See also* Answer to Question 10 of *Frequently Asked Questions*, http://www.usdoj.gov/fbci/faq.html.

This determination by the Department of Justice was hardly a surprise. The ruling had long been presaged within the Bush Administration. The White House Office of Faith-Based and Community Initiatives advised faith-based organizations that, notwithstanding restrictive statutory language in some federal funding programs, faith-based organizations may resort to RFRA for protection. *See* White House Office of Faith-Based and Community Initiatives, *Protecting the Civil Rights and Religious Liberty of Faith-Based Organizations: Why Religious Hiring Rights Must Be Preserved* 4 n.3, 5 n.4 (June 23, 2003), http://www.whitehouse.gov/government/fbci/booklet.pdf. Further, the regulations for several of the Substance Abuse and Mental Health Services Administration's (SAMHSA) drug-abuse prevention and treatment programs looked to RFRA to nullify a statutory limitation on faith-based recipients of the funding. *See* "Charitable Choice Regulations Applicable to States Receiving Substance Abuse Prevention and Treatment Block Grants, Projects for Assistance in Transition From Homelessness Formula Grants, and to Public and Private Providers Receiving Discretionary Grant Funding From SAMHSA for the Provision of Substance Abuse Services Providing for Equal Treatment of SAMHSA Program Participants," 68 Fed. Reg. 56430, 56435 (Sept. 30, 2003) (codified at 42 C.F.R. pt. 54.6). The preamble to the final rule for the Department of Justice's Equal Treatment regulations noted that RFRA applies even though it is not specifically mentioned in the regulations. *See* "Participation in Justice Department Programs by Religious Organizations; Providing for Equal Treatment of All Justice Department Program Participants," 69 Fed. Reg. 2832, 2836 (Jan. 21, 2004). The preamble to the final rule for the Department of Health and Human Service's Equal Treatment regulations included a similar comment. *See* "Participation in Department of Health and Human Services Programs by Religious Organizations; Providing for Equal Treatment of All Department of Health and Human Services Program Participants," 69 Fed. Reg. 42586, 42591 (July 16, 2004).

16 *Effect of the Religious Freedom Restoration Act on Faith-Based Applicants for Grants* (Oct. 2007), http://www.usdoj.gov/fbci/effect-rfra.pdf. The posting is careful to note that religious providers may never discriminate on the basis of religion with respect to the ultimate beneficiaries of the program. *Id.* And, of course, RFRA addresses religious burdens only and thus does not permit employment discrimination on the basis of race, national origin, sex, age, or disability. *Id.*

17 *Id*. In one sense RFRA is case-specific, responding to each organization's sincerely held claim of a religious burden. But for faith-based organizations that staff on a religious basis, RFRA will presumptively grant relief from generally applicable employment laws prohibiting discrimination on the basis of religion. Because RFRA will grant relief almost without fail to faith-based organizations with sincerely held religious staffing practices (*see, infra* notes 36-41, and accompanying text), it is correct to suggest that there is a presumption that RFRA relieves faith-based organizations from the religious burden imposed by these program-embedded nondiscrimination clauses. As with any presumption, the government can inquire into the bona fides of the faith-based organization's claim and rebut the operation of RFRA by evidence of insincerity.

18 42 U.S.C. §§ 3711 to 3797ee-1. Sec. 3789d(c)(1) prohibits employment discrimination on several bases, including religion.

19 *Id*. § 2000bb-1(a).

20 White House Office of Faith-Based and Community Initiatives, *Protecting the Civil Rights and Religious Liberty of Faith-Based Organizations: Why Religious Hiring Rights Must Be Preserved* 3 (June 23, 2003), http://www.whitehouse.gov/government/fbci/booklet.pdf.

21 *See, e.g.*, Charitable Choice Provisions Applicable to the Temporary Assistance for Needy Families Program, 45 C.F.R. § 260.34(c)-(d) (2007).

22 Corp. of the Presiding Bishop v. Amos, 483 U.S. 327, 340 (1987) (Brennan, J., concurring in judgment) (upholding constitutionality of the § 702(a) exemption in Title VII for religious organizations staffing on a religious basis).

23 *Id*. at 342.

24 *Id*.

25 *Id*.

26 RFRA, of course, guards only against burdens on religious exercise. It does not deal with race. Opponents of the faith-based initiative argue that using RFRA to overcome embedded program restrictions on religious staffing will excuse racial discrimination rooted in religious belief. This has not been the case, as there are no reported RFRA cases where racial discrimination is excused under the guise of religion. And in all events, the Supreme Court has already held that the denial of benefits to a religious organization in the interest of eradicating racial discrimination is a compelling governmental interest. *See* Bob Jones Univ. v. United States, 461 U.S. 574, 602-04, 604 n.29 (1983). RFRA claims are overridden, of course, by compelling governmental interests. 42 U.S.C. § 2000bb-1(b).

27 *See* 42 U.S.C. § 2000bb-2(4) (incorporating by reference the definition of "religious exercise" in 42 U.S.C. § 2000cc-5(7)).

28 *See id.* § 2000bb-4 ("Granting government funding... shall not constitute a violation of this chapter.... [T]he term 'granting,' used with respect to government funding... does not include the denial of government funding....").

29 374 U.S. 398 (1963). *See also* S. Rep. No. 103-111, at 13 (1993), *as reprinted at* 1993 U.S.C.C.A.N. 1892 ("parties may challenge, under the Religious Freedom Restoration Act, the denial of benefits to themselves as in *Sherber[t]*"); *id*. at 15 ("the denial of [government] funding... may constitute a violation of the act, as was the case under the free exercise clause in *Sherbert v. Verner*").

30 RFRA states, as one of its purposes, "to restore the compelling interest test" of *Sherbert v. Verner*. 42 U.S.C. § 2000bb(b)(1). The denial of funding in *Sherbert* was slightly different from the denial of a social service grant to a faith-based organization. But RFRA was not drafted to restore the holding of a single case. *Sherbert* was illustrative of the problem, not the whole problem. The terms of RFRA read as general principles, with the object being the provision of a remedy for a variety of religious burdens—no matter how or where the burdens occur.

31 42 U.S.C. § 2000bb-2(4) (incorporating by reference the definition of "religious exercise" in 42 U.S.C. § 2000cc-5(7)).

32 In *Brusca v. Bd. of Educ.*, 405 U.S. 1050 (1972), *aff'g* 332 F. Supp. 275 (D.C. Mo. 1971), the Supreme Court affirmed that a state's provision of free public school education does not compel the state to provide an equal benefit to religious school parents. Luetkemeyer v. Kaufmann, 419 U.S. 888 (1974), *aff'g* 364 F. Supp. 376 (D.C. Mo. 1973), likewise affirmed that a state may choose to provide free bussing to government schools without providing a like benefit to religious schools. But *Brusca* and *Luetkemeyer* are inapposite to the situation here where the government has elected to involve private charities in the delivery of social services.

33 42 U.S.C. § 2000bb-1(a).

34 As part of the *prima facie* case, RFRA requires proof of a substantial burden on a claimant's religion. 42 U.S.C. § 2000bb-1(a) and (b).

35 *See, e.g.,* Jimmy Swaggart Ministries v. Bd. of Equalization of Cal., 493 U.S. 378, 391-92 (1990) (upholding uniform state levy of sales and use taxes on sale of material, including religious material).

36 The dollar amount, large or small, of any particular available grant is relevant but not essential to meeting RFRA's "substantial burden" requirement. A promise to comply with program-embedded employment nondiscrimination provisions is an essential criterion of grant eligibility. To fail to accommodate sincerely held religious employment practices is thus a categorical bar to a faith-based organization's eligibility for any such federal grant program. That unquestionably is a substantial burden or cruel choice, and the burden is uniquely religious rather than monetary. *See Sherbert*, 374 U.S. 398, 404, 406 (1963) (holding that an individual refusing to take a job entailing work on her Sabbath could not be put to the cruel choice of either forfeiting her unemployment benefits or violating her religious Sabbath).

A prohibition on religious staffing may also lead to a lawsuit alleging employment discrimination on the basis of religion. Such a lawsuit may be brought by the government or by a private litigant, the defense of which would be a substantial burden. *See Amos*, 483 U.S. at 336 (noting burden of defending religious organization sued for employment discrimination on basis of religion). RFRA regards any such lawsuit as a burden on religion. If the lawsuit is by a private claimant, RFRA still applies. RFRA applies not only to "all federal law" but also to "the implementation of that law." 42 U.S.C. § 2000bb-3(a). Every discrimination lawsuit by a private litigant is an "implementation" of a federal embedded nondiscrimination clause.

37 42 U.S.C. § 2000bb-1(b).

38 546 U.S. 418 (2006).

39 *Id*. at 430-31.

40 *Id*. at 432.

41 *Id*. at 434-37.

42 *Id*. at 430-31. *See also id*. at 428, 432-34.

43 *See* John D. Ashcroft, Statement on Charitable Choice, Proceedings and Debates of the 105th Cong., 2d Sess., 144 Cong. Rec. S12686, S12687 (Oct. 20, 1998) ("Demanding that religious ministries 'secularize' in order to qualify to be a government-funded provider of services hurts intended beneficiaries of social services, as it eliminates a fuller range of provider choices for the poor and needy, frustrating those beneficiaries with spiritual interests.").

44 *See, e.g.*, Widmar v. Vincent, 454 U.S. 263, 269 n.6, 271 n.9, 272 n.11 (1981) (stating that the principle of no-establishment is served when government avoids entanglement in religious events, practices, and doctrine); Walz v. Tax Comm'n, 397 U.S. 664, 674-76 (1970) (holding that no-establishment is served when government avoids excessive entanglement in the affairs of religious organizations).

45 483 U.S. 327, 332 n.9 (1987) (quoting Senator Ervin).

46 Authentic pluralism is when religious differences are openly acknowledged to make a difference. That is, religious particularity in a civil society that lives with these differences amicably while not denying that the particulars truly matter is authentic pluralism. On the other hand, *pseudo* pluralism is when religious people are expected to hide their religious differences in the public square so as not to offend others of a different persuasion.

47 Stephen V. Monsma, Putting Faith in Partnerships: Welfare-to-Work in Four Cities 62-63, 188-94 (U. of Mich. Press 2004).

Homeschooling Battle in California

*By Raymond J. Tittmann**

A constitutional crisis hit California this spring, although it was noticed more for its social than legal significance. On February 28, 2008, in the case *In re Rachel L.*, the Second District Court of Appeal in Los Angeles, California, found that the state's truancy statute prohibited homeschooling without a teaching credential for the grade taught: under California case law "parents *do not have* a constitutional right to home school their children."[1]

California's truancy statute obliges all children to attend public school unless (1) the child is enrolled in a private full-time day school, (2) the child is tutored by a person with a state teaching credential for the grade being taught, or (3) other limited exemptions apply.[2] The father of the homeschooled children argued that homeschooling was protected by the first exemption because he had enrolled his children in a private school that directed the homeschooling. But the court disagreed, finding that homeschooling was governed by the second exemption, and the parents therefore had to obtain a teaching credential for the grade being taught.

Newspapers in California reported "shock waves" rocking the homes of some 166,000 California homeschooled children.[3] Days later, the executive branch—both Governor Arnold Schwarzenegger and Superintendent of Public Instruction Jack O'Connell—announced its refusal to follow the decision of the appellate court. Superintendent O'Connell declared that Department of Education policy "will not change in any way as a result of this ruling."[4] He said that the Department policy would remain, as before, pro-homeschooling: "Parents *still have* the right to home school in our state."[5] The Department of Education blatantly repudiated the court's decision—even using the same language but replacing the phrase "do not" with "still."[6] Governor Schwarzenegger did not pull any punches either: "This outrageous ruling must be overturned by the courts and if the courts don't protect parents' rights then, as elected officials, we will."[7]

Like the Cuban Missile Crisis, this constitutional standoff was resolved in a mere two weeks when the instigator withdrew. On March 25, 2008, the court of appeal granted rehearing, in effect vacating the precedential value of the opinion.[8] Parties submitted briefs for the rehearing on April 28, 2008, and the court invited amicus parties to file briefs by May 19, 2008. However, like the Cuban Missile Crisis, the short-term resolution does not settle the long-term problem. The court is expected to issue another decision in the summer or fall, followed by an appeal to the California Supreme Court.

I. The Legality of Homeschooling Was Not Properly at Issue

The court's decision in *In re Rachel* was remarkable for its lack of judicial restraint, with respect to both the legal issue and the factual issues framed by the court. Judicial restraint requires courts to avoid unnecessary constitutional issues when other grounds are dispositive.[9] The father here argued that the proper legal issue was not the constitutionality of homeschooling at all, but rather the safety of the children.[10]

The father's argument is based on the procedural history. The case arose from a juvenile court proceeding in which attorneys appointed to represent the children argued that the court should intervene and order the children to school. The California Welfare & Institutions Code grants a court narrow authority to intervene in the parents' decisions concerning the education of their children; it is limited to situations where intervention is "necessary to protect the child."[11] Courts have interpreted this section to require, as a prerequisite to intervention, evidence establishing a "substantial risk" of "*serious physical harm or illness.*"[12] Thus, for example, one court declined to intervene when a transient parent's child consistently failed to appear at her enrolled school because the parent's conduct did not cause serious physical harm.[13]

Given this onerous standard, the juvenile court predictably ruled that homeschooling did not present a "substantial risk" of "*serious physical harm or illness*," and declined to order the children into school.[14] The proper standard on appeal was abuse of discretion. Consequently, to grant a writ ordering the children out of their home school, the court of appeal had to find that homeschooling unquestionably posed a risk of "*serious physical harm or illness*"—nothing more and nothing less. The court, however, did not decide or even consider the safety issue and instead broadly addressed the legality of homeschooling under the state's truancy statute, deciding at the same time both too much and too little.

In similar circumstances, the First District Court of Appeal recently reversed a decision by the family court for its failure to exercise judicial restraint. It found that the family court did not need to decide the constitutionality of the Indian Child Welfare Act because it failed to first apply the statutory provisions to determine whether a constitutional issue actually existed.[15]

Likewise here, the court should have applied the provisions of the Welfare & Institutions Code to determine if it had the authority to intervene. If the court lacked the statutory authority to intervene (i.e., if the children's safety was not threatened), the constitutional and statutory legality of homeschooling was irrelevant.

II. The Court Outlawed Homeschooling Practices of Non-Parties

The court also demonstrated a lack of judicial restraint by deciding factual situations not before it. The parents in *In re Rachel* homeschooled through a private school, but several other methods of compliance, implicating different statutory provisions, were not at issue and were improperly addressed by the court.

* *Raymond J. Tittmann is a board member of Catholics for the Common Good, dedicated to defending Catholic social doctrine in the public square, and a partner at the law firm of Carroll, Burdick & McDonough LLP. The views expressed here are his own.*

Public charter home schools (not at issue): California permits state-funded and state-run charter schools to establish home independent study programs. The state not only knows about this homeschooling, it sponsors, funds, and monitors the homeschooling pursuant to a statutory framework. The children in *In re Rachel* were not enrolled in such a charter school, and the parents therefore had no incentive to present the statutory framework authorizing public charter home schools. Nevertheless, the court called homeschooling through public charter schools illegal when the parent teacher is not credentialed precisely because "the parents present[ed] no authority" to justify such public charter schools, ignoring that the parents had no interest in presenting such authority.[16] Technically this finding is dicta and not enforceable, but it is nevertheless a dangerous statement that could—if it survives rehearing—be misunderstood to ban established state schools.

Home private schools (not at issue): The Home School Legal Defense Association (HSLDA) and several other home school organizations have long asserted that parents in California can declare their home a private school by filling out a Form R-4 affidavit.[17] California law does not prohibit private schools from operating on a residence, using parents as teachers, or enrolling children living in the residence. Accordingly, California law does not prohibit private home schools. The parents in *In re Rachel* did not submit an R-4 affidavit declaring their home a private school. Accordingly, the parents had no incentive to argue the mechanics and legality of the R-4 affidavit in this context, and the HSLDA and other affected non-parties had no opportunity to argue this issue. The court nevertheless rejected the argument that a home school could avoid credentialing requirements by calling itself a private school: "parents instructing their children at home" do not "come within the private full-time day school exemption."[18]

Private schools directing home school (only method at issue): Numerous private schools allow homeschooled students to enroll and then the school assists parents in the education of their children. Like the state-funded charter schools, private home schools may provide curricula, books, resources, grading, advice, tutoring, and consultation. The children in *In re Rachel* were enrolled in Sunland Christian School, a private school of this type. The court found that this method also "does not qualify for the private full-time day school… exemption[] from compulsory education in a public full-time day school."[19] However, the entities with the most interest in this issue, such as the Sunland Christian School and the numerous associations representing such schools, were not parties and therefore were unable to brief the issue.

The decision in *In re Rachel* therefore not only decided the wrong issue, it also outlawed practices without hearing from the parties whose conduct was being outlawed. The court's presumably granted rehearing to correct at least some of these procedural improprieties.

III. The Court Failed To Analyze the Statutory and Constitutional Issues Thoroughly

Critics of the *In re Rachel* decision disagree not only on procedural grounds but on substantive grounds as well. The court relied heavily on *People v. Turner*:[20]

> These provisions of the Education Code [prohibiting homeschooling, under the court's interpretation] were held to be constitutional in *People v. Turner* (1953) 121 Cal. App. 2d Supp. 861, and an appeal to the United States Supreme Court from that decision was dismissed for want of a substantial federal question in *Turner v. People of the State of California* (1954) 347 U.S. 972.[21]

The careless reader might think the decision in *In re Rachel* was, legally speaking, nothing new—homeschooling has been illegal for at least fifty years, under binding California authority implicitly endorsed by the United States Supreme Court. In fact, there is no such binding authority in California. *Turner* was decided by the appellate department of the *superior court*, which decides appeals of such matters as parking tickets and small claim disputes. It is not binding precedent.[22] Nor is the citation to the U.S. Supreme Court helpful; the Court devoted a mere sentence to the matter: "The appeal is dismissed for the want of a substantial federal question."[23] The court's decision finding homeschooling illegal and not constitutionally protected was not only shocking to homeschoolers, it broke new legal ground as well.

Though California had not specifically resolved the homeschooling issue before *In re Rachel*, federal and state constitutional law has generally recognized parents' "right of control" and "natural duty" concerning their children's education.[24] Given the constitutional rights implicated, the court should have construed the truancy statute to "avoid constitutional infirmities."[25] Here, the statute could have been construed to avoid compromising the parents' rights. The Education Code, strictly speaking, is silent as to whether homeschooling may qualify as a "private school" under Education Code sections 48222 and 33190. Those sections require "the owner or other head" of the private school to file an affidavit with the Superintendent of Public Instruction. Nothing in either section precludes application of the statute to homeschools or private schools that enroll homeschooled children.

The court followed *Turner* in finding that these sections addressing private schools were not "intended" to apply to home schools because the legislature must have intended the tutor exception (which, importantly, requires state teaching credentials) for homeschoolers.[26] Parents should not be allowed to escape the credentialing requirement by purporting to enroll their homeschooled children in a private school. The court's interpretation is plausible, but it invites rather than avoids a constitutional issue. The alternate interpretation, that home schools are legal under the private school exemption, is also reasonable and does not compromise fundamental parental rights.

After construing the Education Code to prohibit homeschooling without a state teaching credential, the court conducted a brief constitutional analysis. The court did not apply the "strict scrutiny" called for when parental rights of this sort are implicated.[27] In fact, the court did not even decide what sort of scrutiny should apply. The word "scrutiny" does not appear anywhere in the decision. The only case cited in the court's constitutional analysis—*Wisconsin v. Yoder*[28]—held thirty-six years ago that an Amish community *did* have the right to school their own children for religious reasons.[29] Constitutional law professors on the left and the right have criticized the decision for its abbreviated constitutional analysis, leaving out the last several decades of First Amendment jurisprudence. [30]

One can only assume that the case was briefed summarily. It is doubtful that the parents or the children (all represented by court-appointed counsel) had significant resources to brief the constitutional issues with any great detail. And, as noted, the groups with greater interest and resources, such as Sunland Christian and the Home School Legal Defense Association, were not parties to the case. Consequently, the court's decision to vacate and invite briefing by interested *amicus* parties does a great service to this important issue.

Thus, as the Cuban Missile Crisis did not end the Cold War, the court's grant of rehearing does not settle the legality of homeschooling. The issue remains.

Endnotes

1 *In re* Rachel L. (Jonathan L. and Mary Grace L. v. Superior Court), 160 Cal. App. 4th 624, 626 (2008) (emphasis added). The Court stated, "California courts have held that under provisions in the Education Code, parents do not have a constitutional right to home school their children."

2 Cal. Educ. Code § 48220 *et seq.*

3 *See, e.g.*, Bob Egelko et al., *Homeschoolers' Setback Sends Shock Waves Through State*, San Francisco Chron., Mar. 7, 2008, at A1 ("Shock Waves").

4 Press Release, Superintendent of Public Instruction Jack O'Connell, Schools Chief Jack O'Connell Issues Statement Regarding Home Schooling In California (Mar. 11, 2008).

5 *Id.*

6 The statement by Superintendent O'Connell is even more remarkable considering that his political constituents, teachers unions such as the California Teachers Association, support the court's ruling. *See* http://www.oconnell2006.org/endorsements.htm; Shock Waves at A1 ("The ruling was applauded by a director for the state's largest teachers union. 'We're happy,' said Lloyd Porter, who is on the California Teachers Association board of directors.").

7 Press Release, Gov. Arnold Schwarzenegger, Governor Schwarzenegger Issues Statement Regarding Court of Appeals Home Schooling Ruling (Mar. 7, 2008).

8 *See* Order (1) Granting Rehearing, (2) Denying Motion To Intervene but Permitting Amicus Briefing; (3) Setting Schedule for Additional Briefing; (4) Setting Tentative Date for Further Oral Argument; (5) Announcing Intent To Solicit Additional Amicus Curiae Briefing, filed Mar. 25, 2008.

9 *See, e.g.*, Lyng v. Northwest Indian Cemetary Protective Ass'n, 485 U.S. 439, 445-46 (1988) (principle of judicial restraint requires courts to avoid reaching constitutional questions in advance of the necessity of deciding them); Santa Clara County Local Transp. Auth. V. Guardino, 11 Cal. 4th 220, 230 [902 P.2d 225] (1995) (court will not decide constitutional questions where other grounds are available and dispositive of the issues).

10 Father's Brief on Rehearing, Apr. 28, 2008, at 4 (identifying the issue precisely as this: "whether the juvenile court abused its discretion in concluding that enrollment in school outside the home was not required to protect the children's safety.").

11 Welfare and Institutions Code § 361.

12 *In re* Janet T., 93 Cal. App. 4th 377, 388-89 (2001) (emphasis in original).

13 *Id.* at 382.

14 The decision of the juvenile court is apparently not a public record, and this version of the court's decision is based on statements in the Father's Brief on Rehearing, April 28, 2008, at 6-8.

15 Erika K. v. Brett D., 2008 WL 963408 (Cal. App. Apr. 10, 2008).

16 *In re* Rachel L., 73 Cal. Rptr. 3d at 84. The court stated, "the parents present no authority to the effect that a charter school can excuse the statutory requirement that tutors be credentialed if their students are to come within the tutor exemption to compulsory public school education."

17 *See, e.g.*, Letter from Debbie Schwarzer, Home School Association of California (HSC), Legal Committee, to California Homeschoolers, Sept. 8, 2002.

18 *In re* Rachel L., 73 Cal. Rptr. 3d at 82.

19 *Id.* at 84.

20 121 Cal. App. 2d Supp. 861 (1953).

21 *In re* Rachel L., 73 Cal. Rptr. 3d at 81

22 Worthington v. Unemployment Ins. Appeals Bd., 64 Cal. App. 3d 384, 389 (1976).

23 Turner v. People of the State of Calif., 347 U.S. 972 (1954).

24 Meyer v. Nebraska, 262 U.S. 390, 400 (1923). The Supreme Court has more recently affirmed the same principle: "[I]t cannot now be doubted that the Due Process Clause of the Fourteenth Amendment protects the fundamental right of parents to make decisions concerning the care, custody, and control of their children." Troxel v. Granville, 530 U.S. 57, 66 (2000). "Parents possess a profound interest in the custody of their children, which includes "the right to direct his activities and make decisions regarding his care and control, education health, and religion." Walker v. Superior Court, 47 Cal. 3d 112, 134 (1988).

25 People v. Whaley, 160 Cal. App. 4th 779, 793 (2008).

26 *In re* Rachel L., 73 Cal. Rptr. 3d at 82.

27 Herbst v. Swan, 102 Cal. App. 4th 813, 819 (2002) (applying strict scrutiny to statute that infringed on parents' right to control their child's upbringing).

28 406 U.S. 205, 214 (1972).

29 *In re* Rachel L., 73 Cal. Rptr. 3d at 85.

30 Professors Erwin Chemerinsky, Dean of U.C. Irvine Law School, and John Eastman, Dean of Chapman University Law School, both predicted reversal by the California Supreme Court and the U.S. Supreme Court, and noted the failure to address several recent decisions since *Yoder*. Dean Chemrinsky specifically added, "And I have read Justice Croskey's opinion. It's a very short opinion for such a major issue." http://hughhewitt.townhall.com/talkradio/transcripts/Transcript.aspx?ContentGuid=781d1498-961b-436f-ab39-09bba04dfd4c.

Telecommunications & Electronic Media

A La Carte Regulation of Pay TV: Good Intentions vs. Good Economics

*By Jeffrey Eisenach & Adam Thierer**

In his insightful works on political economy, Professor Thomas Sowell warns of the dangers of lawmakers allowing good intentions to trump good economics when crafting public policy. It is a theme which Friedrich Hayek,[1] Milton Friedman[2] and others articulated before him, of course, but Sowell has more fully developed this cautionary principle in books like *A Conflict of Visions* and *The Vision of the Anointed*. Sowell teaches us that noble intentions alone do not necessarily translate into sound public policy, and cautions against the hubris that leads policymakers to believe that they can easily improve on market outcomes. Even the best-intentioned policies can spawn unintended consequences, giving rise to still more regulatory interventions as policymakers seek to rectify past mistakes.

The Federal Communications Commission (FCC) would be wise to heed Sowell's advice in the ongoing debate over "a la carte" regulation of cable and satellite television networks and programming. The notion—giving consumers the right to pay for only the cable TV channels they want, without having to purchase a full bundle—is highly appealing on the surface, and well-intended advocates on both sides of the political divide, including the Consumer Federation of America's Mark Cooper and FCC Chairman Kevin Martin, are no doubt acting out of the best of intentions. But a closer look suggests that a la carte regulation would be a classic case of what we refer to as Sowell's Law of Wishful Thinking. Indeed, it would likely have *the exact opposite effects* of what its proponents intend, leaving consumers and families worse off than they are today.

I. Today's Pay TV Marketplace

On the face of it, the pay TV marketplace appears to be functioning efficiently. By any conceivable statistical measure, consumers today have access to more video outlets and options than at any time in history, and 86% of Americans subscribe to cable, satellite or telco-provided television services.

Certainly there is no shortage of programming. Indeed, the long-heralded "500-channel" cable universe is now a reality. The overall number of video programming channels available in America has skyrocketed, from just seventy channels in 1990 to 565 channels in 2006.[3] The resulting diversity "on the dial" has been astounding . There is hardly any human interest or hobby that is not covered by some video network. As the FCC concluded in its 2003 Media Ownership Proceeding, "We are moving to a system served by literally hundreds of networks serving all conceivable interests."[4] Exhibit 1 below shows the sheer diversity of programming on pay TV today.

Cable and Satellite TV Programming Options

News: CNN, Fox News, MSNBC, C-Span, C-Span 2, C-Span 3, BBC America

Sports: ESPN, ESPN News, ESPN Classics, Fox Sports, TNT, NBA TV, NFL Network, Golf Channel, Tennis Channel, Speed Channel, Outdoor Life Network, Fuel

Weather: The Weather Channel, Weatherscan

Home Renovation: Home & Garden Television, The Learning Channel, DIY

Educational: The History Channel, The Biography Channel (A&E), The Learning Channel, Discovery Channel, National Geographic Channel, Animal Planet

Travel: The Travel Channel, National Geographic Channel

Financial: CNNfn, CNBC, Bloomberg Television

Shopping: The Shopping Channel, Home Shopping Network, QVC

Female-oriented: WE, Oxygen, Lifetime Television, Lifetime Real Women, Showtime Women

Family / Children-oriented: Animal Planet, Anime Network, ABC Family, Black Family Channel, Boomerang, Cartoon Network, Discovery Kids, Disney Channel, Familyland Television Network, FUNimation, Hallmark Channel, Hallmark Movie Channel, HBO Family, KTV – Kids and Teens Television, Nickelodeon, Nick 2, Nick Toons, Noggin (2-5 years), The N Channel (9-14 years), PBS Kids Sprout, Showtime Family Zone, Starz! Kids & Family, Toon Disney, Varsity TV, WAM (movies for 8-16-year-olds)

African-American: BET, Black Starz! Black Family Channel

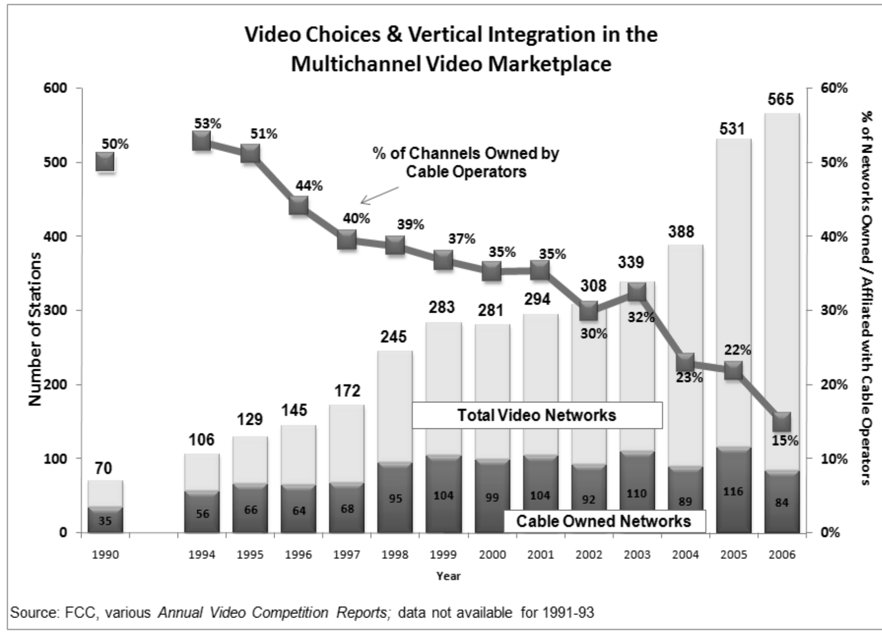
Foreign / Foreign Language: Telemundo (Spanish), Univision (Spanish), Deutsche Welle (German), BBC America (British), AIT: African Independent Television, TV Asia, ZEE-TV Asia (South Asia) ART: Arab Radio and Television, CCTV-4: China Central Television, The Filipino Channel (Philippines), Saigon Broadcasting Network (Vietnam), Channel One Russian Worldwide Network, The International Channel, HBO Latino, History Channel en Espanol

Religious: Trinity Broadcasting Network, The Church Channel (TBN), World Harvest Television, Eternal Word Television Network (EWTN), National Jewish Television, Worship Network

Music: MTV, MTV 2, MTV Jams, MTV Hits, VH1, VH1 Classic, VH1 Megahits, VH1 Soul, VH1 Country, Fuse, Country Music Television, Great American Country, Gospel Music Television Network

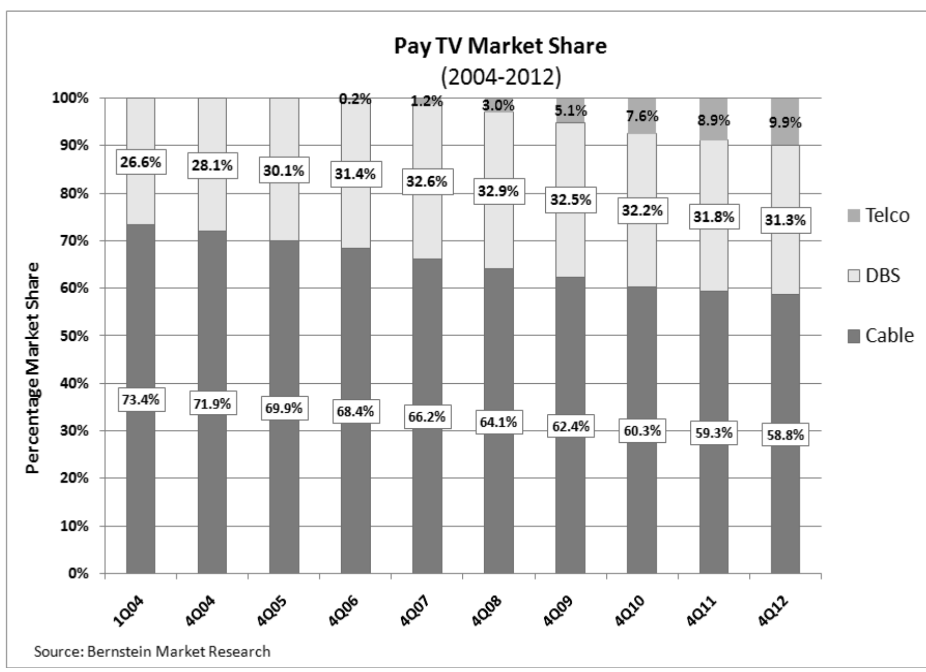
Movies: HBO, Showtime, Cinemax, Starz, Encore, The Movie Channel, Turner Classic Movies, AMC, IFC, Flix, Sundance, Bravo (Action, Westerns, Mystery, Love Stories, etc.)

Not only is programming diverse, so, too, is the universe of programmers. While the FCC and cable industry critics have often expressed concern about vertical integration between cable operators on the one hand and programmers on the other, the reality is that vertical integration in the video marketplace has plummeted. Since 1990, the number of cable-owned or affiliated channels has increased only slightly, while the number of independently-owned and operated video networks has exploded. Thus, as shown in Exhibit 2 (top), the percentage of cable channels owned and operated by cable operators has dropped from 50% in 1990 to just 14.9% today.



Source: FCC, various *Annual Video Competition Reports;* data not available for 1991-93

Competition is also growing on the program delivery front. While local cable operators were once monopolies, competition from satellite, cable overbuilders, and, most recently, telephone companies like AT&T and Verizon is cutting deeply into cable's market share. As shown in Exhibit 3 (bottom), competition in the program delivery market is increasing rapidly.



Source: Bernstein Market Research

Growing competition is not only offering consumers more choice, but also increased quality. Cable, satellite, and telephone companies have invested hundreds of billions of dollars in recent years to provide more channels, digital delivery, video-on-demand, and, most recently, high-definition. As the FCC itself concluded in its most recent report on the video programming marketplace, "competition in the delivery of video programming services has provided consumers with increased choice, better picture quality, and greater technological innovation."[5]

Given these results, what is the case for regulation? Proponents of an a la carte mandate suggest that they can improve on the market in two primary ways. First, since people would no longer be forced to pay for channels they do not watch, they would pay less for cable television. Second, since people could choose not to buy certain channels, they would no longer be forced to subsidize programming (particularly "racy" programming) of which they disapprove. Thus, an a la carte mandate is presented as both economic regulation, designed to reduce prices, and social regulation, designed to "clean up the airways."

II. A La Carte as Economic Regulation

The economic case for an a la carte mandate rests on two premises. First, it is argued, cable prices are rising faster than inflation, and government action is therefore appropriate to give consumers lower prices. Second, a la carte regulation would reduce the prices people pay for cable programming. Neither premise withstands scrutiny.

* *Jeffrey Eisenach is Chairman of Criterion Economics and Adam Thierer is a Senior Fellow with the Progress & Freedom Foundation.*

If cable prices were rising faster than inflation, there might well be a strong political (though not an economic) case for regulation. Indeed, much of the case for a la carte has been premised on this notion, with advocates pointing, for example, to the most recent FCC report on cable prices, which shows that the average price paid for the basic tier of cable channels rose 93% over the ten-year period between 1995 and 2005, compared with the consumer price index, which increased only 28%. But there are two problems with this data.

First, the FCC's data is three years old. More recent data from the Bureau of Labor Statistics shows that cable price increases have moderated and, indeed, that cable prices in 2007 and 2008 have actually risen less rapidly than the Consumer Price Index (CPI). Perhaps this change is due to increased competition, or to efficiencies associated with the increasing ability of cable companies to spread the fixed costs of their infrastructures over multiple services, such as broadband and telephony. Whatever the reason, cable price increases have slowed, and the political case for regulation has thus weakened.

The second problem with the price increase argument is more fundamental, and ultimately more important: the price increase figures cited by a la carte advocates fail to account for changes in quality, which can be captured by taking into account either the number of television channels included in the basic tier or the time people spend watching cable TV. When cable prices are looked at in terms of the price per channel, the real price of pay television has fallen throughout the past decade.[6] And because television viewing overall is increasing, and cable's share of viewing hours is also going up (relative to over-the-air TV), the price paid per viewing hour has actually declined *even in nominal* terms.[7] Thus, the basic underlying political argument for economic regulation of cable is, to be blunt, a lie: by the most accurate measure, cable prices are not only not rising faster than inflation, they are actually going down!

Most people would agree, of course, that rising prices—even if they were rising—would not by themselves constitute a basis for regulation. Rather, regulation should be considered only if a case can be established for market failure—in which case it might be possible, at least in theory, for regulation to improve on the market outcome and lead to lower prices in the long run. But a la carte advocates have failed to demonstrate that bundling constitutes a market failure of any sort.

Bundling is, course, pervasive throughout the economy, and while the economics of bundling are complex, economists universally agree that it is generally efficient and beneficial to consumers. Bundling improves economic efficiency in a variety of situations, including when there are economies of scope and scale. One particularly significant and relevant efficiency motivation, advanced many years ago by Nobel Prize winner George Stigler,[8] occurs when there are high fixed costs of production and consumers have differing preferences for various "flavors" of a product. A simple example illustrates the point.[9]

Suppose there are two cable TV channels, "sports" and "business," each of which costs $10 to produce. Suppose further that there are two consumers, one of whom is willing to pay $7 for the sports channel and $4 for the business channel, while the other is willing to pay only $4 for sports, but will pay $7 for business. If the two channels are offered separately, there is no price at which demand will be sufficient to cover cost: if each is offered for $10 (its cost), no one buys either channel; if each is offered at $7 and is purchased by one consumer, revenue is $7 and each channel loses $3; and, if each is offered at $4 and purchased by both consumers, revenue is $8, and each channel loses $2. In short, in an a la carte world, neither channel is produced.

If bundling is permitted, on the other hand, the two channels can be offered together for $10, and both consumers (each of whom values the two channels at a total of $11) will purchase. Revenues are now $20, covering the costs of both channels, and each consumer receives $1 in consumer surplus.

This argument is only one of several that explain why bundling of cable TV channels is economically efficient. Bundling also provides a means for cable channels to expand their distribution, thereby increasing advertising revenues (and defraying costs that would otherwise be passed on to consumers in the form of higher subscription fees); it allows consumers to sample cable channels, thereby reducing marketing costs; and, it reduces transactions costs by avoiding the need for cable operators to constantly add and subtract channels from individual consumers' feeds.[10]

Economists also recognize that bundling can, in certain (very limited) circumstances, be a sign of market power, and a la carte advocates have suggested this is the case in cable TV.[11] But they have failed miserably to prove their case. Indeed, as noted above, competition in pay TV is growing: virtually all consumers now have a choice among at least one cable TV provider and two satellite providers, and many can also choose to get their service from cable overbuilders or telephone companies. If a la carte was an economically efficient business model, we would expect to see at least some of the firms in a competitive market to offer it voluntarily, yet none have done so.

Back in 2004, when the FCC first considered (and rejected) a la carte regulation, a group of respected economists wrote to the agency's Media Bureau warning that the proposal would not achieve its purported objectives. Their conclusion: "(1) mandatory a la carte distribution would very likely raise overall prices; (2) consumers' viewing decisions would very likely be distorted and their ability to sample alternative networks and shows would very likely be suppressed; and (3) mandatory a la carte distribution would very likely harm new and niche networks, which would result in fewer viewing options for consumers."[12]

In short, when it comes to a la carte, the economics of cable TV are clear: Rather than reducing prices and increasing choice, as proponents hope, it would do precisely the opposite.

III. A La Carte as Social Regulation

A la carte proponents point to another supposed benefit of regulation: it could help "clean up" the character of pay TV. Many policymakers have expressed a desire to extend content controls to cable and satellite TV, but realize that *direct* efforts to regulate subscription-based media platforms would likely be held to be in violation of the First Amendment.[13] A la carte regulation has been pitched by some of those lawmakers as an *indirect* method of accomplishing that same objective.[14] But will it really work? It is highly unlikely, at least in the fashion many lawmakers and family groups hope for.

The reason is two-fold. First, as made clear above, a la carte regulation threatens the wonderful diversity of programming on television today. That also explains why a la carte proponents are wrong when they suggest that it would "clean up" pay TV and allow us to purchase just the "good stuff." The "good stuff" is not likely to survive in a world of mandatory a la carte regulation. Most family-focused/children's networks, female-oriented channels, and religious programmers oppose a la carte mandates for this reason. They understand that their programs attract only a small subset of the overall universe of viewers. If their networks are not bundled alongside other channels, they might disappear entirely. Colby May, director of the Faith and Family Broadcasting Coalition, which represents religious broadcasters, last year called a la carte regulation "a dagger aimed at the heart of religious broadcasting in America," and predicted that it would "decimate religious broadcasting and the wholesome, family-oriented programming carried on niche cable channels."[15]

Second, the channels that some lawmakers want driven off basic cable—MTV, F/X, Comedy Central, Spike, and so on—will likely continue to do just fine. They are all among the Top 20 networks on cable and satellite TV today and have a strong following on DVD and the Internet. Even under a new regulatory regime, people will still flock to these networks in fairly large numbers.

So the "choice" consumers will be left with in a world encumbered by a la carte regulation is one of fewer choices of television programming. If smaller, niche-oriented networks begin to disappear, lawmakers will be dismayed obviously, but they will be absolutely furious if the channels that they *really* wanted to see vanish end up surviving anyway.

Again, wishful thinking cannot change the basic rules of markets and economics. Policymakers might like to see "racy" programming disappear and "family" programming rule the "airwaves," but imposing a la carte regulation would likely have precisely the opposite effect.[16]

CONCLUSION

From a policy perspective, a la carte regulation is worse than a solution in search of a problem; it is a problem waiting to happen. As prices rose and programming became less diverse in the wake of an a la carte mandate, policymakers would find themselves besieged by consumers and interest groups demanding yet another "solution." Perhaps they would frankly admit error and reverse course, repealing the misguided policy they had so recently put in place. On the other hand, it is just possible that they would proffer still more regulatory solutions—price controls, for example—which would be politically attractive on the surface, but equally flawed in their underlying economics. No doubt the new rules would also be motivated by the best of intentions.

Endnotes

1 *See generally* F.A. Hayek, The Fatal Conceit, (1989).

2 *See generally* Milton Friedman, Capitalism and Freedom (1962).

3 "FCC Adopts 13th Annual Report to Congress on Video Competition and Notice of Inquiry for the 14th Annual Report," Fed. Commc'ns Comm'n, *Press Release,* Nov. 27, 2007, at 3, http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-278454A1.pdf.

4 Fed. Commc'ns Comm'n, *In the Matter of 2002 Biennial Regulatory Review – Review of the Commission's Broadcast Ownership Rules and Other Rules Adopted Pursuant to Section 202 of the Telecommunications Act of 1996*, FCC 03-127, June 2, 2003, at 48-49, http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-03-127A1.pdf.

5 *See* Fed. Commc'ns Comm'n, *In the Matter of Annual Assessment of the Status of Competition in the Market for the Delivery of Video Programming*, Twelfth Annual Report, MB Docket No. 05-255 (Mar. 3, 2006), at ¶5.

6 *See, e.g.,* Jeffrey A. Eisenach & Douglas A. Trueheart, "Retransmission Consent and Cable Television Prices," The CapAnalysis Group, LLC (March 2005).

7 *See* Steven S. Wildman, *Assessing Quality-Adjusted Changes in the Real Price of Cable Service*, Mich. St. Univ. (Sept. 10, 2003).

8 George J. Stigler, *A Note on Block Booking*, *in* The Organization of Industry (1968).

9 For a more complete discussion, see Jeffrey A. Eisenach & Richard Ludwick, *The FCC's Further Report on A La Carte Pricing of Cable Television*, The CapAnalysis Group, LLC (March 6, 2006).

10 For an excellent discussion of the economic efficiency rationales for cable bundling, see Federal Communications Commission, Report On the Packaging and Sale of Video Programming Services to the Public (December 2004) at 20-26.

11 For example, Chairman Martin has suggested that bundling might constitute an illegal "tying" arrangement. Ted Hearn, *Martin: Expanded Basic a Tying Arrangement*, Multichannel News, Sept. 14, 2006.

12 Joint Economists Letter to Kenneth W. Ferree, MB Docket 04-207, November 4, 2004.

13 *See* Adam Thierer, "Thinking Seriously About Cable and Satellite Censorship: An Informal Analysis of S. 616, The Rockefeller-Hutchison Bill," Progress & Freedom Foundation *Progress on Point* no. 12.5, Apr. 2005, http://www.pff.org/issues-pubs/pops/pop12.6cablecensorship.pdf; Robert Corn-Revere, "Can Broadcast Indecency Regulations Be Extended to Cable Television and Satellite Radio?" Progress & Freedom Foundation *Progress on Point* 12.8 (Progress and Freedom Foundation May 2005), http://www.pff.org/issues-pubs/pops/pop12.8indecency.pdf.

14 *See, e.g.,* Chairman Martin's remarks at "Open Forum on Decency," in the Senate Commerce Committee in November 2005 (http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-262484A1.pdf) as well as his comments on Capitol Hill on June 14, 2007 applauding the introduction of an a la carte legislative proposal (http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-274169A1.pdf).

15 "Religious Broadcasters Oppose FCC Chairman's Call for Pay-Per-Channel Legislation," *Press Release*, Faith and Family Broadcasting Coalition, April 17, 2007, http://www.ncta.com/DocumentBinary.aspx?id=575. The late the Rev. Jerry Falwell also opposed a la carte regulation for the same reason, saying "Though well-intentioned, the fact is that a la carte would threaten the very existence of religious broadcasting and the vital ministry conducted over the television airwaves." Quoted in Ted Hearn, "Falwell's 11th Commandment: No a La Carte," *Multichannel News*, Nov. 17, 2004, http://www.multichannel.com/article/CA481169.html.

16 The better approach would be to encourage the continued development of even more family-friendly and educational fare and then let parents simply block those channels they do not desire with the parental control tools available in every set-top box today. That's a win-win strategy that demands no further regulation of pay TV markets while ensuring parents access to an expanding array of enriching programming for themselves and their children. *See* Adam Thierer, "Parental Controls and Online Child Protection: A Survey of Tools & Methods," Progress & Freedom Foundation *Special Report*, Version 3.0, Spring 2008, http://www.pff.org/parentalcontrols.

NARUSCA v. FCC: Federalism, Line Item Billing, and the Future of Wireless Consumer Protection

*By Seth Cooper**

States can regulate line items in customer billing for cellular wireless services. Thus spoke the U.S. Supreme Court, it seems, when it refused to hear *National Association of State Utility Consumer Advocates v. F.C.C.* The denial of certiorari left standing a Eleventh Circuit Court of Appeals decision that state truth-in-billing rules for wireless are *not* preempted under federal law. Here follows a look back at the Eleventh Circuit's ruling, as well as a short take on the past, present, and possible future of wireless consumer protection regulation.

Congressional Wireless Policy and the FCC's Truth-in-Billing Proceeding

The Communications Act of 1934 vested the Federal Communications Commission with authority to regulate radio frequencies used in cellular wireless services.[1] In 1993, Congress amended the Act to grant the federal government the exclusive authority to the "rates charged" and the "entry" of wireless carriers.[2] In significant respects, this cordoned off wireless carriers from state regulation. Under the amendment, however, states were permitted to continue regulating "other terms and conditions" of wireless services.[3]

In May, 1999, the Commission adopted Truth-in-Billing Rules "to ensure that consumers are provided with basic information they need to make informed choices in a competitive telecommunications marketplace, while at the same time protecting themselves from unscrupulous competitors."[4] The Commission applied those rules to wireline services, but exempted wireless providers from several such rules.

The State Consumer Advocates later petitioned the Commission for a declaratory ruling that prohibits wireless providers "from imposing any separate line item or surcharge on a customer's bill that was not mandated or authorized by federal, state, or local law."[5] A line item is "a discrete charge identified separately on an end user's bill."[6] The State Consumer Advocates argued that those line items do not allow customers to accurately assess what they are being billed for or whether they are being billed for government-mandated taxes and fees.

In March, 2005, the Commission issued its "Second Report and Order," amending and clarifying the application of truth-in-billing rules to wireless providers. It concluded that wireless providers were no longer exempt from requirements that billing descriptions be "brief, clear, non-misleading and in plain language."[7] The Commission also issued a "Declaratory Ruling," in which it denied the Consumer Advocates petition and declared state laws requiring or prohibiting use of line items on bills for wireless services preempted by Section 332(c)(3)(A) of the Communications Act of 1934, as amended.[8]

The Commission's preemption of state regulations for wireless line-item billing "rates" included "rate levels," "rate structures," and "rate elements."[9] It concluded that line items are "rate elements" and that state regulations prohibiting or requiring line items directly affect how wireless providers structure their rates. On the other hand, the Commission also concluded that state taxes, state universal service support charges, and other state regulations only have an "indirect effect… on a company's behavior."[10] The Commission's preemption ruling cited "the pro-competitive, deregulatory framework [for wireless providers] prescribed by Congress" in 1993.[11] According to the Commission, different state laws regulating line items would result in a variety of conflicting rules.

In preempting state regulations for wireless line item billing, the Commission left undisturbed state authority to impose taxes on wireless services, assess state universal service support charges, and enact other disclosure laws. However, it also requested comments about the role of states in regulating truth-in-billing issues and whether federal law preempted other state regulations of billing practices.

National Association of State Utility Consumer Advocates v. F.C.C.

The State Consumer Advocates, joined by the National Association of Rate Utility Commissioners (NARUC), petitioned for legal review of the Commission's ruling. Wireless providers Sprint Nextel and Cingular Wireless were granted intervenor status in support of the Commission.

In *National Association of State Utility Consumer Advocates v. F.C.C.* ("NASUCA v. FCC"), a unanimous panel of the Eleventh Circuit reversed the Commission's preemption ruling.[12] The court's opinion was penned by Judge William Pryor. A federalism-minded jurist, Judge Pryor's analysis of the Commission's preemption order began with citation of the Supremacy Clause of Article VI of the U.S. Constitution:[13]

> This Constitution, and the Laws of the United States which shall be made in Pursuance thereof; and all Treaties made, or which shall be made, under the authority of the United States, shall be the supreme Law of the land; and the Judges in every State shall be bound thereby, any Thing in the Constitution or Laws of any State to the Contrary notwithstanding.

The basics of modern preemption doctrine followed.[14] In the context of administrative law, federal agencies acting within their scope of congressionally delegated authority may preempt state regulation. Agencies entrusted with discretionary powers must not exceed their statutory authority or act arbitrarily.

Judge Pryor cited case authorities for a clear-statement rule of federal preemption: "the assumption that the historic police powers of the states are not superseded by federal law unless preemption is the clear and manifest purpose of Congress."[15] "[T]his presumption, he wrote "guides our understanding of the statutory language that preserves the power of the States to regulate 'other terms and conditions.'"[16]

Judge Pryor cited the Commission's finding that "'Congress did not specifically define 'rates,' 'entry,' or other key terms in section 332(c)(3)(A), but explained that 'rate regulation

* *Seth Cooper is the Director of the Telecommunications and Information Technology Task Force at the American Legislative Exchange Council in Washington, D.C., the nation's largest nonpartisan, individual membership organization of state legislators.*

extends to regulation of 'rate levels' and 'rate structures' for' wireless service providers.[17] But the court took exception to the Commission's rationale.

In the panel's view, "[t]he language of section 332(c)(3)(A) unambiguously preserved the ability of the States to regulate the use of line items in cellular wireless bills."[18] This determination was made through an examination of the defined terms of the Communications Act, as well as the meaning those defined terms and common dictionary definitions give to the undefined terms of the Act. Deeming a "rate" to be an amount of charge or payment, Judge Pryor wrote that "[t]he prohibition or requirement of a line item affects the presentation of the charge on the user's bill, but it does not affect the amount that a user is charged for service."[19] Whereas states may regulate billing practices of wireless providers, but not the amounts charged to consumers, "the presentation of line items on a bill is not a 'charge or payment for service'… it is an 'other term or condition' regulable by the states."[20]

Moreover, the panel found fault with the Commission for failing to follow the definition of "rates" that it relied upon in its previous rulings. In prior proceedings, the Commission defined "rates" as an "amount of payment or charge based on some other amount,"[21] and defined "rates charged" as prohibiting states from "prescribing, setting or fixing rates" of wireless providers.[22]

In seeking to preempt state line item billing requirements for its alleged effect upon rates, the Commission's stance was complicated by its own decisions upholding state universal service charges on wireless customers as an "other term or condition."[23] The Commission argued that line items had a "direct effect" on rates, whereas universal service charges only have an "indirect effect" on rates. The panel flatly rejected this argument as unavailing and without logical distinction.

The panel also concluded that the legislative history of the 1993 amendments to the Communications Act "shows that Congress intended to leave the authority to regulate line items with the states." If line items were a matter of "rates," concluded the panel, the Commission could preempt almost any form of state regulation of wireless service. Wrote Judge Pryor, "[t]he failure of the Commission to delineate the proper scope of rate regulation allows the Commission indefinitely to expand its authority without regard to the mandate by Congress that 'other terms and conditions' remain the realm of state regulation." In so concluding, the panel vacated the Commission's preemption ruling, remanding the case to the Commission.

Sister Circuit Case Comparison

The Eleventh Circuit panel's opinion in *NASUCA v. FCC* did not include any close examination of related decisions from other circuit courts of appeal. But its ruling against any easy preemptive presumptions for state laws relating in some way to wireless rates under 47 U.S.C. § 332(c)(3)(A) nonetheless finds some consistency in circuit court case law.

For instance, *Fedor v. Cingular Wireless Corporation* considered an argument to "interpret the preemption provision as covering any claim that touches on the rates charged in any manner."[24] The Seventh Circuit rejected that argument, concluding that position "overstates the scope of the preemption and in fact is a position that has been repeatedly rejected by the Courts and the FCC."[25] The Seventh Circuit panel reviewed prior rulings by the FCC concerning wireless regulations and preemption—including rulings also analyzed by the Eleventh Circuit in *NASUCA v. FCC*. "Those decisions," ruled the Seventh Circuit, "reject the argument that any claims related to the billing amount are automatically preempted under section 332.[26] Instead, the Seventh Circuit concluded that courts must "examine whether the claims require the state court to assess the reasonableness of the rates charged, or impact of market entry."[27]

In *Fedor*, the Seventh Circuit held that state law claims in contract and under state consumer protection laws for a wireless providers' improper attribution of calls and charges did *not* address rates themselves, but only the conduct of a provider in failing to adhere to those rates. Ultimately, the state law claims at issue in *Fedor* were not preempted but were "preserved for the states under §332 as the 'terms and conditions' of commercial mobile services."[28]

The Future of Line Item Billing Regulation

The U.S. Supreme Court declined to hear, and thereby upheld the Eleventh Circuit panel's decision, in January.[29] (Defendant intervenors Sprint Nextel filed the petition with the Supreme Court, as the FCC declined to pursue further litigation.) The FCC is therefore now forbidden from preempting the states on line item billing requirements for wireless carriers as "other terms and conditions" under Section 332(c)(3)(A). But the future suggests the possibility of regulatory reform in consumer protection—including line item billing—for wireless customers.

At present, state regulation of consumer protection in wireless services coincides with private self-regulation. In the past few years, wireless carriers have embraced some self-policing efforts. For instance, in 2003 the wireless industry adopted the "CTIA Consumer Code for Wireless Service."[30] Established through the industry's primary trade association, the CTIA Consumer Code is a ten-point set of best practices that its members agreed to for marketing services and billing customers. Also, in 2004, several wireless carriers entered into the "Assurance of Voluntary Compliance" with thirty-three state attorneys general.[31] The agreement set out nationwide consumer protection standards that wireless carriers agreed to follow.

However, the wireless industry has recently become more vocal in articulating its own vision for a National Regulatory Framework for Wireless. CTIA President Steve Largent has urged Congress to "close the 'other terms and conditions loophole,'" and establish "a clear, regulatory framework for all wireless consumers in all states."[32] According to Largent, states should continue to regulate wireless just like any other industry through its generally applicable state consumer protection laws.[33] But specific requirements for wireless carrier consumer protection would be federalized.

Moreover, legislation in the 110th Congress proposes express rulemaking authority for the FCC to address billing requirements for wireless carriers. For instance, Senator Amy Klobuchar has introduced the Cell Phone Consumer Empowerment Act of 2007.[34] Among other things, Sec. 5(f) of the bill requires the Commission to initiate proceedings under the Communications Act to establish regulations for line item billing. The Klobuchar bill includes a section expressly preempting state laws inconsistent with those of the legislation, except for "any State laws that provide additional protection to subscribers of wireless telephone service."[35] In addition, Senator Mark Pryor has introduced the Uniform Wireless Consumer Protection Act.[36] The bill requires the Commission to adopt consumer protection regulations for wireless customers within one year of the legislation's enactment. But Pryor's bill does not contain an express preemption section.

More recently, House Representative Ed Markey has circulated a draft bill titled the Wireless Consumer Protection and Community Broadband Empowerment Act of 2008.[37] This draft bill includes a section requiring the Commission to issue line item billing regulatory requirements for wireless carriers. But Markey's draft bill also does not include any express preemption section. These legislative proposals have not gone uncriticized.[38]

CONCLUSION

In sum, the Supreme Court's denial of certiorari in *NASUCA v. FCC* appears to bring some finality to the issue of whether states can regulate line items in wireless service. The Eleventh Circuit opinion in the case is a straightforward federal preemption ruling. States are not preempted from adopting line item requirements, and any action undertaken by the FCC on remand from the Eleventh Circuit must be made in that light. The judiciary has expounded on Congress's purposes in the 1993 amendments to the Communications Act. Whether Congress will take a different path through future legislation remains to be seen.

Endnotes

1 47 U.S.C. § 303.

2 *Id.* § 332(c)(3)(A).

3 *Id.*

4 *In re* Truth-in-Billing and Billing Format, 14 F.C.C.R. 7492 (1999).

5 *See In re* Truth-in-Billing and Billing Format, Nat'l Ass'n of State Util. Consumer Advocates' Petition for Declaratory Ruling Regarding Truth-in-Billing ("Second Report and Order"), 20 F.C.C.R. 6448, at 6449 (2005).

6 *Second Report and Order*, 20 F.C.C.R. at 6462.

7 *Id.* at 6454-6458.

8 *Id.* at 6458-6467.

9 *Id.* at 6456. *See also* 47 C.F.R. § 64.2401(b).

10 *Id.* at 6463.

11 *Id.* at 6466.

12 Nat'l Ass'n of State Util. Consumer Advocates v. F.C.C., 457 F.3d 1238 (11th Cir. 2006), *modifying opinion on denial of panel reh'g*, 468 F.3d 1272 (11th Cir. 2006), *reh'g en banc denied* (Nov. 29, 2006).

13 *See, e.g.*, William H. Pryor, *Madison's Double Security: In Defense of Federalism, the Separation of Powers, and the Rehnquist Court*, 53 Ala. L. Rev. 1167 (2002); William H. Pryor Jr., *The Demand for Clarity: Federalism, Statutory Construction, and the 2000 Term*, 32 Cumb. L. Rev. 361 (2002); William H. Pryor, "Fighting for Federalism," Remarks Before the Federalist Society (Mar. 28, 2001), *available at* http://www.fed-soc.org/chapters/id.83/default.asp (May 5, 2008).

14 To briefly summarize, federal preemption of state law is typically analyzed through a three-part taxonomy: (1) express preemption of state laws contained in federal statutes; (2) conflict preemption when federal and state laws cannot be reconciled; and (3) field preemption where Congress appears to have occupied the whole scope of a subject area and mutual compliance with state and federal law is practically impossible. *See* Wardair Canada, Inc. v. Fla. Dept. of Rev., 477 U.S. 1, 6 (1986). Congressional purpose is central of preemption analysis. La. Pub. Serv. Comm'n v. F.C.C., 476 U.S. 355, 369 (1986). For a more detailed background discussion of federal preemption doctrine, *see* David E. Engdahl, Constitutional Federalism (2d ed) (1986) 74-92, 332-56.

15 *Nat'l Ass'n of State Util.Consumer Advocates,* 457 F.3d at 1252.

16 *Id.* (cite omitted).

17 *Second Report and Order*, 20 F.C.C.R. at 6462-63 (cite omitted).

18 *Nat'l Ass'n of State Util. Consumer Advocates,* 457 F.3d at 1254.

19 *Id.*

20 *Id.*

21 *Id*. at 1254-55 (citing *In re* Sw. Bell Sys., Inc., 14 F.C.C.R. 19898, 19901 (1999)).

22 *Id.* (citing Cellular Telecomms. Indus. Ass'n v. FCC, 168 F.3d 1332, 1336 (D.C.Cir.1999) (quoting *In re* Pittencrieff Commc'ns, Inc. 13 F.C.C.R. 1735, 1745 (1997)).

23 *See Pittencrieff*, 13 F.C.C.R. at 1742, *aff'd sub nom. Cellular Telecomms. Indus.*, 168 F.3d at 1332 (cited in NARUSCA v. FCC, 457 F.3d at 1255-1256.

24 *Id*. at 1072.

25 *Id.*

26 *Fedor*, 355 F.3d at 1072-1073 (citing *In re* Sw. Bell Mobile Sys., Inc., 14 F.C.C.R. 19898 (1999) for the proposition that "state law claims stemming from contract or consumer fraud laws governing disclosure of rates or rate practices are not generally preempted under §332," and that the FCC held such claims fall within the 'other terms and conditions' regulable by the states); *Id.* (citing *In re* Wireless Consumers Alliance, Inc., 15 F.C.C.R. 17021 (2000) as holding that damage awards against commercial mobile service providers based on state court tort or contract claims are not preempted by §332 but that the FCC holds "such claims are generally preempted only where they involve the court in rulemaking").

27 *Id.* at 1074.

28 *Id.*

29 Sprint Nextel Corp. v. Nat'l Ass'n of State Util. Consumer Advocates, 128 S.Ct. 1119 (2008).

30 http://files.ctia.org/pdf/The_Code.pdf (May 5, 2008).

31 http://www.nasuca.org/CINGULAR%20AVC%20FINAL%20VERSION.pdf (May 5, 2008).

32 http://files.ctia.org/pdf/Testimony_Largent_WirelessConsumerProtection_2_27_08.pdf (May 5, 2008).

33 *Id* ("States should exercise their role in consumer protection to the same extend they do for competitive industries, no more and no less, by enforcing generally applicable consumer protection laws, but *not* through the promulgation of wireless-specific economic regulations").

34 S. 2033, 110th Congress, 1st Session (introduced Sept. 07, 2007), *available at* http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_bills&docid=f:s2033is.txt.pdf (May 5, 2008).

35 *Id.*, § 12.

36 S. 2171, 110th Congress, 1st Sess. (introduced Oct. 17, 2007), *available at* http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_bills&docid=f:s2171is.txt.pdf (May 5, 2008).

37 H.R. _____ [Staff Discussion Draft], 110th Congress, 2d Session (February 15, 2008), *available at* http://markey.house.gov/docs/telecomm/draft_wireless_legislation.pdf (May 5, 2008).

38 *See, e.g.*, *supra* note 32.

Book Reviews

How Judges Think

By Richard A. Posner

Constitutional Conscience

By H. Jefferson Powell

*Reviewed by Paul Horwitz**

What do we want of our judges? And what can we reasonably expect from them? However much these questions bedevil us, we rarely ask them so directly. The first question has featured prominently in every contested federal judicial nomination and presidential campaign since Robert Bork, but the public and politicians generally neglect the second question. Constitutional theorists often act as if the role and limitations of judges are a mere nuisance; surely, they will suggest *sotto voce*, the best judge is the one who most closely tracks my own thinking—institutional constraints be damned.

Happily, two recent developments in legal scholarship have led some to consider these questions more closely and more clearly. The ever-widening gyre of law and economics, and its cousin, behavioral law and economics, has led an increasing number of scholars to put the judge, with all his motivations, incentives, and cognitive limitations, in the spotlight. And a small but promising band of legal scholars, influenced by the flourishing study of virtue ethics elsewhere in the academy, has begun to consider the role of virtue in judging.

These contrasting approaches are on vivid display in two rich new books. H. Jefferson Powell's *Constitutional Conscience: The Moral Dimension of Judicial Decision* takes a page from virtue ethics in offering an expansive and ruminative vision of the ethical virtues and vices that characterize the judge in a constitutional case. Judge Richard A. Posner, in *How Judges Think*, brings his economist's toolkit (supplemented, as usual, by a variety of disciplinary adjuncts and a bracing dose of pragmatism) to many of the same questions—although his object is somewhat different from Powell's and his conclusions are strikingly so.

Powell begins his exploration of the role and duty of the constitutional judge in territory that has been oft-explored but still leaves much to be discovered and discussed: John Marshall's opinion in *Marbury v. Madison*. In mustering his arguments for judicial review, Marshall asked of the Constitution:

> Why otherwise does it direct the judges to take an oath to support it? ... How immoral to impose it on them, if they were to be used as the instruments, and the knowing instruments, for violating what they swear to support! ...
>
> Why does a judge swear to discharge his duties agreeably to the constitution of the United States, if that constitution forms no rule for his government? ... If such be the real state of things, this is worse than solemn mockery. To prescribe, or to take this oath, becomes equally a crime.

This passage has been criticized and sometimes mocked almost since its inception. In his famous dissent in *Eakin v. Raub* (1825), Judge Gibson wrote that the oath is "taken indiscriminately by every officer of the government, and is designed rather as a test of the political principles of the man, than to bind the officer in the discharge of his duty." More recently, Judge Posner, who serves as a foil in Powell's book, called the oath "a loyalty oath rather than a direction concerning judicial discretion."

Thankfully, and quite rightly, Powell—reaching through the mists of history to a time in which oaths had far-reaching consequences based on a wider sense of honor and virtue at large in the society—sees something more to it than that. Marshall's appeal to the oath, he writes, suggests a "juxtaposition of the judiciary's governmental role and the judge's personal conscience." Thus, the practice of judicial review is related not only to the broader constitutional structure that demands it but "flows as well from the judge's individual obligations as a moral actor." The inescapable aspect of moral obligation at the center of the judge's duties thus "implies a closer connection than is sometimes acknowledged between how we understand constitutional law and how individual judges understand the moral circumstances in which they carry out their duties."

From this departure point, Powell derives two central themes. The first is that it should be possible to derive a set of standards for evaluating the work of the judge or justice in moral or ethical terms. The second is that we can understand constitutional law from within this ethical perspective as an attempt to establish a fair process of resolving inevitable political and social conflicts without expecting "consensus or even broad agreement on many issues." What we think of as the actual practice of constitutional law—the "game" whose rules, Justice Holmes famously observed, one must "play by," rather than simply "doing justice"—is the suite of acceptable and traditionally well-worn forms of arguments which constitute that practice. These practices involve much that is careful, precise, and "technical" about reading the Constitution; but even a good-faith attempt to reach a constitutional ruling on this basis ultimately may drive us "beyond the realm of professional expertise and algorithmic reasoning" into "a sphere in which intellectual and moral integrity are essential." On this basis, he again presses his central point: the elements of good and bad faith that we can derive from an examination of sound constitutional judging simultaneously make evident that constitutional judges enter "the realm of moral obligation to which Marshall appealed in his discussion of the judicial oath," and provide us with a metric to see how well or poorly they traverse that territory.

What are the virtues that constitute a moral constitutional judge, in Powell's view? It is not unfair to say that they are few, and fairly abstract—although Powell is a skilled exegete and makes the most of them. The first is good faith: the belief that the Constitution has an intelligible meaning, and the will to discern and honor that meaning as best one can. Second, given the open spaces in that document, judges must display candor, by which Powell means a willingness to fully and transparently express in words the judge's sincere effort to grapple with the meaning of the Constitution, and integrity: "the virtue

* *Paul Horwitz is an Associate Professor of Law at the University of Alabama.*

of seeking in any given situation that interpretation of the Constitution that honestly seems to the interpreter the most plausible resolution of the issues in the light of the text and constitutional tradition." Powell's description of the remaining virtues is often more substantive than simply procedural or even ethical as such. He describes humility as being not about the judge's own frailties, but about the judge's willingness to accept that the Constitution leaves some divisive issues to be settled by the political process. Moreover, he names "acquiescence" as a constitutional virtue—a judge's willingness to accede to long-standing precedents of constitutional law and practice even if the judge might not reach those conclusions independently. Without these virtues, Powell bluntly concludes, "American constitutionalism is a fraud."

Powell concedes, as he must after such a stirring, but distinctly romantic, view of the judge's task and our own shared duty to "live out in the political and moral life of the Republic at large the virtues which the Constitution expects of its official interpreters," that his ideals "can be accused of fantasy, a failure to see that the political enjoys priority in a much harsher sense than I have conceded, that there is not and cannot be anything other than the agonistic struggle of political preferences."

That is as good a place as any to say: "Enter Richard Posner." Powell would drape the figure of Lady Justice in new and glorious robes, albeit robes of ancient design. Given his druthers, Posner would criticize the cut of Justice's robes, scoff that they are not warm enough to have any useful function, and digress to note that the taboo against nudity is itself a historically contingent and only locally applicable social norm. (Seriously. Doubters may consult the index entries on "nudity" in Posner's *Sex and Reason*.)

Posner begins by doubting that most judges themselves are especially reliable authorities on the role of the judge. They are too apt to spout "the loftiest Law Day rhetoric" and be "cagey, even coy, in discussing what they do. They tend to parrot an official line about the judicial process... , and often to believe it, though it does not describe their actual practices." Like their brethren in the wider legal profession, they have developed "a mystique" about judging "that exaggerates not only the professional's skills but also his disinterest. Judges have been doing this for thousands of years and have become quite good at it."

Posner will have none of this. Judges are not, he says, "moral or intellectual giants (alas), prophets, oracles, mouthpieces, or calculating machines. They are all-too-human workers, responding as other workers do to the conditions of the labor market in which they work." They are not "legalists" in a strict sense, working at formal proofs like so many logicians at a blackboard. Belief in legalism as the solution to the judicial puzzle is "the falsest of false dawns." Rather, Posner says, judges are "political," a loaded term by which he simply means that their decisions on divisive moral issues that "cannot be resolved by expert analysis, let alone by conventional reasoning," perforce must be influenced by a range of factors including ideology, background, personal preference, and so on. They are "occasional legislators" who use these and other factors to reach decisions in the open spaces afforded them by statutes and by a 200-year-old Constitution of broad reach and limited clarity.

Above all, in his view, they are "*constrained* pragmatists": "rule pragmatists" who decide cases with regard to their consequences, with all the potentially free-floating policy considerations that entails, but who are subject to a host of internal and external constraints on their decision-making. A constrained pragmatist judge may well decide cases in "legalist" fashion, but she will do so because legalism can have systemic beneficial consequences, not because legalism is compelled in some deeper sense. Like Holmes, Posner's constrained pragmatist "must play by the rules of the judicial game, just like other judges."

With barely concealed glee, Posner unapologetically warns readers that they will have to "brace themselves" for such analytic jargon as Bayesian decision theory, "reversal aversion," "utility function," and "agency costs." Although judges might not think in those terms, he emphasizes that "we must consider what judges *want*. I think they want the same basic goods that other people want, such as income, power, reputation, respect, self-respect, and leisure." Indeed, much of his book is given over to a consideration of the elements that make up the "judicial utility function"—a congeries of personal, psychological, broadly economic, and other factors that influence both the supply and demand curve for judges and the way they work.

He does not disdain the kinds of legal craft constraints that surely form a large part of Powell's picture of the judge. These craft considerations are an important part of the judge's makeup and constitute a significant part of his pleasure in the job at hand. Neither, though, does Posner romanticize legal craft or believe that it offers much clarity or direction in difficult cases. The law student, he writes, must learn all the conventional techniques of the lawyer—"en route to transcending them. But transcend them (or at least begin the process of transcending them) he must." Posner does not mock reason, but he believes firmly in its limits, particularly in constitutional law, whose open-endedness and emotionally charged content can turn reason into a mere patina, providing a gloss on "reasons" that reason knows not: "A Supreme Court Justice—however questionable his position in a particular case might seem to be—can, without lifting a pen or touching the computer keyboard, but merely by whistling for his law clerks, assure himself that he can defend whatever position he wants to take with enough professional panache to keep the critics at bay."

It is no labor to review a writer as eminently quotable as Judge Posner. Here he is downgrading the "committee of lawyer aristocrats" that make up the Court: "Cocooned in their marble palace, attended by sycophantic staff, and treated with extreme deference wherever they go, Supreme Court Justices are at risk of acquiring an exaggerated opinion of their ability and character." He makes similarly short work of both Antonin Scalia *and* Stephen Breyer, neither of whose apologia for their preferred brands of judging he finds persuasive. He derides Anthony Kennedy—in many respects the acme of a Powellian Justice, obsessed with the moral character of judging—as a mystic and messiah manqué in judge's robes, and adds, of his opinion in *Gonzales v. Carhart*, "What does it tell us about the commitment to legalism of the four most conservative Justices of the Supreme Court that they should have joined such a wild opinion?" Under his pen fall Henry Hart, Erwin Griswold, the former President of the Israeli Supreme Court, the liberal law professors who contested the Solomon Amendment in *Rumsfeld v. FAIR*, and legions besides.

The only difficult task for a reviewer is to thread a path through this book, which is strewn with divagations on topics as diverse as the rules of baseball as applied to Chief Justice Roberts's infamous "umpire" metaphor, the citation of foreign law, and the elasticity of salary incentives for judges. Still, Posner's endpoint is clear enough. However much judges and others may wish to deny it, there is an unavoidable personal and political element in their decisions on the most contested areas of law, especially constitutional law. Although even constrained pragmatism is no antidote, it can at least encourage us to reach our decisions—and analyze the virtues of those decisions—in a clear-eyed way that takes into account all of the internal and external factors that drive them.

There is no doubt that Powell and Posner differ in their approaches and at least some of their conclusions. The books' indexes tell the story well. Posner observes early on that a classic text on American judges lacks entries for "politics" or "ideology." Posner's own book likewise lacks entries for words such as "oath," "honor," and "duty," while Powell neglects a raft of social and economic terms that might have added some leavening realism to his rhapsodic account. Surprisingly, both lack an entry for "Weber, Max." This is a shame (and, where Posner is concerned, a shock), because both might have done well to consider the extent to which either the constitutionally virtuous judge or the brilliant and wide-ranging, if constrained, pragmatist judge has much of a role left to play in an increasingly routinized, bureaucratic, and disenchanted age.

However, despite their wildly divergent paths, Posner and Powell in fact share many common ties. Both, for example, share interesting and important views on how we might reform legal education—in Powell's case, to focus more on "how constitutional questions can be resolved with integrity and their resolution expressed with clarity"; in Posner's, to master conventional legal skills and then move beyond them and focus on the actual underpinnings of judicial decision. Although Posner might reject Powell's approach as tending towards the "moral vanguardism" of a Justice Kennedy, he repeatedly emphasizes that some of the most important constraints on the judicial task are, "first, the desire for self-respect and for respect from other judges and legal professionals generally, which a judge earns by being a good judge, and, second (and closely related), the intrinsic satisfactions of judging, which usually are greater for a good judge than for a bad one." So even a Posnerian judge may have recourse to the kinds of quasi-moral constitutional "virtues" that are at the heart of Powell's work.

Both also share some of the same blind spots. Foremost among them is their regrettable focus on federal appellate judges, particularly the Justices of the Supreme Court. To be sure, as Powell writes, the Court "has by a very large margin the loudest institutional voice in constitutional debate," and it often writes in the most open margins of the law, rendering it a fit subject for Posner's thesis. But if we are to take seriously the importance of virtuous judging (in Powell's case), or judging as a mixture of subtle constraints and motivations (in Posner's case), and of good and bad faith judging in both writers' case, perhaps it would be better to focus on judges whose every action is not in the public spotlight. Those judges may fall prey more easily to the temptation to act in bad faith despite the loose shackle of precedent—especially since, as Posner notes, most lower court decisions are no longer subject to serious review. Both writers, for example, have something to say about race and *Brown v. Board of Education*. But if they are to think seriously about the courts' role and constraints in this area, and to examine whether judges act in good faith in such a controversial field, perhaps they would have been better off examining what the district courts and the old Fifth Circuit did in implementing, extending, and, in the case of many district courts, resisting *Brown*. Powell and Posner's approach is like asking about the honesty of the constabulary by examining the probity of the Chief of Police instead of looking at the lowly, lonely, constantly tempted desk sergeant in the evidence room of the local precinct.

My head is with Posner. Too much is missing from Powell's account. Nowhere in his book do we find a serious treatment of the many real world factors that influence judicial decision making: the role of law clerks as canned reasoners for judges who perforce need do little reasoning of their own; the extent to which judging is a social and collegial process rather than the purely solitary and deliberative act of a cloistered monk in a cell; the host of human motivations and limitations that drive and hem in a judge and impede any effort to set down a reliable instruction manual for any would-be judicial Hercules; and any number of broader institutional factors that might add depth to his romantic view of the judiciary. His constant refrain that his picture of constitutional virtue must be true because "much of what we do and say and do in constitutional interpretation" would otherwise be meaningless (a "solemn mockery") begs the question: What if it is? You cannot prove God's existence by saying that life would be bleak if God did not exist. No more can Powell prove the necessity of his approach by arguing that it would be depressing to think otherwise. Posner writes that "for judges to acknowledge even just to themselves the political dimension of their role would open a psychologically unsettling gap between their official job description and their actual job"; Powell's apparent response is: "So let us then not acknowledge it." But that is not a proof. Powell is mixing his 'is' and his 'ought.'

One might also pick a number of fights with the substantive "virtues" Powell has selected, and some of the vices he deplores. His vision of "acquiescence" is in more tension than he acknowledges with the role of the oath—which requires a judge to exercise in good faith some *independent* judgment about what the Constitution requires. To be sure, precedent plays a part. But a conscientious judge cannot utterly displace his own obligation finally to make a judgment of his own. His privileging of what he calls "the priority of the political"—that is, his view that judges must respect the priority of decision-making in the public and political realm rather than in the courts—may be an unquestionable good as a default position in constitutional judging, but that does not make it a virtue in a deeper ethical sense. Powell's attack on "instrumentalist" judges such as Posner is flatly unconvincing in light of the fuller picture Posner himself paints in *How Judges Think*. As he points out, there is no reason why pragmatism cannot be considerably bounded and constrained, both by the Constitution itself and by the host of traditions and practices in which it is embedded.

Posner has his flaws, too. His non-formalist position on foreign judicial materials, which seems to boil down to "read but don't cite," is close to incoherent. He might also have done well to look for examples of legal and judicial virtue beyond the canonical and unrepresentative Holmes, as Powell does with a 19th century legal opinion by Attorney General Amos Akerman. But as a description of what judges do and what they are likely to continue to do in anything other than the best of all possible worlds, Posner easily wins this match on points.

Still, my heart, I must confess, lies substantially (although not entirely) with Powell. Posner, it would seem, regards naivete as *the* cardinal sin: he would rather explain (and revel in) judges' limits than seek even a tentative path beyond them. This is unfortunate. Virtue ethics' status as a growth stock in the academy has slowly but surely fostered some important efforts to think about virtue and its relationship to the judicial task, as exemplified in recent treatments by Lawrence Solum, Suzanna Sherry, and Daniel Farber. These early efforts are still more shallow than deep, still largely nostrums—more "Law Day rhetoric" than deep treatments of the issue. But they are, I believe, the right path. In that sense, Powell's book, with its yearning and eloquent reach in that direction, is a good step forward.

But much more than idealism is required. A paean to the judicial virtues will have little impact, in light of the much more solid evidence mustered by Posner, without a substantial shift in the culture surrounding it—if not the whole culture, then at least the legal culture. If we are to have the judges Powell wants, honoring and inhabiting the kinds of virtues he lauds, we are unlikely to get them piecemeal. We need generally to recapture a sense of public and private virtue that has largely been lost in the face of the rationalistic and disenchanted spirit of our age—a spirit, it should be said, that is not without virtues of its own. We need to recapture and adapt a sense of the virtues in our own time, and internalize that sense in a way that can actually serve as a meaningful constraint on judicial action. Powell, is right, I think, to hearken back to *Marbury*'s invocation of the oath, for surely obedience to the oath requires something more than shallow loyalty. For those who take it seriously, every judicial act should partake of a quality of commitment and obligation that is distinctly moral; the judge should constantly ask herself whether she is acting in accord with the deeper sense of honor that she puts at stake by taking the oath. But honor is largely an obsolete value today, and it will take considerable imaginative work to retrieve it.

We should not fool ourselves: however Edenic Powell's world may be, it is not the world we live in, and likely never was. We live in Posner's world, with all its human imperfection. No single judicial approach, whether one of legalism or loose constructionism, and no lofty Law Day rhetoric, can bring us out of it. But we can *wish* for Powell's world, and take small steps to make it our own. To bring the judicial virtues to life, however, we will have to take stock honestly of the world we actually live in. Ironically, to get to Powell's virtue, and the virtue of judges and lawyers at the center of Powell's affections, like Marshall and Akerman, we may need more than a little dose of Holmes and Posner and their acid bath. We can and should emulate Powell, but we ought to make sure that at least a small corner of our minds has learned well and retained Posner's lessons.

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., Timothy Lynch
& Peter Thomson, *Vice Chairmen*

Environmental Law & Property Rights
Donald Kochan, *Chairman*

Federalism & Separation of Powers
Ilya Somin, *Chairman*
Geoffrey William 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 Tepp & Adam Mossoff, *Vice Chairmen*

International & National Security Law
Nathan Sales, *Chairman*
Soraya Rudofsky, *Vice Chairman*

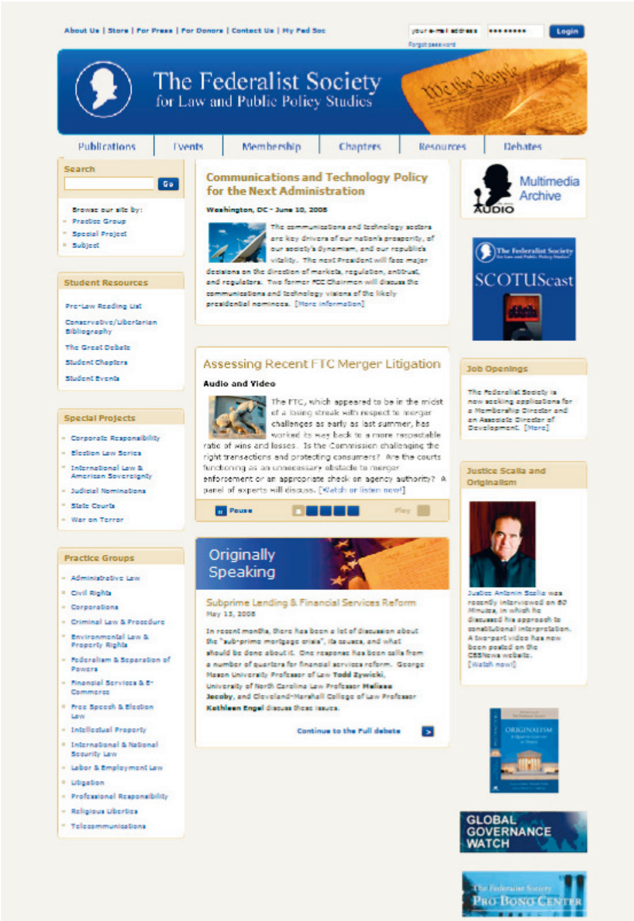
Labor & Employment Law
James A. Haynes, *Chairman*
Raymond J. LaJeunesse, Jr., *Vice Chairman*

Litigation
Margaret A. Little & Christopher C. Wang
Co-Chairmen

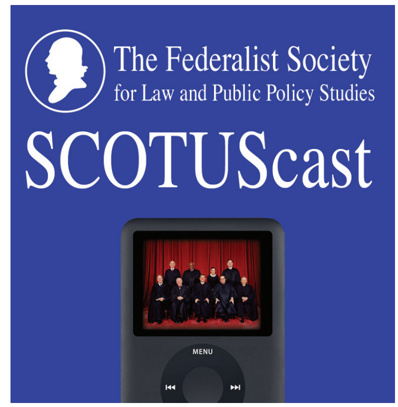
Professional Responsibility & Legal Education
Jack Park, *Chairman*

Religious Liberties
Becky Dummermuth, *Chairman*

Telecommunications & Electronic Media
Raymond L. Gifford, *Chairman*



New Features at The Federalist Society www.fed-soc.org



SCOTUScast
Originally Speaking
Multimedia Archive

The Federalist Society
for Law & Public Policy Studies
1015 18th Street, N.W., Suite 425
Washington, D.C. 20036